

Toetsen op School Hoger onderwijs

Piet Sanders (redactie)



zeker weten

Toetsen op School Hoger onderwijs

Piet Sanders (redactie)

Toetsen op School hoger onderwijs is een uitgave van Cito

- Redactie: Piet Sanders
- Auteurs: Piet Sanders, Bernard Veldkamp, Gerard Straetmans, Brenda Zandsteeg, Marieke van der Schaaf, Casper Hulshof, Frans Prins, Jan Adema
- Tekstredactie: Erna Meijer
- Opmaak: Service unit, MMS
- Foto omslag: Ron Steemers

© Stichting Cito Instituut voor Toetsontwikkeling Arnhem (2016)

Alle rechten voorbehouden. Niets uit dit werk mag zonder voorafgaande schriftelijke toestemming van Stichting Cito Instituut voor Toetsontwikkeling worden openbaar gemaakt en/of verveelvoudigd door middel van druk, fotokopie, scanning, computersoftware of andere elektronische verveelvoudiging of openbaarmaking, microfilm, geluidskopie, film- of videokopie of op welke wijze dan ook. Stichting Cito Instituut voor Toetsontwikkeling Arnhem heeft getracht alle rechthebbenden te achterhalen. Indien iemand meent als rechthebbende in aanmerking te komen, kan hij of zij zich tot Cito wenden.

Inhoud

	Inleiding	8
1	Het doel van toetsen	10
1.1	Toetsen voor het beoordelen van studenten	10
1.1.1	Selecteren van studenten	10
1.1.2	Plaatsen van studenten	12
1.1.3	Certificeren van studenten	12
1.2	Toetsen voor het beoordelen van het onderwijsleerproces	16
1.3	Toetsen voor het beoordelen van groepen studenten en scholen	19
2	De inhoud en constructie van toetsen	22
2.1	Stappenplan voor het ontwikkelen van toetsen	22
2.1.1	Stap 1: Maak een toetsplan	23
2.1.2	Stap 2: Definieer de inhoud	24
2.1.3	Stap 3: Stel de toetsmatrijs op	24
2.1.4	Stap 4: Ontwikkel de items	25
2.1.5	Stap 5: Stel de toets samen	26
2.1.6	Stap 6: Neem de toets af	27
2.1.7	Stap 7: Scoor de antwoorden	27
2.1.8	Stap 8: Bepaal de zak/slaaggrens	27
2.1.9	Stap 9: Rapporteer de scores	28
2.1.10	Stap 10: Documenteer de voorgaande stappen	28
2.2	Combineren van verschillende deeltcijfers tot één eindcijfer	29
2.3	Ten slotte	30
3	Het beoordelen van competentie	32
3.1	Competentie: een verwarrend concept	33
3.2	Beoordelen van competentie: de obstakels	33
3.2.1	Bewijs schiet kwalitatief tekort	34
3.2.2	Te weinig bewijs	35
3.2.3	Gebrekkige scoring	36
3.2.4	Ondeskundige examinatoren	37
3.2.5	Geen of slechte cesuur	39
3.2.6	Onvoldoende zicht op authenticiteit van het bewijs	39
3.3	Competentiebeoordeling vraagt om een assessmentprogramma	40
3.3.1	Meervoudig bewijs	40
3.3.2	Methodemix	41
3.3.3	Inzet van meerdere examinatoren	43
3.3.4	Gestandaardiseerde en transparante scoringsprocedure	44
3.3.5	Transparante beslissingsprocedure	47
3.4	Een assessmentprogramma leidt tot een verantwoord bewijsdossier	48
3.5	Ten slotte	49
4	Het beoordelen van afstudeeropdrachten	52
4.1	Onderzoeken en professioneel handelen	52

4.2	Kenmerken van afstudeeropdrachten	53
4.3	Dilemma's bij de beoordeling van afstudeeropdrachten	54
4.4	Functies van beoordelen bij afstudeeropdrachten	55
4.5	Eisen waaraan de beoordeling van afstudeeropdrachten moet voldoen	55
4.6	Formuleren van beoordelingscriteria	56
4.7	Formuleren van beoordelingsstandaarden	57
4.8	Beoordelingsmodellen voor het combineren van criteria	57
4.9	Individueel beoordelen of groepsbeoordeling	58
4.10	Gebruik van rubrics, peer beoordeling en zelfbeoordeling	59
4.11	Beoordelen en begeleiden bij afstudeeropdrachten	60
4.12	Portfoliobeoordeling	61
5	De examencommissie in het hoger onderwijs	64
5.1	Examencommissie in ontwikkeling	65
5.2	Rol van de examencommissie in de organisatie	67
5.2.1	Examencommissie volgens het NVE-functieprofiel	67
5.2.2	Relatie met andere actoren in de organisatie	68
5.3	Taken van de examencommissie	71
5.3.1	Wettelijke taken	71
5.3.2	Niet-wettelijke taken	73
5.4	Borgen van de kwaliteit van examinering	75
5.4.1	Operationaliseren van de kwaliteit van examinering	75
5.4.2	Risicoanalyse	78
5.4.3	Onderzoeken en analyseren	80
5.4.4	Adviseren van het bevoegd gezag en management	81
5.5	Samenstelling van de examencommissie	81
5.6	Regeling examencommissie en huishoudelijk reglement	82
5.7	Ten slotte	84
6	De kwaliteit van toetsen en examens	86
6.1	NVAO-accreditatie van opleidingen in het hoger onderwijs	86
6.1.1	Beoordelingskader instellingstoets kwaliteitszorg	87
6.1.2	Beoordelingskader beperkte opleidingsbeoordeling	87
6.1.3	NVAO-databank beoordeelde opleidingen	90
6.2	RCEC-beoordelingssysteem voor de kwaliteit van toetsen en examens	93
6.2.1	Criterium 1: Doel en gebruik	93
6.2.2	Criterium 2: Toets- of examenmateriaal	95
6.2.3	Criterium 3: Representativiteit	97
6.2.4	Criterium 4: Betrouwbaarheid	97
6.2.5	Criterium 5: Standaardbepaling en normhandhaving	98
6.2.6	Criterium 6: Afname en beveiliging	100
6.3	NVAO-beoordelingskader en RCEC-beoordelingssysteem	100
	Geraadpleegde literatuur	104
	Afkortingen	114
	Index	116

Inleiding

Inleiding

In de publicatie *Toetsen op School* (herziene versie, 2013) bespreken we in tien hoofdstukken een groot aantal aspecten van toetsen en examineren. De inhoud van *Toetsen op School* behoort volgens Cito tot de basiskennis van iedere (beginnende) docent. Het is de bedoeling om *Toetsen op School* aan te vullen met verschillende publicaties die zich richten op het toetsen en examineren in de verschillende onderwijssectoren. Reeds verschenen zijn de aanvullende publicaties *Toetsen op School primair onderwijs* (2011) en *Toetsen op School voortgezet onderwijs* (2012). De hier voorliggende publicatie betreft *Toetsen op School hoger onderwijs*. *Toetsen op School* middelbaar beroepsonderwijs moet nog verschijnen.

Toetsen op School hoger onderwijs behandelt in zes hoofdstukken verschillende aspecten van het toetsen en examineren in het hoger onderwijs. In hoofdstuk 1 bespreken we het doel van toetsen en examens die in het hoger onderwijs worden afgenomen. Hierbij maken we onderscheid tussen toetsen voor het beoordelen van studenten, toetsen voor het beoordelen van het onderwijsleerproces en toetsen voor het beoordelen van groepen studenten en scholen. Hoofdstuk 2 gaat over de inhoud en constructie van toetsen. De bespreking van beide onderwerpen vindt plaats aan de hand van een stappenplan voor het ontwikkelen van toetsen. Hoofdstuk 3 heeft betrekking op het beoordelen van competentie. Onderwerpen die aan de orde komen zijn het begrip ‘competentie’, de obstakels bij het beoordelen van competentie en de noodzaak voor een assessmentprogramma. Hoofdstuk 4 gaat in op het beoordelen van afstudeerwerkstukken. Aan bod komen onder andere de eisen waaraan de beoordeling van afstudeeropdrachten moet voldoen, beoordelingsmodellen en het gebruik van rubrics. Hoofdstuk 5 gaat over de steeds belangrijkere rol van de examencommissie in de organisatie, de taken van de examencommissie en de relatie met andere actoren in de organisatie. Hoofdstuk 6 gaat nader in op de kwaliteit van toetsen en examens, eerst aan de hand van het accreditiekader van de NVAO en vervolgens, als alternatief of mogelijke aanvulling, aan de hand van het RCEC-beoordelingssysteem.

Voor de lezer die meer wil weten over het toetsen met andere toetsvormen en het toetsen van andere vaardigheden dan die in deze publicatie besproken worden, verwijzen we naar het boek ‘*Toetsen in het hoger onderwijs*’ van Van Berkel, Bax en Joosten-ten Brinke (2014).

Toetsen op School op internet

De publicaties in de reeks *Toetsen op School* zijn online beschikbaar. Ga voor meer informatie en het downloaden van de elektronische versies naar www.toetsenopschool.nl.

1 Het doel van toetsen

1 Het doel van toetsen

Piet Sanders

Ook dit deel van de reeks Toetsen op School begint in paragraaf 1.1 met een bespreking van toetsen die voor verschillende doelen gebruikt worden. Onder een toets verstaan we 'een instrument voor het meten van iemands kennis en vaardigheden (praktische vaardigheden en houdingen) die door middel van studie en/of onderwijs op een of ander vakgebied zijn verworven' (Cito, 2010). We beginnen de bespreking met toetsen die tot doel hebben studenten te beoordelen, dat wil zeggen studenten te selecteren, plaatsen of certificeren. Toetsen voor het beoordelen van studenten worden wel summatieve toetsen of 'toetsen van het leren' genoemd. In paragraaf 1.2 bespreken we toetsen die als doel hebben inzicht te geven in het onderwijsleerproces en dit waar nodig bij te sturen. Deze toetsen worden ook wel formatieve toetsen of 'toetsen voor het leren' genoemd. In paragraaf 1.3 gaan we kort in op het gebruik van toetsen die tot doel hebben de onderwijsprestaties van groepen studenten en scholen te beoordelen. In de publicaties van Toetsen op School voor het primair en voortgezet onderwijs is ook het gebruik van toetsen besproken die tot doel hebben de kwaliteit van het onderwijs te beoordelen. Dit laatste doel van toetsen komt in het hoger onderwijs echter niet voor.

1.1 Toetsen voor het beoordelen van studenten

In het hoger onderwijs kunnen toetsen die tot doel hebben studenten te beoordelen, onderscheiden worden in toetsen waarmee we studenten kunnen:

- Selecteren
- Plaatsen
- Certificeren

1.1.1 Selecteren van studenten

We spreken van selecteren als niet alle studenten die zich voor een opleiding of onderwijsprogramma aanmelden ook worden toegelaten. In het hoger onderwijs dient men over diploma's te beschikken om te worden toegelaten. Wie een opleiding in het hoger beroepsonderwijs (hbo) wil volgen, moet in het bezit zijn van een havo-diploma, een vwo-diploma of een mbo 4-diploma. Voor toelating tot een universiteit is een vwo-diploma of een propedeuse van een hbo-opleiding vereist. Sommige opleidingen hanteren naast diploma-eisen nog extra toelatingseisen. Bijvoorbeeld de eis dat een bepaald profiel onderdeel is van het examen van de vooropleiding van de student. Sommige opleidingen mogen aanvullende eisen stellen, zoals auditie doen of een toelatingstest maken. Zie bijvoorbeeld de toelatingseisen van het Conservatorium van Amsterdam, www.ahk.nl/conservatorium. Informatie over toelatingseisen is te vinden in de 'Regeling nadere vooropleidingseisen hoger onderwijs' op de website van DUO, <http://instellingsinformatie.duo.nl> > Regelgeving OCW, en ook op te vragen bij de onderwijsinstelling.

Het havo- of vwo-diploma geeft recht op toelating tot de meeste opleidingen of studierichtingen maar niet voor studierichtingen met een numerus fixus (Latijn: vastgesteld aantal). Er zijn drie manieren om toegelaten te worden tot een numerus-fixusopleiding:

- Directe toelating tot hbo of universiteit indien de student een gemiddeld eindexamencijfer van 8 of hoger heeft voor vwo, havo of mbo-niveau 4.
- Centrale selectie of gewogen loting. Bij gewogen loting krijgt de student een lotnummer. Op basis van het gemiddelde eindexamencijfer valt de student in een bepaalde lotingsklasse.

Hoe hoger de lotingsklasse, hoe groter de kans op inloting. Een overzicht van lotingstudies is te vinden op de website van DUO, www.duo.nl.

- Decentrale selectie waarbij hogescholen en universiteiten zelf de studenten voor een numerus-fixusopleiding selecteren.

De centrale loting voor numerus-fixusopleidingen wordt vanaf het studiejaar 2017-2018 afgeschaft. In plaats daarvan gaan de onderwijsinstellingen zelf studenten selecteren die een numerus-fixusopleiding willen volgen. Bij de nieuwe, decentrale selectie mag er niet meer alleen gekeken worden naar examencijfers, maar moet ook aandacht zijn voor persoonlijkheidskenmerken en de motivatie van de aspirant-student.

Een voorbeeld van een numerus-fixusopleiding met 100% decentrale selectie is de bachelor Voeding en Gezondheid van Wageningen University, www.wur.nl. In 2015 zijn er in totaal 130 plaatsen voor startende studenten beschikbaar. De decentrale selectie gebeurt op basis van drie onderdelen:

- Het cijfer van het eerste tijdvakexamen vwo-biologie, dit weegt voor 30% mee in de selectie.
- Het cijfer van het eerste tijdvakexamen vwo-scheikunde, dit weegt voor 30% mee in de selectie.
- De uitkomst van de kennismakings- en selectiedag waarbij door middel van een motivatietoets en een inhoudelijke toets motivatie respectievelijk lerend vermogen in kaart worden gebracht. Dit weegt voor 40% mee in de selectie.

Wie 21 jaar of ouder is en niet over de juiste diploma's beschikt, kan toegelaten worden tot hbo of universiteit op basis van een toelatingsonderzoek. Dit toelatingsonderzoek heet in het hbo de 21+-toets en bij de universiteit het colloquium doctum. Bij de Hogeschool Windesheim, www.windesheim.nl, kent het toelatingsonderzoek bijvoorbeeld een algemeen en een opleidingsspecifiek deel. Het algemene deel bestaat uit een capaciteitentest en een Engelse taalvaardigheidstoets. Het opleidingsspecifieke deel varieert per opleiding en bestaat doorgaans uit (kennis)toetsen op eind havo-niveau. Omdat er over het 21+-toelatingsonderzoek geen landelijke afspraken of richtlijnen bestaan, is het resultaat op het toelatingsonderzoek alleen geldig bij de hogeschool waar het toelatingsonderzoek voor een bepaalde opleiding is afgelegd. Bij de Rijksuniversiteit Groningen, www.rug.nl, bestaat het colloquium doctum voor een aantal opleidingen uit een toetsing in drie vakken. Welke vakken dat zijn, hangt af van de opleiding die de student wil volgen. Voor de opleiding Communicatie- en Informatiewetenschappen worden bijvoorbeeld de vakken Nederlands, Geschiedenis en Frans of Duits getoetst. Voor de opleiding Wiskunde wordt geen toets afgenomen maar dient de student over een deelcertificaat wiskunde van het vwo te beschikken. Ook voor het colloquium doctum bestaan er geen landelijke afspraken of richtlijnen.

In het Nederlandse hoger onderwijs wordt een bindend studieadvies (BSA) gehanteerd. Een BSA is een beslissing van de universiteit of hogeschool over de voortgang van de opleiding van de student. Alle studenten krijgen aan het einde van het eerste studiejaar een studieadvies. Bij een negatief advies moet de student met de opleiding stoppen. Studenten ontvangen een negatief advies als zij niet genoeg studiepunten hebben en er geen bijzondere omstandigheden gelden. De Universiteit Twente, www.utwente.nl, hanteert een BSA voor alle opleidingen. Dit betekent dat studenten die in het eerste jaar minder dan 45 studiepunten behalen (van de te behalen 60), de opleiding niet kunnen voortzetten. Het BSA heeft als hoofddoel studenten sneller op de juiste plek te krijgen en de Universiteit Twente is van mening dat een student die in het eerste jaar minder dan 45 van de 60 studiepunten behaalt, niet op de juiste plaats zit en de opleiding moet verlaten. De Universiteit Twente helpt de student die moet stoppen met zijn studie bij het vinden van een meer geschikte opleiding. Bij de Universiteit Leiden, www.leidenuniv.nl, moet de student ook aan het eind van het eerste jaar 45 van de 60

studiepunten behaald hebben. Aan het eind van het tweede jaar moet het eerste jaar behaald zijn en ook minimaal 30 studiepunten uit de twee volgende jaren van de bacheloropleiding. Wordt hier niet aan voldaan, dan mag de student de studie aan de Universiteit Leiden niet voortzetten.

In het verlengde van het BSA is er in het wetenschappelijk onderwijs sinds 2012 ook sprake van een zogenaamde ‘harde knip’. Dit betekent dat het bachelordiploma behaald moet zijn voordat begonnen kan worden met de masteropleiding.

Een BSA kan voordelen hebben, zoals het verbeteren van de prestaties van studenten, het verbeteren van de studiediscipline en studenten stimuleren meer verantwoordelijkheid te nemen. Studentenorganisaties als de Landelijke Studenten Vakbond (LSVb), www.lsvb.nl, verzetten zich tegen het negatief BSA omdat het volgens hen contraproductief werkt en studenten niet de kans geeft om zich te bewijzen. De LSVb stelt dat intensivering van het onderwijs, waarbij de groepsgrootte afneemt en er meer plaats is voor individuele aandacht, een veel betere oplossing is.

1.1.2 Plaatsen van studenten

We spreken van het plaatsen van studenten als studenten afhankelijk van de resultaten op een toets verschillende onderwijsprogramma's volgen die tot hetzelfde diploma of certificaat leiden.

Bij de ITV Hogeschool voor Tolken & Vertalen, www.itv-hogeschool.nl, kan een opleiding tot vertaler en een opleiding tot tolk gevolgd worden. Iedereen die de opleiding tot vertaler wil volgen, wordt aangeraden een plaatsingstoets af te leggen. De resultaten van deze toets geven een goed beeld van de kennis en vaardigheden en een inschatting van de mogelijkheden van de student. Heeft de student de studietaal niet als examenvak gedaan maar hij/zij is wel in het bezit van een havo-, mbo 4- of vwo-diploma, dan is de plaatsingstoets verplicht. De plaatsingstoets is ook verplicht bij instroming in een hoger jaar. De plaatsingstoets bestaat uit diverse onderdelen, zoals het maken van een korte vertaling en het schrijven van een opstel. Het gebruik van een woordenboek is toegestaan bij de toets. De uitslag volgt binnen drie weken, met daarbij een beschrijving van het behaalde niveau. Ook wanneer iemand niet van plan is direct aan de opleiding deel te nemen, kan de toets worden gemaakt. De uitslag blijft een jaar geldig. Tolk kan men alleen worden wanneer men over een zeer goede actieve kennis beschikt van de studietaal en het Nederlands. Daarom wordt eerst drie jaar de opleiding tot vertaler gevolgd en in het derde jaar de voorbereidende cursus algemene tolktechnieken. De verplichte plaatsingstoets voor de opleiding tot tolk bestaat uit drie mondelinge tolkoefeningen die op een digitale voicerecorder worden opgenomen. Daarna geeft het hoofd van de sectie tolken een advies over het al dan niet verder gaan met de opleiding tot tolk.

Bij Wageningen University wordt de Oxford Online Placement Test (OOPT) afgenomen. Deze plaatsingstoets Engels wordt gebruikt om het niveau Engels van studenten vast te stellen. Aan de hand van het resultaat op deze computertoets krijgen studenten advies over het niveau waarop zij het beste kunnen instromen. De resultaten van de OOPT zijn gekoppeld aan het Common European Framework of Reference for Languages (CEFR), een referentiekader voor taalniveaus. Aan de hand van deze indeling kunnen docenten studenten indelen op hun eigen niveau.

1.1.3 Certificeren van studenten

We spreken van het certificeren van studenten als afhankelijk van de resultaten op een of meer toetsen beslist wordt of studenten een onderwijsprogramma al of niet met succes afgesloten hebben. We presenteren hierna een aantal voorbeelden uit het hoger onderwijs.

Examens

De bekendste summatieve of afsluitende toetsen zijn examens. De Onderwijsraad, www.onderwijsraad.nl, gebruikt de term examens voor elke vorm van afsluitende of tussentijdse

toetsing voor het vaststellen van leerresultaten met enig civiel effect. Van de vele functies die aan examens toegedicht worden, beschouwt de Onderwijsraad de kwalificerende functie als de belangrijkste. Bij de kwalificerende functie gaat het met name om het vaststellen van de kennis, vaardigheden en competenties die een student heeft opgedaan aan het eind van een opleiding. Aan de kwalificerende functie ontleen examens hun civiel effect, dat wil zeggen dat de examens erkend worden door de maatschappij, de arbeidsmarkt, vervolgopleidingen of de overheid en dat aan het slagen voor examens en het behalen van het bijbehorende diploma rechten te ontleen zijn. Zo kan de student die de master tandheelkunde heeft behaald zelfstandig het beroep van tandarts uitoefenen of zich specialiseren in een deelgebied van de tandheelkunde zoals parodontologie.

In het hoger onderwijs is bij vele opleidingen sprake van examens op basis waarvan studenten al of niet diploma's verkrijgen. Voor een goed begrip van de vele opleidingen en diploma's is kennis van de structuur van het hoger onderwijs gewenst.

Structuur van het hoger onderwijs

In het hoger onderwijs bestaan de volgende opleidingen:

- associate degree aan hogeschool (2 jaar);
- bacheloropleiding aan hogeschool (4 jaar);
- bacheloropleiding aan universiteit (3 jaar);
- masteropleiding aan hogeschool of universiteit (minimaal 1 jaar).

Associate degree, bachelor en master zijn internationaal erkende titels die gebruikt mogen worden als een opleiding aan een hogeschool of universiteit is behaald, zie www.rijksoverheid.nl > Onderwerpen > Hoger onderwijs. De associate degree wordt behaald aan een hogeschool en een bachelor- of masteropleiding op een hogeschool of universiteit. De associate degree, bachelor- en masteropleidingen aan de hogeschool zijn gericht op beroepsuitoefening terwijl de bachelor- en masteropleidingen aan de universiteit een minder praktisch en meer wetenschappelijk georiënteerd uitgangspunt hebben.

Voor de associate degree, bachelor en master gelden de volgende toelatingseisen:

- associate degree: diploma op minimaal mbo 4-niveau of havo-niveau;
- bacheloropleiding aan hogeschool: diploma op minimaal mbo-niveau of havo-niveau;
- bacheloropleiding aan universiteit: vwo-diploma of een hbo-propedeuse;
- masteropleiding aan hogeschool of universiteit: de hogeschool of universiteit bepaalt de toelatingseisen (meestal is dit een afgeronde opleiding bachelor aan de hogeschool of universiteit).

Studiepunten worden sinds de invoering van de bachelor-masterstructuur (bama) in ECTS uitgedrukt. ECTS is een afkorting van European Credit Transfer and Accumulation System. Een studiepoint of credit staat voor 28 uur studie-inspanning. Een studiejaar komt overeen met 60 studiepunten (ECTS) of 1680 studie-uren; een masteropleiding van vier jaar bestaat dus uit 240 studiepunten. In Nederland is de bamastructuur in het hoger onderwijs ingevoerd in het academisch jaar 2002-2003.

Het doctoraat is het hoogste diploma dat aan een universiteit behaald kan worden.

De bijbehorende titel is doctor. Het behalen van een doctoraat heet promotie, en de persoon die dat doet heet promovendus of promovenda (Latijn: hij of zij die moet promoveren). Promotie in academische zin is het behalen van de academische graad van doctor door het schrijven en publiekelijk verdedigen van een proefschrift, al dan niet met stellingen, aan een universiteit, of althans onder supervisie van een hoogleraar die als promotor optreedt. Het onderzoek voor een promotie wordt veelal aan een universiteit uitgevoerd, maar noodzakelijk is dat niet. Het is ook mogelijk als 'buitenpromovendus' te promoveren, bijvoorbeeld als medewerker van een niet-

universitaire onderzoeksinstelling, zoals een researchlaboratorium in het bedrijfsleven of een ziekenhuis. Een doctoraat kan ook toegekend worden voor bijzondere verdienste, zonder dat de kandidaat een wetenschappelijk werk moet verdedigen. We spreken dan van een eredoctoraat. Om tot de promotie toegelaten te worden, moet de kandidaat eerst het doctoraalexamen, meester in de rechten of ingenieur, dan wel master of science of master of arts, of een daarmee gelijk te stellen (bijvoorbeeld buitenlands) examen hebben behaald. Hoewel niet gebruikelijk, zijn hierop soms uitzonderingen mogelijk. Iemand die geen doctoraalexamen heeft gedaan, kan – in grote uitzondering – toch tot de promotie worden toegelaten als aannemelijk kan worden gemaakt dat de promovendus in staat is zelfstandig wetenschappelijk onderzoek te verrichten en met gerede kans op succes een proefschrift kan voltooien. De promovendus/promovenda dient hiervoor wetenschapsbeoefenaren en een promotor te vinden die (schriftelijk) garant voor hem of haar staan.

Welke titel iemand mag voeren na het afronden van een hbo- of universitaire opleiding, hangt af van het niveau van de opleiding en de studierichting, zie ook www.onzetaal.nl > Taaladvies. Voordat de bamastructuur werd ingevoerd, werden de volgende titels gebruikt: ing. (hbo ingenieur), bc. (overige hbo-opleidingen), ir. (wo ingenieur), mr. (wo rechten), drs. (wo overige opleidingen), dr. (na promotie). Na invoering van de bamastructuur worden onder andere de volgende titels gebruikt: B (hbo-bachelor), M (hbo-master), BA (wo bachelor of arts), BSc (wo bachelor of science), LLB (wo bachelor of laws), MA (wo master of arts), MSc (wo master of science), LLM (wo master of laws), D of PhD (wo na promotie). Merk op dat de internationale titels achter de naam gezet worden en dat Nederlandse en internationale titels niet gecombineerd mogen worden.

Erkenning van Verworven Competenties (EVC)

Behalve door het volgen van een opleiding aan een hogeschool of universiteit kan een diploma ook behaald worden via een EVC-procedure. EVC staat voor het erkennen van verworven competenties. Het gaat hier om kennis en vaardigheden die niet in een formele opleiding zijn verworven. Met behulp van een EVC-procedure is het mogelijk het curriculum vitae voor een sollicitatie te valideren en (nieuwe) loopbaan- en scholingsbeslissingen voor te bereiden. Daarnaast kan EVC als opstap dienen voor herintredende werknemers en werkzoekenden en is het een instrument binnen het personeelsbeleid van werkgevers. Tegelijkertijd is het een instrument voor de overheid om het principe van ‘een leven lang leren’ een stap dichterbij te brengen en de verdere ontwikkeling van de kennissamenleving te stimuleren. Voor examencommissies van opleidingen in het hoger onderwijs vormt de EVC-rapportage (het ervaringscertificaat) de basis om beslissingen te nemen over vrijstellingen als kandidaten na een EVC-procedure besluiten een opleiding te gaan volgen.

Op de website van het Nationaal Kenniscentrum EVC, www.nationaal-kenniscentrum-evc.nl, is een overzicht te vinden van aanbieders van EVC-procedures. Onder de aanbieders bevinden zich, naast de Open Universiteit, geen andere universiteiten en volgens het register zouden er slechts drie hogescholen EVC-procedures aanbieden. Het register is echter niet volledig, hebben we geconstateerd, want de Hogeschool van Amsterdam (HvA) ontbreekt bijvoorbeeld als aanbieder. Op de website van de HvA, www.hva.nl, staat de EVC-procedure beschreven, die bestaat uit vier fasen: 1. oriëntatie en advies (quickscan); 2. samenstellen van een portfolio; 3. assessmentgesprek; 4. EVC-rapportage en besluit examencommissie die vaststelt welke vrijstellingen EVC oplevert als de kandidaat zich inschrijft voor de opleiding en wat het nog te volgen opleidingstraject is. De HvA biedt EVC aan voor 13 deeltijd hbo-bacheloropleidingen.

Informatie over het onderwijs en de examinering van een opleiding in het hoger onderwijs staat beschreven in de onderwijs- en examenregeling (OER) van de betreffende opleiding. Op grond van artikel 7.13 van de Wet op het hoger onderwijs en wetenschappelijk onderzoek (WHW), zie <http://wetten.overheid.nl>, is elke opleiding of groep van opleidingen verplicht om een OER vast te stellen. De OER bevat de hoofdlijnen van vorm en opzet van het onderwijs en de examinering

binnen een opleiding en vormt daarmee het basisdocument voor student en docent. In de OER komen onder meer de volgende onderwerpen aan de orde: de inhoud van de opleiding en de daaraan verbonden examens, de eindtermen, het aantal en de volgtijdelijkheid van de tentamens, de momenten waarop de tentamens afgelegd kunnen worden, de geldigheidsduur van de tentamens en wanneer de student (cum laude) geslaagd is. Een voorbeeld van een OER is te vinden op de website van de Hogeschool Leiden, zie www.hsleiden.nl/toegepaste-psychologie/onderwijs-en-examenregeling.html. Bekijk op deze website ook de bijlagen waar voor elk studiejaar onder andere de onderwijseenheden/vakken vermeld worden, het bijbehorende aantal studiepunten (ECTS) en de naam en vorm van de toets. Voor meer informatie zie ook hoofdstuk 5.

Overall Toets Accountancy (OAT)

De hbo-opleiding Accountancy kent al meer dan 25 jaar landelijke examens. Tot 2004 waren dat 10 verschillende vakgerichte examens, maar sinds 2004 zijn de vakgerichte landelijke examens vervangen door één landelijk OAT. Dat er een landelijk examen is, heeft ermee te maken dat:

- accountancy een wettelijk geregeld beroep is;
- er Europese eisen geformuleerd zijn die vertaald zijn naar eindtermen die bewaakt worden door de Commissie Eindtermen Accountantsopleiding (CEA), www.ceaweb.nl;
- landelijke examens worden gezien als een effectief middel om aan te tonen dat de opleidingen aan de te stellen eisen voldoen (kwaliteitswaarborg).

De OAT bestaat uit een praktijkcase met een conceptjaarrekening. Alle hogescholen met de opleiding Accountancy nemen de OAT in het vierde jaar twee keer af, op hetzelfde tijdstip. De jaarlijkse kosten voor de OAT bedragen rond de 45.000 euro voor 1.600 examenkandidaten.

Interuniversitaire Voortgangstoets Geneeskunde (iVTG)

Vijf medische opleidingen in Nederland passen gezamenlijk voortgangstoetsing toe: interuniversitaire Voortgangstoets Geneeskunde, www.ivtg.nl. De voortgangstoets wordt ieder kwartaal schriftelijk afgenomen, bij vijf instellingen tegelijk, bij alle studenten geneeskunde (± 10.000). De iVTG is een meetinstrument om de kennisprogressie van studenten geneeskunde gedurende hun studie te meten. De voortgangstoets kan gezien worden als een afsluitend examen dat de einddoelen van het curriculum test. Aan de voortgangstoetsen nemen dan ook niet alleen de laatstejaars maar alle studenten van de opleiding deel. Het verschil in kennisniveau uit zich in het verschil in toetsscores; de scores van de laatstejaars studenten zullen over het algemeen hoger zijn dan die van jongerejaars studenten. Deze voortgangstoetsen hebben een aantal tekortkomingen.

Tekortkomingen van de iVTG

- Voortgang of achteruitgang kan niet goed gemeten worden omdat niet bekend is of de items/examens op dezelfde schaal liggen of dat de verschillende examens dezelfde moeilijkheidsgraad hebben. Vanwege dat laatste is waarschijnlijk ook gekozen voor relatieve normering. Dit kan echter betekenen dat een relatief sterke kandidaat kan zakken als hij/zij deel uitmaakt van een relatief sterke jaargroep, terwijl dezelfde kandidaat kan slagen als hij/zij deel uitmaakt van een relatief zwakke jaargroep. Dit is uit het oogpunt van billijkheid onwenselijk.
- Kandidaten kunnen aangeven dat het antwoord op het item juist of onjuist is, of dat zij het antwoord niet weten. De volgende scoringsregel wordt gehanteerd: de score 1 als het antwoord goed is, de score -1 als het antwoord fout is en de score 0 als de kandidaat aangeeft het antwoord niet te weten. Uit onderzoek blijkt dat vrouwen door deze scoringsregel benadeeld worden omdat vrouwen meer dan mannen ervoor kiezen geen antwoord te geven om de eenvoudige reden dat ze het antwoord namelijk niet kunnen weten (Baldiga, 2012). Dit laatste maakt dat dergelijke examens niet valide zijn omdat deze bij vrouwen iets anders meten dan bij mannen.

- De betrouwbaarheid van de examens is relatief laag (0,70 tot 0,80) gezien het grote aantal van 200 items waaruit de toetsen bestaan.

Vanwege deze tekortkomingen is in het SURF-programma Toetsing en Toetsgestuurd Leren onderzocht op welke wijze computergestuurd adaptief toetsen (CAT) bij voortgangstoetsing mogelijk is en kan bijdragen aan kwaliteitsverbetering van deze toets. Uit het SURF-onderzoek 'VGTogether, samenwerken in voortgangstoetsing' (2013) blijkt onder andere dat de betrouwbaarheid van de adaptieve toets gelijk is aan de betrouwbaarheid van een papieren toets met twee keer zo veel vragen en dat de CAT in veel minder tijd afgenomen kan worden dan de papieren toets.

Landelijke hbo-toetsen

Als antwoord op de vraagtekens die gezet werden bij de kwaliteit van de examens en dus de kwaliteit van de hbo-diploma's, stelt de commissie Bruijn in het rapport 'Vreemde ogen dwingen' (2012) voor om te streven naar instellingsoverstijgende toetsing op alle plaatsen waar dat mogelijk is en waar dat niet mogelijk is te werken met 'vreemde ogen'. Dit laatste kan invulling krijgen door bij de toetsing bijvoorbeeld gebruik te maken van tweede beoordelaars of door externe deskundigen deel uit te laten maken van de examencommissie. De commissie Bruijn is van oordeel dat centrale, top down ontwikkelde toetsing – zoals bij de centrale examens voortgezet onderwijs het geval is – niet alleen inhoudelijk onwenselijk is, maar ook organisatorisch aanzienlijk complexer is dan andere vormen van externe validering van toetsing. Bovendien veroorzaakt centrale toetsing meer bureaucratie en administratieve lasten. De minister van Onderwijs onderschrijft het voorstel van de commissie en heeft voor de periode 2013-2016 middelen ter beschikking gesteld om de doelstelling te realiseren dat in 2016 bij alle opleidingen sprake is van een versterking van externe validering van toetsing en examinering, zie ook www.rijksoverheid.nl > Onderwerpen > Hoger onderwijs.

1.2 Toetsen voor het beoordelen van het onderwijsleerproces

In de vorige paragraaf hebben we voorbeelden van summatieve toetsen besproken. In deze paragraaf gaat het over toetsen voor het beoordelen van het onderwijsleerproces. In navolging van de ToetsSpecial 'Leren van toetsen' op de website Toetswijzer, zie <http://www.toetswijzer.nl/berichten/toetspecial-leren-van-toetsen-een-cyclisch-proces>, maken we in deze paragraaf een onderscheid tussen formatieve evaluatie en formatief assessment. We spreken van formatieve evaluatie als het evalueren van de kwaliteit van het onderwijs wordt bedoeld. De term die in dit verband in de onderwijskundige literatuur gebruikt wordt, is opbrengstgericht werken (OGW). Van formatief assessment is sprake als het toetsproces gericht is op leerprocessen binnen de klas en het verbeteren van instructie aan de klas en individuele leerlingen. De termen die in dit kader gebruikt worden, zijn: assessment for learning (AfL) en diagnostisch toetsen (DT).

Hierna presenteren we eerst de belangrijkste kenmerken van de benaderingen OGW, AfL en DT. Die kenmerken zijn grotendeels ontleend aan voornoemde ToetsSpecial. Vervolgens gaan we na hoe deze benaderingen in de praktijk gerealiseerd worden. Daarvoor gebruiken we de publicatie 'Ontwikkelingen in de Delftse Toetspraktijk: de (digitale) toets als diagnose, leermiddel, voortgangscntrole en kwalificatieinstrument' van Keijzer-de Ruiter (2011).

Opbrengstgericht werken (OGW)

'OGW volgt een vaste cyclus van doelen bepalen, data verzamelen, resultaten registreren, resultaten interpreteren en het onderwijs aanpassen. Deze onderwijsaanpassingen kunnen gericht zijn op student-, klas- of schoolniveau. Voor de dataverzameling worden voornamelijk

objectieve, betrouwbare en gestandaardiseerde instrumenten gebruikt, met name wanneer de consequenties van de beslissingen grote gevolgen hebben voor degene over wie de beslissing wordt genomen. Beslissingen die op schoolniveau worden genomen, hebben vaker grotere gevolgen dan beslissingen die leerkrachten dagelijks in de klas maken. Ook zak/- slaagbeslissingen hebben verstrekken gevolgen voor studenten en worden daarom ondersteund met objectieve en betrouwbare datafeedback. OGW is hoofdzakelijk gericht op het verhogen van de uitkomsten van het leerproces en wordt minder gebruikt voor het verbeteren van het leerproces zelf.'

OGW bij de TU Delft

De TU Delft is in 2012 gestart met het hervormen van het onderwijs. De reden hiervoor was het lage rendement van de opleidingen. De lange gemiddelde studieduur van de Delftse student in combinatie met een verhoogde instroom van studenten, legde een zware druk op de beschikbare middelen. De TU Delft is geheel digitale opleidingen gaan aanbieden, omdat daar aanzienlijke besparingen van verwacht worden, zie www.tudelft.nl
> Actueel > Online onderwijs. Bij een geheel digitale opleiding behoort het digitaal afnemen van toetsen (anders leren, anders toetsen) en dat zelfs ook op afstand. Bij de TU Delft kunnen we spreken van OGW omdat hier op basis van informatie over de studieduur besloten is het onderwijs en de toetsing aan te passen. Er is hierbij gekozen voor digitalisering van zowel onderwijs als toetsing. De aanpassingen vinden plaats op het niveau van de onderwijsinstelling.

Assessment for learning (Afl)

'Bij Afl gaat het om het continu verzamelen van bewijs over leervooruitgang van studenten in de alledaagse klassituaties en het inspelen hierop. In Afl is assessment geïntegreerd in de dagelijkse lespraktijk. Het assessmentproces is een sociale activiteit waarbij de nadruk ligt op het verbeteren van het leren binnen de klas. Studenten nemen actief deel aan hun leerproces en kunnen hun eigen leerproces reguleren. Afl heeft dus betrekking op het student- en klasniveau. Afl kan op vele manieren uitgevoerd worden en hier worden niet noodzakelijk schriftelijke toetsen bij gebruikt. Afl vindt ook niet noodzakelijk plaats binnen een bepaalde activiteitscyclus, het is voornamelijk een flexibel en interactief proces. De nadruk ligt vooral op het veelvuldig verzamelen van informatie op allerlei manieren en hier direct op inspelen met aangepaste instructie en feedback. Omdat het in Afl gaat om het creëren van een optimaal leerproces speelt feedback die aansluit op de leerbehoefte van leerlingen een essentiële rol.'

AfL bij de TU Delft

Bij projectgericht leren kunnen docenten studenten stimuleren om de theorie te bestuderen door tussentijds toetsen aan te bieden die voor een deel meetellen voor het eindcijfer. Zo heeft de TU Delft in het vak Façade Design Plus blokonderwijs gerealiseerd in combinatie met de inzet van toetsen als leermiddel.

Bij het vak Façade Design Plus is de theoretische lesstof onderverdeeld in vijf modules die elk afgesloten worden met een toets. Daarnaast maken de studenten als eindopdracht (groepsopdracht) een ontwerp. De afsluitende toetsen bestaan uit open vragen die door gebruikmaking van de peer assessment module uit Blackboard niet door de docent maar door medestudenten beoordeeld worden. De docent heeft voor elke vraag een modelantwoord, beoordelingscriteria en een scoringsmodel gemaakt. De studenten beoordelen toetsen van twee anonieme medestudenten. Indien de scores van de twee beoordelingen te veel verschillen, wordt de betreffende toets nader bekeken.

De gemiddelde score op de vijf toetsen telt voor 25% mee voor het eindcijfer van het vak waardoor de studenten de toetsen ook serieus nemen. Ongeveer 90% van de studenten neemt aan alle toetsen deel. Daarmee is voor de docent het doel – ervoor zorgen dat de studenten de stof bijhouden en er ook nog iets van opsteken – bereikt. De toetsing van de theoretische lesstof bij het vak Façade Design Plus kan opgevat worden als AfL. Door toetsen bij alle modules af te nemen en studenten de toetsen van medestudenten te laten beoordelen, mogen we verwachten dat de studenten de lesstof beter zullen beheersen. Dat de cijfers op de vijf modules meetellen voor het eindcijfer, motiveert de studenten hun best te doen en dat is dan ook een goede toetspraktijk. Geen consequenties verbinden aan toetsen die als leermiddel of voor voortgangscontrolle worden ingezet, is volgens Kamminga (2011) een van de valkuilen van (digitaal) toetsen. Andere valkuilen die volgens Kamminga bij digitaal toetsen vermeden moeten worden, zijn onder andere: uitgebreide oefentoetsen geven, een onbeperkt aantal pogingen aanbieden, en de drempel voor beheersing van de lesstof bij 60% leggen. Het advies is om korte toetsen van vijftien minuten te geven (zie ook Karpicke & Blunt, 2011), het aantal pogingen te beperken tot vier en de drempel voor beheersing van de lesstof bij 80% te leggen.

Diagnostisch toetsen (DT)

‘DT is niet alleen gericht op het meten van het leerproces en de leerprestaties van studenten, maar ook op de mogelijke verklaringen van deze leeropbrengsten. Deze verklaringen kunnen worden gezocht in de voorkennis die studenten hebben of in de manier waarop studenten redeneren. De verklaringen betreffen over het algemeen student- en taakkenmerken die beïnvloeden of een student bepaalde taken (al dan niet) zelfstandig kan uitvoeren. Daarnaast kan een deel van de leerprestaties worden verklaard aan de hand van oplossingsstrategieën en systematische fouten die studenten maken. DT is een cyclisch proces waarbij het leerproces en de leerprestaties van de student op een steeds gedetailleerder niveau beschreven worden. Hierbij gaat het om individuele studenten; de beslissingen die genomen worden over de inrichting van de leeromgeving gaan over individuele studenten. Bij DT wordt veelal gebruik gemaakt van objectieve en gestandaardiseerde toetsen die worden afgenomen op formele toetsmomenten. Er zijn in Nederland nog niet veel diagnostische toetsen ontwikkeld die geautomatiseerd het leerproces en de leerprestaties van studenten in kaart brengen. Docenten kunnen echter ook gebruikmaken van andere assessmentmethoden, zoals diagnostische interviews.’

DT bij TU Delft

In de publicatie van Keijzer-de Ruiter (2011) is sprake van drie verschillende praktijk-situaties waarbij de toets als diagnostisch instrument wordt ingezet. De instaptoets Wiskunde voor eerstejaars bachelorstudenten is bedoeld om de wiskundekennis van het voortgezet onderwijs te reactiveren. Bij het verplichte eerstejaars vak Transport and Separation in de masteropleiding Biochemical Engineering is de instaptoets Wiskunde bedoeld om deficiënties in de voorkennis bij studenten te identificeren en, afhankelijk van de toetsresultaten, het programma aan te passen. De Placement Test Engels wordt gebruikt om de cursisten op basis van hun startniveau in te delen in verschillende cursusgroepen. Een ander voorbeeld van een diagnostische talentoets is DIALANG, European System for Diagnostic Language Assessment, zie www.lancaster.ac.uk/researchenterprise/dialang/about. Deze online diagnostische toets is geënt op het Common European Framework of Reference for Languages (CEFR) dat in Nederland bekend staat onder de naam Europees Referentiekader (ERK). Voor veertien talen zijn diagnostische toetsen beschikbaar voor lezen, luisteren, schrijven, woordenschat en grammatica. Gebruikers krijgen naast de feedback bij hun toetsprestaties (gekoppeld aan het CEFR) ook advies over hoe zij hun huidige vaardigheidsniveau kunnen verhogen. Voor de drie diagnostische digitale toetsen van de TU Delft geldt dat gebruik-gemaakt wordt van het digitale toetssysteem MapleTA (Testing & Assessment). De instaptoets Wiskunde van de TU Delft en DIALANG kunnen als diagnostische toetsen beschouwd worden. Of de Placement Test Engels als een diagnostische toets aangemerkt kan worden, is niet duidelijk. Volgens de terminologie die in dit hoofdstuk gehanteerd wordt, zou het eerder als een plaatsingstoets dan als een diagnostische toets aangemerkt worden.

1.3 Toetsen voor het beoordelen van groepen studenten en scholen

In de publicatie Toetsen op School (Sanders, 2013) presenteren we vijf voorbeelden van toetsen die gebruikt worden voor het beoordelen van groepen leerlingen en scholen. De toetsen die daarvoor gebruikt worden, zijn veelal dezelfde toetsen als die gebruikt worden voor het beoordelen van individuele leerlingen. Geen van die voorbeelden heeft betrekking op de beoordeling van groepen studenten en hogescholen of universiteiten. Daarvoor wordt als reden veelal aangevoerd dat de toetsen van hogescholen of universiteiten niet vergelijkbaar zouden zijn. Dit laatste is echter niet helemaal waar, omdat twee toetsen die we eerder in dit hoofdstuk besproken hebben – de Overall Toets Accountancy en de interuniversitaire Voortgangstoets Geneeskunde – wel identiek zijn over de opleidingen heen. Bij deze twee toetsen zou het dus heel goed mogelijk zijn om de verschillende opleidingen op basis van de behaalde toetsresultaten te vergelijken. Die vergelijking zou het onderwijs bijvoorbeeld goede diensten kunnen bewijzen in het geval opleidingen extreem hoog of laag presteren.

Met de introductie van landelijke opleidingsoverstijgende toetsen in het hbo wordt het mogelijk de prestaties op gelijkwaardige opleidingen op verschillende hogescholen onderling te vergelijken. Dergelijke vergelijkingen zouden kunnen bijdragen aan de externe validering van de toetsen, maar ook aan het vaststellen van verschillen in kwaliteit van de instroom op verschillende hogescholen of de kwaliteit van het onderwijs binnen die hogescholen.

2 De inhoud en constructie van toetsen

2 De inhoud en constructie van toetsen

Bernard Veldkamp

Dit hoofdstuk gaat in op methoden die gebruikt kunnen worden voor de constructie van toetsen. Een belangrijk uitgangspunt hierbij zijn de doelen van de toetsing. Toetsing binnen het hoger onderwijs kan, zoals ook al in hoofdstuk 1 aangegeven is, een formatief of een summatief doel hebben. Formatieve toetsen worden afgenomen om op basis van de resultaten feedback te geven aan studenten, terwijl op basis van summatieve toetsen beslissingen te nemen zijn over het eindniveau van een student aan het einde van een module, vak of opleiding.

We beschrijven het ontwikkelen van toetsen aan de hand van een stappenplan dat uit tien stappen bestaat: van het maken van een toetsplan (stap 1) tot het documenteren van de voorafgaande stappen (stap 10). Dit stappenplan is gericht op het ontwikkelen van individuele toetsen. Binnen het hoger onderwijs worden vakken of modules echter veelal afgerond met een combinatie van verschillende toetsen en opdrachten. Daarom gaan we aan het eind van dit hoofdstuk in op methodes voor het combineren van scores van verschillende toetsen om tot een eindoordeel te komen.

2.1 Stappenplan voor het ontwikkelen van toetsen

Effectieve toetsontwikkeling vraagt om een systematische, gedetailleerde aanpak die ervoor zorgt dat er valide uitspraken mogelijk zijn over het niveau van de studenten. De onderwijs- en examenregeling (OER) geeft hiervoor de kaders. In een OER legt de onderwijsinstelling vast dat voor elk vak, module of onderwijseenheid een functionaris aangewezen moet worden die verantwoordelijk is voor de examinering van het vak. Het is zijn of haar taak om te zorgen voor de toetsontwikkeling en het afleggen van verantwoording hierover aan de buitenwereld. Het gebruik van een stappenplan draagt bij aan een effectieve toetsontwikkeling. Bekende stappenplannen voor het ontwikkelen van toetsen komen aan de orde in Downing en Haladyna (2006) en in Baartman, Bastiaens, Kirschner en Van der Vleuten (2006). Voor het ontwikkelen van toetsen in het hoger onderwijs kunnen de volgende tien stappen onderscheiden worden:

- 1 Maak een toetsplan
- 2 Definieer de inhoud
- 3 Stel de toetsmatrijs op
- 4 Ontwikkel de items
- 5 Stel de toets samen
- 6 Neem de toets af
- 7 Scoor de antwoorden
- 8 Bepaal de zak/slaaggrens
- 9 Rapporteer de scores
- 10 Documenteer de voorgaande stappen

We lichten de verschillende stappen toe en illustreren deze met voorbeelden uit het hoger onderwijs.

2.1.1 Stap 1: Maak een toetsplan

De eerste stap is het vastleggen van het doel van de toets. Welk construct wil je gaan meten en hoe moeten de scores geïnterpreteerd worden? Toetsen kunnen onderscheiden worden in summatieve en formatieve toetsen.

In hoofdstuk 1 is al ingegaan op de drie summatieve doelen:

- selectie;
- plaatsing;
- certificering.

Binnen het hoger onderwijs zijn de meeste toetsen gericht op het certificeren van studenten, dat wil zeggen, op het beslissen of een student een vak of studie al dan niet met succes afgesloten heeft.

Bij formatief toetsen kunnen we, zoals we in hoofdstuk 1 gezien hebben, onderscheid maken tussen:

- opbrengstgericht werken (OGW);
- assessment for learning (AfL);
- diagnostisch toetsen (DT).

Bij opbrengstgericht werken gebruikt de docent de toetsresultaten om de inhoud van het onderwijsprogramma aan te passen. Vanwege de module- of vakkenstructuur van opleidingen in het hoger onderwijs geven de meeste docenten slechts een beperkte periode les aan de studenten en wordt deze vorm van formatief toetsen maar weinig toegepast.

Bij assessment for learning is het de bedoeling dat het leerproces van de student gestimuleerd wordt door het ontvangen en verwerken van de feedback op de toetsresultaten. Deze feedback kan mondeling, schriftelijk of digitaal van aard zijn en afkomstig van medestudenten, docenten of software. Binnen het hoger onderwijs wordt deze vorm van formatief toetsen bijvoorbeeld vaak toegepast binnen (onderzoeks)projecten, waar tussenresultaten of conceptversies van rapporten worden voorzien van commentaar.

Diagnostisch toetsen brengt in kaart op welk onderdeel van de leerstof de student nog deficiënties vertoont. Binnen sommige opleidingen worden academische vaardigheden – zoals schrijven, literatuur zoeken, onderzoeksmethoden – diagnostisch getoetst. De resultaten daarvan laten zien wat een (aankomend) student nog niet beheerst en waar bijscholing nodig is voordat de student bijvoorbeeld kan instromen in een master.

Een van de belangrijkste doelen van toetsing binnen het hoger onderwijs is om aan te tonen dat de student de leerdoelen heeft behaald. Daarom worden de toetsdoelen bij het formuleren direct gerelateerd aan de leerdoelen. Het verschil tussen leerdoelen en toetsdoelen is dat de toetsdoelen ook aangeven hoe aangetoond kan worden dat de leerdoelen behaald zijn.

De toetsdoelen worden vastgelegd in het toetsplan. Daarnaast beschrijft het toetsplan per vak of module op welke manier de student kan aantonen dat hij of zij de aangeboden leerstof beheerst. Het toetsplan kan bijvoorbeeld bestaan uit één afsluitend tentamen dat zowel ontwikkeld als nagekeken wordt door de docent die het vak geeft, en waarbij de student minimaal een zes moet scoren om het vak succesvol af te ronden. Bij modules, die vaak uit meerdere onderdelen bestaan, wordt in het toetsplan beschreven welke tussentoetsen, deelttoetsen, opdrachten en afsluitende tentamens gemaakt moeten worden, wie ze ontwikkelt, wie ze afneemt, wie ze nakijkt en scoort en hoe de verschillende toetsresultaten onderling gecombineerd worden om tot een eindcijfer te komen. Een gedetailleerde uitwerking van het toetsplan zorgt ervoor dat het hele toetsontwikkelingsproces op een efficiënte manier kan verlopen en mogelijk verder te verbeteren valt.

De examinerator van het vak of de module is verantwoordelijk voor het ontwikkelen van het toetsplan. Dit gebeurt vaak in afstemming met het opleidingsdirectoraat. De examencommissie ten slotte, is verantwoordelijk voor het goedkeuren van het toetsplan en voor het afstemmen

van de verschillende toetsplannen binnen een examenprogramma. De examencommissie zorgt ervoor dat er binnen een examenprogramma sprake is van een evenwichtige verdeling van de verschillende (soorten) toetsen over het studiejaar.

2.1.2 Stap 2: Definieer de inhoud

Het toetsplan (stap 1) beschrijft de inhoud op hoofdlijnen in de vorm van toetsdoelen. In stap 2 wordt de inhoud van de toets verder gespecificeerd. Daarbij wordt ook de link tussen de leerstof en de toets vastgelegd. Een duidelijke omschrijving van de inhoud zorgt ervoor dat de docenten die verantwoordelijk zijn voor het schrijven van de toetsvragen of de opdrachten goed geïnstrueerd zijn, zodat zij deze taak ook uit kunnen voeren. Op het moment dat het schrijven van de items en het samenstellen van de toets door dezelfde persoon geschiedt, kan een minder expliciete formulering van de inhoud volstaan.

Vanuit een streven naar een gelijkwaardig niveau van opleidingen in Europa, zijn in 2004 de Dublin-descriptoren (NVAO, 2004) opgesteld die een globale beschrijving geven van het niveau dat een student moet bereiken voor de onderdelen:

- kennis en inzicht;
- toepassen van kennis en inzicht;
- oordeelsvorming;
- communicatie;
- leervaardigheden.

Binnen het hoger onderwijs worden leerdoelen veelal geformuleerd met behulp van de Dublin-descriptoren. Een ander kader voor het formuleren van leerdoelen zijn de Meijers-criteria, die vooral gebruikt worden bij technische opleidingen (Meijers, Borghuis, Mutsaers, Overveld & Perrenet, 2005). Deze criteria expliciteren ook het onderzoeken en het ontwerpen als specifieke criteria van wetenschappelijk onderwijs. Deze descriptoren en criteria zijn vrij globaal gedefinieerd en elke opleiding vult ze in passend bij de inhoud van de verschillende vakken. Bij het vaststellen van de leerdoelen kan uitgegaan worden van het beschikbare lesmateriaal, maar binnen het hoger onderwijs is ook input van experts van belang, omdat zij op de hoogte zijn van de meest recente ontwikkelingen.

Een voordeel van een systematische aanpak en documentatie van de inhoud is dat de inhoudsvaliditeit of de representativiteit van de toets voor het examenprogramma te onderbouwen valt. Omdat de meeste curricula in het hoger onderwijs uit losse vakken of modules bestaan die afzonderlijk getoetst worden, is de inhoudsvaliditeit van groot belang bij een visitatie.

2.1.3 Stap 3: Stel de toetsmatrijs op

Als de inhoud van de toets bekend is, kunnen de specificaties vastgelegd worden. Hierbij valt te denken aan:

- de vorm van de toets;
- het aantal vragen;
- het gebruik van een classificatiesysteem, bijvoorbeeld de taxonomie van Bloom;
- het gebruik van stimuli, zoals grafieken, plaatjes, filmpjes, etc.;
- het antwoordformat;
- de scoring;
- de interpretatie van de scores;
- de benodigde tijd voor het afnemen.

Ook hierbij geldt dat de specificaties gedetailleerder vastgelegd moeten worden naarmate er meer personen bij het ontwikkelen van de toets betrokken zijn en het belang van de toets groter is.

Een van de hulpmiddelen bij de ontwikkeling van een toets is de toetsmatrijs. Een toetsmatrijs is een schematische weergave van de inhoud van de toets waarin voor de verschillende inhoudscategorieën is vastgelegd op welk niveau de toetsing ervan plaatsvindt. Zie figuur 2.1 voor een voorbeeld van een toetsmatrijs.

Inhoudsgebied	Kennis	Toepassing	Evaluatie	Totaal
Toetsplan	4	8	8	20
Inhoud	7	5	7	19
Specificaties	7	9	5	21
Items schrijven	8	10	6	24
Scoring	4	8	4	16
Totaal	30%	40%	30%	100%

Figuur 2.1: Voorbeeld van een toetsmatrijs voor een examen Toetsontwikkeling

Deze matrijs geeft voor de verschillende inhoudsgebieden, zoals vermeld in het examenprogramma, aan hoeveel items er per cognitieve categorie deel moeten uitmaken van de toets. Een groot voordeel van het gebruik van een toetsmatrijs is dat verschillende versies van een toets die erop gebaseerd zijn inhoudelijk vergelijkbaar zijn. Dit is met name van belang als een vak een aantal opeenvolgende leerjaren aangeboden wordt of als bij een examen parallelle versies van een toets gebruikt worden om examenfraude te voorkomen.

2.1.4 Stap 4: Ontwikkel de items

Binnen het hoger onderwijs is een scala aan uiteenlopende toetsvormen te onderscheiden. Populaire vraagtypes zijn meerkeuzevragen, open vragen, essayvragen en opdrachten (Van Berkel, Bax & Joosten-ten Brinke, 2014). Voor al deze vraagtypes geldt, dat het belangrijk is ze ondubbelzinnig te formuleren, dat het taalgebruik aangepast moet zijn aan de populatie en dat de ene vraag geen informatie weg mag geven over de andere vraag. Bij stapelvragen, waarbij het antwoord op de ene vraag nodig is voor het beantwoorden van de volgende vraag, is het bovendien van belang om te kijken of er niet onevenredig veel afhangt van het correcte antwoord op de eerste vraag, of dat er ook met incorrecte antwoorden zinvol verder te werken is. Omdat bij steeds meer studies in het hoger onderwijs Engels de voertaal is, mogen we van de betrokken itemschrijvers verwachten dat zij de vragen in correct Engels kunnen formuleren. Bij meerkeuzevragen is het vervolgens van belang dat de afleiders van voldoende kwaliteit zijn. Ten slotte geldt voor alle vraagvormen dat het aan te raden is om een tweede docent of inhoudsexpert de kwaliteit van de vragen te laten beoordelen vóór het afnemen ervan. Over het formuleren van items is veel literatuur beschikbaar. We verwijzen naar Haladyna en Rodriguez (2013) en de hoofdstukken 6, 7 en 8 van Toetsen op School, zie www.toetsenopschool.nl.

In het derde jaar van de bachelor of in een masteropleiding worden vakken geregeld afgerond met het schrijven van papers, verslagen of scripties. De werkwijze bij het ontwikkelen van deze opdrachten is anders dan bij het ontwikkelen van toetsen waarbij elke deelnemer dezelfde vragen beantwoordt. Het gaat immers vaak om geïndividualiseerde opdrachten. Maar ook bij deze vorm van toetsing is het van belang dat de instructies duidelijk en uniform omschreven zijn en dat van tevoren voor de studenten duidelijk is op welke aspecten ze beoordeeld gaan worden. Bij stages, bachelor- of masteropdrachten worden deze instructies vastgelegd in een

stage-, bachelor- of masterhandleiding. In deze handleiding staat nauwkeurig omschreven wat van de student wordt verwacht en op welke manier de opdracht wordt beoordeeld. Hierbij kan gebruikgemaakt worden van een beoordelingsschaal (rubriek/rubric) waarin voor de verschillende beoordelingscriteria wordt aangegeven hoe het gedrag of de prestatie leidt tot een bepaalde score. Meer informatie hierover is te vinden in hoofdstuk 4 en hoofdstuk 7 van Toetsen op School, zie www.toetsenopschool.nl.

2.1.5 Stap 5: Stel de toets samen

Bij het samenstellen van de toets worden items geselecteerd en op volgorde gezet. Uitgangspunt hierbij is dat voldaan wordt aan de randvoorwaarden die opgesteld zijn in stap 3. Als er één versie van een tentamen of toets gemaakt wordt, die op papier wordt afgenomen, dan wordt de toets meestal handmatig samengesteld door de examinerator. Op het moment dat er parallelle versies gebruikt worden, kan voor het samenstellen van toetsen gebruikgemaakt worden van gespecialiseerde software, zoals de toetsservicesystemen Questionmark Perception en TestVision.

In alle fasen van het productieproces blijft het bewaken van de toetskwaliteit een belangrijke taak van de examinerator. De examinerator dient te letten op de inhoud van de items, maar ook op de psychometrische eigenschappen van een toets, zoals de betrouwbaarheid en de validiteit. Bestaat de toets bijvoorbeeld wel uit voldoende vragen om betrouwbaar te kunnen meten? Een vuistregel die vaak gehanteerd wordt, is dat een toets minstens 60 meetpunten moet hebben om tot een betrouwbare beoordeling te komen.

Het samenstellen van toetsen verloopt geavanceerder als ze adaptief aangeboden worden, bijvoorbeeld de WISCAT-pabo, zie www.cito.nl > Hoger onderwijs > Toetsen voor pabo's. Bij deze vorm van toetsen worden de items op maat geselecteerd tijdens de toetsafname. Op basis van de gegeven antwoorden, wordt een inschatting gemaakt van het beheersingsniveau van de kandidaat. Vervolgens wordt bij dit niveau een item geselecteerd dat de meeste informatie geeft. Hierbij wordt gebruikgemaakt van (itemresponsie) modellen die een psychometrisch kader bieden om computer adaptieve toetsen te ondersteunen. Bij computer adaptief toetsen (CAT) beoordeelt de computer de antwoorden van de student. Deze vorm van toetsing is daarom alleen geschikt voor vragen die objectief scorebaar zijn. Het grote voordeel van CAT is een reductie in de lengte van de toets met 40%. Daarnaast raakt een student niet verveeld of gefrustreerd door vragen die veel te makkelijk of juist veel te moeilijk zijn. Voor het afnemen van adaptieve toetsen is gespecialiseerde software beschikbaar. Het grootste nadeel is echter dat er veel items van uiteenlopende moeilijkheid nodig zijn om alle kandidaten op maat te bedienen. Een vuistregel die hierbij gehanteerd wordt, is dat het aantal items in de itembank minimaal gelijk is aan 12 keer de toetslengte. Voor een toets van 50 vragen heb je dus een itembank nodig met 600 vragen. Deze vorm van toetsen is daarom met name interessant bij grootschalige toetsen die vaker afgenomen worden bij grote aantallen studenten. Een voorbeeld hiervan zijn de in hoofdstuk 1 genoemde voortgangstoetsen (iVTG) die bij verschillende geneeskundefaculteiten jaarlijks worden afgenomen.

Het samenstellen van de toets is over het algemeen geen ingewikkelde stap. Wel is het voor de kwaliteit van de toets van belang dat nauwkeurig gecontroleerd wordt dat de resulterende toets voldoet aan de randvoorwaarden die opgelegd zijn in stap 3. Het verdient dan ook aanbeveling dat een onafhankelijke expert, bijvoorbeeld een collega-docent, mede beoordeelt of aan deze voorwaarden voldaan is. Mocht het zo zijn dat de uiteindelijke toets aan de voorwaarden voldoet, maar naar de mening van de examinerator toch niet optimaal voldoet aan het toetsplan, dan moet stap 3 opnieuw doorlopen worden om de randvoorwaarden aan te passen of verder te expliciteren.

De laatste stap bij het samenstellen van de toets is het maken van de toetsmaterialen. Bij schriftelijke toetsen moeten de materialen voor iedereen goed leesbaar afgedrukt worden. Zie hiervoor ook de criteria zoals die bijvoorbeeld in het RCEC-beoordelingssysteem voor toetsen opgenomen zijn, zie hoofdstuk 6 en www.rcec.nl. Ook voor digitale toetsing zijn richtlijnen opgesteld voor de kwaliteit van het materiaal, zie de standaarden van de International Test Commission (ITC, 2013) of de criteria van het RCEC-beoordelingssysteem. Op het moment dat gebruik wordt gemaakt van de diensten van derden om de toetsmaterialen te produceren, ontstaan bovendien extra kwaliteitsrisico's op bijvoorbeeld menselijke fouten of het voortijdig lekken van de inhoud van de toets richting de kandidaten.

2.1.6 Stap 6: Neem de toets af

De afname van de toets is vaak het meest zichtbare onderdeel van toetsing. Om ervoor te zorgen dat de toetsing zo eerlijk mogelijk verloopt, is het van belang de condities waaronder de toetsing plaatsvindt zo veel mogelijk te standaardiseren. Deze standaardisering geldt niet alleen voor de inhoud van de toets, maar ook voor de testmaterialen, de afnametijd en de omgeving waarbinnen de toets afgenomen wordt. Surveillance speelt daarbij een belangrijke rol. De surveillanten dienen er niet alleen voor te zorgen dat er tijdens de toets niet te frauderen is, zij zijn ook verantwoordelijk voor het bewaken van de condities waaronder de toets plaatsvindt. Als er bijvoorbeeld storend omgevingsgeluid is, kunnen zij proberen de oorzaak hiervan weg te nemen of kunnen ze eventueel zelfs de toetsafname stilleggen. Het bestrijden en voorkomen van examenfraude is niet alleen een taak van de surveillanten. Bij het opstellen van het toetsplan moet al overwogen worden of het, gegeven de afnamecondities, nodig is om te werken met parallelle versies van de toets. Recentelijk zijn ook in het hoger onderwijs de eerste proeven met cameratoezicht gehouden. Hiervan gaat een preventieve werking uit. Maar de opleiding moet zich wel afvragen in hoeverre dit middel nodig is en of het niet te veel inbreuk maakt op de privacy van de niet-frauderende kandidaten.

Ten slotte is het bij de toetsafname van belang om rekening te houden met kandidaten met beperkingen. De meeste opleidingen in het hoger onderwijs kennen regelingen voor kandidaten met bijvoorbeeld dyslexie, dyscalculie of RSI. In het algemeen kan gesteld worden dat er bij het opstellen van dit soort regelingen uitgegaan wordt van het belang van de student en het streven naar vergelijkbare condities voor alle deelnemers.

2.1.7 Stap 7: Scoor de antwoorden

Om de antwoorden te kunnen scoren is een scoringsvoorschrift nodig. Voor meerkeuzevragen geeft het scoringsvoorschrift aan welke van de alternatieven correct is. Bij open vragen wordt een beschrijving van het correcte antwoord gegeven. Binnen het hoger onderwijs wordt ook steeds meer gebruikgemaakt van beoordelingsschalen of rubrics.

Na afname van de toets is het nodig het scoringsvoorschrift te valideren. Daarvoor worden de toetsen, of een steekproef van de toetsen op het moment dat het om grote aantallen gaat, beoordeeld met het scoringsvoorschrift. Bij de validatie wordt gekeken of het scoringsvoorschrift eenduidig en volledig is. Ook wordt de toets- en itemanalyse gebruikt om te kijken of individuele vragen/items geen afwijkend patroon laten zien (Boulet, McKinley, Whelan & Hambleton, 2003). Eventueel worden op basis van deze validatieronde het beoordelingsvoorschrift en de scores aangepast.

2.1.8 Stap 8: Bepaal de zak/slaaggrens

Bij summatieve toetsen en examens dient een *s* te worden vastgelegd. Deze grens geeft aan hoeveel punten een kandidaat gescoord moet hebben om te slagen. Bij veel toetsen in het hoger onderwijs wordt uitgegaan van het principe dat de kandidaat de helft van de vragen correct moet beantwoorden, dan wel de helft van het totale aantal punten moet scoren, om een

voldoende te behalen. Deze keuze is echter niet vanzelfsprekend en eigenlijk volkomen arbitrair. Om de grens te bepalen zou men eigenlijk altijd gebruik moeten maken van standaardbepalingsmethodes. Hierbij wordt op een systematische manier bepaald welk aantal behaalde punten recht geeft op een voldoende. Er bestaan zowel relatieve als absolute standaardbepalingsmethodes.

Bij relatieve methodes wordt een representatieve groep kandidaten geselecteerd. De examinerator bekijkt hun prestaties en identificeert op basis van zijn/haar expertoordeel welk gedeelte van de groep geslaagd en welk gedeelte gezakt is. Blijkt dat 30% van de kandidaten van deze representatieve groep gezakt is dan geldt dit percentage ook als zak/slaagregel voor de totale groep kandidaten. Bij absolute methodes wordt, los van de prestaties van de kandidaten, door een groep inhoudsexperts bepaald hoeveel punten een kandidaat voor een gegeven toets of examen moet behalen om een voldoende te scoren. Veel absolute methodes gaan uit van een hypothetische 'grenskandidaat'. Voor elk van de vragen in de toets wordt ingeschat hoeveel punten de grenskandidaat op deze vraag zou scoren. De som van deze scores wordt vervolgens gebruikt als zak/slaaggrens. Binnen het hoger onderwijs bepaalt de examinerator meestal de zak/slaaggrens met een absolute methode en communiceert die van tevoren met de kandidaten. Hoofdstuk 9 van Toetsen op School, www.toetsenopschool.nl, geeft een overzicht van veel gebruikte standaardbepalingsmethodes en een uitgebreid overzicht is te vinden in Cizek (2001).

Als de zak/slaaggrens vastgelegd is, kan een scoretabel gemaakt worden die voor elk aantal behaalde punten aangeeft welk cijfer erbij hoort. Dit kan op verschillende manier gebeuren. De meest gebruikte methodes zijn een lineaire transformatie of een lineaire met knik. Stel dat er bijvoorbeeld 20 punten van de 27 gescoord moeten worden voor een voldoende, dan kunnen de verschillende scores omgezet worden in cijfers zoals weergegeven in figuur 2.2.

Software voor het omzetten van toetsscores in cijfers is te vinden in hoofdstuk 9 van Toetsen op School, zie www.toetsenopschool.nl.

2.1.9 Stap 9: Rapporteer de scores

Bij het rapporteren van de scores is het van belang dat dit accuraat en op tijd gebeurt. Voor veel toetsen en examens geldt dat de uitslag binnen een concrete periode bekend gemaakt moet zijn. Omdat de cijfers veel impact kunnen hebben voor de kandidaat is een eenduidige en heldere communicatie gewenst. Als er gebruikgemaakt wordt van deelttoetsen of een rubric, moet bovendien duidelijk worden gecommuniceerd hoe de eindscore berekend is op basis van de deelscores.

Ten slotte geldt voor veel toetsen het recht van inzage. Kandidaten kunnen na afloop van de toets de beoordeling inzien en uitleg vragen aan de examinerator. Om dit proces goed te laten verlopen, is het van belang dat duidelijk beschreven is hoe de score samenhangt met de prestatie van de kandidaat. Een uitgewerkt scoringsvoorschrift is hierbij een belangrijk hulpmiddel.

2.1.10 Stap 10: Documenteer de voorgaande stappen

In het kader van de verantwoording is het belangrijk alle stappen goed te documenteren, om vast te leggen waarom bepaalde keuzes gemaakt zijn en om in alle stappen van het proces argumenten te verzamelen die de validiteit en betrouwbaarheid van de toets onderbouwen. Deze documentatie kan door examencommissies en tijdens visitaties gebruikt worden om aan te tonen dat de kwaliteit van de examinering in orde is.

Score	Cijfer lineair	Cijfer lineair met knik
0	1	1
1	1	1,3
2	1	1,5
3	1	1,8
4	1	2
5	1	2,3
6	1	2,5
7	1	2,8
8	1	3
9	1	3,3
10	1	3,5
11	1,5	3,8
12	2	4
13	2,5	4,3
14	3	4,5
15	3,5	4,8
16	4	5
17	4,5	5,3
18	5	5,5
19	5,5	5,8
20	6	6
21	6,5	6,5
22	7	7
23	7,5	7,5
24	8,5	8,5
25	9	9
26	9,5	9,5
27	10	10

Figuur 2.2: Voorbeeld van een scoretabel

2.2 Combineren van verschillende deeltijfers tot één eindcijfer

Bij veel vakken in het hoger onderwijs worden meerdere deeltijfers gecombineerd tot één eindcijfer. De deeltijfers kunnen afkomstig zijn van vergelijkbare onderdelen, maar binnen bijvoorbeeld het projectonderwijs kan het eindcijfer ook gebaseerd zijn op een combinatie van een kennistoets, een individuele literatuurstudie en een project. De kernvraag bij het combineren van deeltijfers is in hoeverre de kandidaat mag compenseren. Vanuit de testtheorie is bekend dat volledige compensatie, waarbij het eindcijfer wordt berekend als het (gewogen) gemiddelde van de deeltijfers, resulteert in het laagste aantal misclassificaties (Douglas & Mislevy, 2010). Als je dus zo weinig mogelijk studenten ten onrechte wilt laten slagen of zakken voor een vak, dan moet je onvoldoendes op deeltolsten laten compenseren binnen het vak. Aan de andere kant wordt het vanuit de opleidingen soms onwenselijk geacht om onvoldoendes toe te staan voor bepaalde kernonderdelen. Denk bijvoorbeeld aan de deeltolsten wiskunde binnen de technische wetenschappen. Ook wordt het soms onwenselijk geacht dat studenten onvoldoendes op inhoudelijke vakken kunnen compenseren met voldoende op vaardigheidsvakken, bijvoorbeeld statistiek, of vice versa. In voornoemde gevallen kan voor het combineren van deeltijfers gebruikgemaakt worden van conjunctieve, complementaire en compensatoire beslisregels. Bij een conjunctieve beslisregel wordt het eindcijfer berekend als een (gewogen)

gemiddelde waarbij alle onderdelen voldoende moeten zijn. Bij complementaire beslisregels moet een aantal onderdelen met een voldoende afgerond zijn. Bij een compensatoire beslisregel hoeft alleen het (gewogen) gemiddelde van de deelscores voldoende te zijn om het vak met een voldoende af te ronden.

Het opleggen van extra eisen aan deoltoetsen, zoals het minimaal moeten behalen van een voldoende voor bepaalde deoltoetsen, of het aanwijzen van specifieke deoltoetsen waartussen gecompenseerd mag worden, resulteert erin dat er meer kandidaten zullen zakken voor het vak (zie ook Van Rijn, Béguin & Verstralen, 2012). Daar komt bij dat elk van de deoltoetsen een meeton nauwkeurigheid heeft. Door een aantal deoltoetsen te combineren wordt de invloed van deze meetfout op het eindcijfer kleiner en reduceert het aantal misclassificaties. Maar als er extra eisen aan de deoltoetsen opgelegd worden, dan heeft dit als gevolg dat het aantal misclassificaties hoger wordt.

2.3 Ten slotte

Binnen dit hoofdstuk is een systematische aanpak beschreven voor het ontwikkelen van toetsen binnen het hoger onderwijs. Daarnaast is aandacht besteed aan verschillende methodes voor het combineren van deeltcijfers tot een eindcijfer. Door al deze aspecten in het oog te houden bij de ontwikkeling van toetsen en de gemaakte keuzes te documenteren, valt aannemelijk te maken dat het eindcijfer voor een vak of module betrouwbaar en valide tot stand gekomen is.

Los van de kwaliteit van de toetsing op het niveau van individuele vakken of modules, blijft het van belang om ook per leerjaar te kijken naar een evenwichtige mix van toetsen en opdrachten, afgestemd op de leerdoelen en competenties die voor de verschillende vakken of onderwijs-eenheden vastgesteld zijn. Zie voor het laatste ook hoofdstuk 3.

Ten slotte is het de taak van de examencommissie (zie hoofdstuk 5) om de kwaliteit van de toetsing binnen een opleiding te monitoren, zowel voor de individuele vakken of modules, als voor de opleiding als geheel. Door toetsen en opdrachten systematisch te screenen kan de kwaliteit van de toetsing worden verantwoord naar de buitenwereld, bijvoorbeeld bij een accreditatie door de Nederlands-Vlaamse Accreditatieorganisatie (NVAO), zie ook hoofdstuk 6. De verschillende stappen, zoals ze zijn beschreven binnen dit hoofdstuk, kunnen hierbij als uitgangspunt dienen.

3 Het beoordelen van competentie

3 Het beoordelen van competentie

Gerard Straetmans

Het motief voor de invoering van competentiegericht opleiden (CGO) is – zeker als het om beroepsonderwijs gaat – altijd helder geweest, namelijk een betere aansluiting van het onderwijs op de beroepspraktijk. Het was een logisch vervolg op projectonderwijs en probleem-gestuurd onderwijs, in die zin dat men met CGO probeerde de theoretische en praktische componenten van de beroepsopleiding nog verder te integreren. Er zijn nu zo'n twintig jaar verstreken sinds de eerste experimenten met CGO en we moeten constateren dat de hooggespannen verwachtingen over het nieuwe onderwijsconcept nog niet zijn waargemaakt. De laatste jaren is er zelfs steeds meer onrust ontstaan over de kwaliteit van het onderwijs in het hoger beroepsonderwijs. Meestal werd die onrust naar buiten gebracht door officiële toezichthouders maar in 2005 waren het zelfs studenten die via de media lieten weten dat ze in hun opleiding nagenoeg niets leerden. Al snel werd het competentiegericht opleiden hiervoor verantwoordelijk gehouden. Met CGO is in principe echter weinig mis. Sinds jaar en dag worden beroepsopleidingen geacht competente beroepsbeoefenaren af te leveren. Dat zijn mensen die adequaat kunnen handelen in taaksituaties waarmee ze in hun beroep of functie geconfronteerd kunnen worden. Zo bekeken is het (hoger) beroepsonderwijs dus altijd al competentiegericht geweest. Waar wél iets mee mis is, is de wijze waarop aan CGO gestalte is gegeven door individuele opleidingen. Het feit dat critici CGO automatisch in verband brengen met zaken als weinig contactonderwijs en weinig theorie, doet vermoeden dat financieel-economische overwegingen zwaarder gewogen hebben bij de opzet van dergelijke curricula dan vakinhoudelijke en didactische overwegingen. En niet zelden is er ook iets mis met de wijze waarop bij CGO getoetst wordt of de opleidingsdoelen bereikt zijn. Zo constateerde de Onderwijsraad (2004, pag. 80) dat de vertaling van competenties naar het curriculum lang niet altijd systematisch en transparant was en vaak gebeurde vanuit vakperspectieven en zich mede daardoor beperkte tot kenniselementen. Niet vreemd dus dat de toetsing voor het leeuwendeel traditioneel van aard bleef en niet aansloot op de nieuwe onderwijsdoelen (competenties). En nu, zo'n tien jaar later, is hierin nog altijd onvoldoende verbetering gekomen. Niet voor niets heeft de Nederlands-Vlaamse Accreditatieorganisatie (NVAO) recent het aspect toetsing en beoordeling tot een kritisch criterium verheven bij de accreditatie van opleidingen. En niet voor niets dat een door de Vereniging Hogescholen ingestelde commissie van experts een protocol heeft opgesteld voor het beoordelen van (kern)werkstukken (Expertgroep Protocol, 2014).

In dit hoofdstuk gaan we in op de specifieke problematiek die speelt bij het beoordelen van competentie en op mogelijke oplossingen daarvoor. Verwacht geen snelle, simpele oplossingen; competentiebeoordeling is geen kwestie van even iemand beoordelen die met een authentieke taak bezig is. Zo'n aanpak lijkt misschien heel valide, maar leidt in de praktijk zelden tot een gerechtvaardigde beslissing over competentie. In paragraaf 3.2 bespreken we de belangrijkste obstakels die het trekken van deugdelijke conclusies over competentie kunnen belemmeren. In paragraaf 3.3 schetsen we de contouren van een oplossing. Maar eerst is het nodig wat woorden te wijden aan het begrip competentie zelf. Valide conclusies over de verwerving van competenties vergen een helder inzicht in de aard en reikwijdte van dit begrip bij alle betrokkenen.

3.1 Competentie: een verwarrend concept

In onderwijskringen wordt niet altijd zorgvuldig omgegaan met de terminologie. Dat is eens te meer duidelijk geworden in de verwarring die is opgetreden rondom het begrip competentie. Geconfronteerd met de zeer uiteenlopende wijzen waarop onderwijsinstellingen dit begrip interpreterden, besloot de Onderwijsraad in 2002 een studie te laten uitvoeren met als doel te komen tot één voor het hele onderwijs geldende definitie. De onderzoekers die deze studie uitvoerden concludeerden dat dit, gegeven de diversiteit van de toepassingen in de diverse onderwijssectoren, niet mogelijk was en stelden daarom een flexibele definitie voor. Het begrip competentie werd beschreven aan de hand van zes dimensies, drie noodzakelijke en drie relevante, die elk in meerdere of mindere mate ‘van toepassing’ kunnen zijn. De onderzoekers illustreerden deze flexibele definitie aan de hand van een audioversterker die de gebruiker in staat stelt de dimensies van het geluid (volume, balans, hoge tonen, lage tonen) elk afzonderlijk af te stemmen op grond van de eigen voorkeur (Onderwijsraad, 2002, pag. 72 e.v.).

Het rapport van de Onderwijsraad heeft ertoe geleid dat onderwijzend Nederland kon berusten in de verschillende visies die onderwijssectoren hadden op het competentiebeprip. Maar het heeft niet kunnen voorkomen dat er nog steeds grote onduidelijkheid is over de aard en reikwijdte van het begrip, vooral bij degenen die primair verantwoordelijk zijn voor het welslagen van CGO: de docenten. Zo roept het woord competentie in het hbo associaties op met: (1) beheersing van de kerntaken van een beroep of functie, (2) Dublin-descriptoren en (3) hbo-kernkwalificaties. Eerstgenoemde associatie heeft betrekking op specifieke opleidingsdoelen, de twee andere zijn generieke opleidingsdoelen die niet zo zeer te maken hebben met wat de gediplomeerde precies moet kunnen, maar vooral met het niveau van zijn functioneren. De specifieke en generieke opleidingsdoelen zijn conceptueel te onderscheiden, maar in praktische zin niet te scheiden. Hbo-kernkwalificaties als ‘toepassing van de wetenschap’ en ‘creativiteit en complexiteit in handelen’ en Dublin-descriptoren als ‘oordeelsvorming’ kunnen alleen in een context verworven en gedemonstreerd worden. Afzonderlijke toetsing van Dublin-descriptoren en hbo-kernkwalificaties is dan ook niet zinvol. Hun belangrijkste toepassing is aan te geven op welke kwaliteiten – Kayzel (2005) heeft het over ‘normen voor adequaat handelen’ – gelet moet worden om te beoordelen of kandidaten bij de uitvoering van de kerntaken het gewenste niveau bereiken. Het zijn dus eerder prestatiecriteria dan competenties. Volgens deze redenering moet het bij competentiebeoordeling in het hbo primair gaan om de vaststelling of de kandidaat de kerntaken van een beroep of functie beheerst. De definitie die we in dit boek gebruiken voor het begrip competentie luidt daarom als volgt:

Competentie

Competentie behelst de bekwaamheid om zodanig te handelen in taaksituaties die kenmerkend zijn voor een kerntaak van een beroep of functie, dat zowel het handelen zelf (proces) als de daaruit voortkomende resultaten (product) voldoen aan de kwaliteitsnormen, te stellen aan een beginnend professional.

3.2 Beoordelen van competentie: de obstakels

De kwaliteitseisen voor competentiebeoordelingen zijn in principe dezelfde als waar alle onderwijskundige meetinstrumenten aan moeten voldoen. Voor elk onderwijskundig meetinstrument geldt in de eerste plaats dat het valide scores moet genereren. Het gaat bij validiteit dus niet om de meetinstrumenten zelf, maar om de interpretatie van de

meetresultaten en het gebruik daarvan (Kane, 2006). Of zoals McMillan zegt: ‘Validity is concerned with the soundness, trustworthiness, or legitimacy of the claims or inferences that are made on the basis of obtained scores. In other words, is the interpretation made from test results reasonable? Is the information that I have gathered the right kind of evidence for the decision I need to make or the intended use?’ (McMillan, 2011, pag. 68). Anders gezegd houdt validiteit in dat we erop kunnen vertrouwen de juiste beslissing te hebben genomen, ofwel dat we er vrij zeker van zijn dat de kandidaat die een voldoende resultaat heeft behaald op een kennis- of praktijktoets adequaat kan functioneren in een of andere vervolgsituatie waarover de toets beoogt een uitspraak te doen. Dat klinkt eenvoudig en voor de hand liggend maar in de praktijk is dat vertrouwen maar moeizaam te verkrijgen, zeker bij ingewikkelde onderwijsdoelen als competenties. Diverse obstakels kunnen opdoemen die het vertrouwen in de deugdelijkheid van de te nemen beslissing doen afnemen. In figuur 3.1 wordt het vertrouwen gesymboliseerd door het schijnen van de zon en de obstakels door de wolken die voor de zon schuiven. Die obstakels komen hierna één voor één aan de orde. Ze geven niet alleen aan wat er mis kan gaan bij het interpreteren en gebruiken van een toetsscore, maar ook waar bij het ontwikkelen van instrumenten en procedures voor het vaststellen van competentie op gelet moet worden.



Figuur 3.1: Obstakels voor deugdelijke beslissingen bij toetsing van competentie

3.2.1 Bewijs schiet kwalitatief tekort

Competent zijn, moet blijken uit het feit dat een persoon adequaat handelt in relevante taaksituaties. Daartoe beoordeelt een deskundige examiner de kwaliteit van de taakuitvoering (proces) en het daaruit voortkomende resultaat (product). De beoordeelde prestaties vormen het bewijs waarmee de kandidaat kan aantonen competent te zijn. De kracht van het bewijs hangt in eerste instantie af van de keuze voor de opdrachten of taken die in het kader van de beoordeling moeten worden uitgevoerd.

Meer dan veertig jaar geleden gaven Fitzpatrick en Morrison (1971) aan dat het bij de kwaliteit van performance assessments feitelijk maar om één ding gaat, namelijk de representativiteit. Daarmee bedoelden ze de overeenkomst tussen wat beoordeeld wordt en dat waarover men een uitspraak wil doen op grond van de beoordeling. De mate van overeenkomst laat zich volgens deze auteurs langs twee dimensies beschrijven: natuurgetrouwheid (fidelity) en volledigheid (comprehensiveness).

De natuurgetrouwheid is een belangrijk kwaliteitsaspect omdat niet de toetsprestatie op zichzelf interessant is, maar de voorspelling die ermee gedaan kan worden over het functioneren in een of andere vervolgsituatie (een werksituatie bijvoorbeeld of een vervolgopleiding). Hoe ‘echter’ de assessmenttaak en -uitvoering, hoe beter de voorspelling. De betere voorspelbaarheid bij een hoge natuurgetrouwheid van de assessmenttaak vindt zijn oorzaak in het feit dat bij natuurgetrouwe assessmenttaken de kans groter is dat alle relevante aspecten van de te beoordelen bekwaamheid aan bod komen. Dat is wat Fitzpatrick en Morrison bedoelden met de dimensie ‘comprehensiveness’.

Hoewel de meeste docenten sympathiseren met natuurgetrouwe assessmenttaken, komt het in het onderwijs regelmatig voor dat competenties worden vastgesteld op basis van assessmenttaken die naar inhoud en vorm maar weinig gelijkenis vertonen met de taken die in de vervolgsituatie uitgevoerd moeten kunnen worden. De competentie om spoedeisende hulp te verlenen, bijvoorbeeld, kan niet adequaat worden vastgesteld met een reeks kennisvragen over spoedeisende hulp. Kennis is zonder twijfel een belangrijk aspect van deze competentie, maar niettemin een onvoldoende basis voor het competente handelen in situaties waar spoedeisende hulp vereist is. Toch was, en wellicht is, dit dé manier waarop veel artsopleidingen deze competentie vaststellen. Zelfs in een opleiding waarin basisartsen zich kunnen specialiseren als verlener van spoedeisende hulp denkt men op deze wijze deugdelijke beslissingen over studenten te kunnen nemen, zoals blijkt uit de volgende passage in het onderwijsreglement van de Nederlandse Vereniging van Spoedeisende Hulp Artsen: ‘De competentie medisch handelen kan goed getest worden door middel van een schriftelijke toets.’ Geen wonder dat enige jaren terug in de pers gemeld werd dat twee van de drie aankomende artsen niet in staat zijn om goede eerste hulp te verlenen. Immers, van de afsluitende toetsing gaat een sturende werking uit op het leren van de studenten en, natuurlijk, ook op het onderwijzen van de docenten (Frederiksen, 1984; Birenbaum & Dochy, 1996). Wie alleen kennis toetst, moet niet verrast zijn als studenten zich bij de voorbereiding op die toets beperken tot het kunnen onthouden en begrijpen van de kenniselementen van een vakgebied.

Een ander voorbeeld van waar het niet helemaal goed gaat, komt uit de lerarenopleiding voor het basisonderwijs. Vierdejaars studenten krijgen als LIO (leraar in opleiding) de integrale verantwoordelijkheid voor een groep leerlingen gedurende een aantal maanden. In die tijd moeten ze een portfolio opbouwen dat voldoende bewijzen bevat om aan te tonen dat ze belangrijke docentcompetenties verworven hebben. Onderzoek doet vermoeden dat het opgevoerde bewijs vaak zwak is (Straetmans, Roelofs & Peters, 2011, pag. 10): ‘Eindoordelen over de didactische bekwaamheid zijn [daardoor] in bijna de helft van de gevallen gebaseerd op zwak bewijs (gesprekken en verslagen blijken populairder dan de methodes die op directe wijze zicht geven op de bekwaamheid om adequaat te handelen in onderwijsleersituaties).’ Natuurlijk gaan die gesprekken en verslagen vaak (maar niet altijd) over het handelen van de student in didactische situaties, maar het probleem hier is dat het didactisch handelen zelf niet meer waar te nemen is door degene die het uiteindelijke oordeel moet geven. Deze baseert zijn oordeel op de interpretaties van anderen, die niet per se even deskundig zijn.

3.2.2 Te weinig bewijs

Veel docenten weten tegenwoordig dat een toetsmatrijs (zie hoofdstuk 2) een krachtig hulpmiddel is om toetsen samen te stellen die zowel betrouwbare als valide scores kunnen produceren. In een toetsmatrijs wordt aangegeven met hoeveel opgaven elke doelstelling of onderwerp getoetst moet worden. Deze handelwijze leidt tot toetsen die dekkend zijn voor de te toetsen leerstofdomeneinen. Maar als het gaat om het beoordelen van competenties, lijkt dit inzicht plotseling vergeten. Te vaak komt het voor dat de conclusie over het al dan niet verworven zijn van een competentie getrokken wordt op grond van één assessmenttaak. Stel dat de competentie op het gebied van het adviseren over hypothecaire leningen beoordeeld

zou worden aan de hand van één geobserveerd adviesgesprek tussen de kandidaat en een cliënt. Daar zou men ernstige bedenkingen bij kunnen hebben. Niet alleen omdat je een keer pech of juist geluk kunt hebben, maar vooral omdat één adviesgesprek een slechte afspiegeling is van het assessmentdomein. Met het assessmentdomein bedoelen we alle verschillende adviesgesprekken die in principe gevoerd zouden kunnen worden tijdens een assessment. Het maakt nogal wat uit of je een advies moet geven over een hypothecaire lening die aflossingsvrij is, of over een waarbij volgens een bepaald schema wordt afgelost. Ook de persoonlijke omstandigheden van de cliënt maken verschil voor de wijze waarop het adviesgesprek verloopt: vast of tijdelijk arbeidscontract, hoog of laag inkomen, alleenverdiener of meeverdienende partner, etc. Ten slotte hangt het advies ook af van economische ontwikkelingen en het daarop gebaseerde overheidsbeleid inzake de aftrekbaarheid van hypotheekrente. De grote variatie in adviesgesprekken maakt het onmogelijk om op basis van één beoordeeld adviesgesprek een uitspraak te doen over de competentie 'adviseren over hypothecaire leningen'. Immers, de benodigde kennis en vaardigheden voor het ene adviesgesprek kunnen sterk afwijken van die welke nodig zijn in een ander.

Uit de literatuur blijkt dat het probleem van de taakspecifieke prestaties in alle domeinen speelt (Linn & Burton, 1994). Taakspecifieke prestaties zijn een gevolg van het feit dat kennis en vaardigheden niet zonder meer overdraagbaar zijn van de ene toepassingscontext naar de andere. Eraut en Du Boulay (2000) verwoorden dit zogeheten transferprobleem als volgt: 'Knowledge is acquired in a particular context and remains situated in that context until it can be transformed and resituated in another context' (pag. 43).

Te weinig tijd en geld zijn de belangrijkste redenen waarom er bij de beoordeling van competenties vaak (te) weinig assessmentopdrachten worden ingezet. Het probleem bij assessmentopdrachten is dat ze niet zelden veel tijd kosten aan voorbereiding en uitvoering. Denk bij de voorbereiding bijvoorbeeld aan de ontwikkeling van een draaiboek waarin nauwgezet staat voorgeschreven hoe een bepaalde taaksituatie te creëren waarin de kandidaat zijn bekwaamheid moet tonen. Bij de uitvoering kan het bijvoorbeeld gaan om de inzet van personen die bepaald gedrag proberen uit te lokken bij de kandidaat en, natuurlijk, om één of meer examinatoren die het gedrag en de resultaten daarvan beoordelen. Soms ook zijn er hoge kosten verbonden aan de uitvoering van assessmentopdrachten die niet direct te koppelen zijn aan de inzet van personen. Bijvoorbeeld als er dure materialen verbruikt worden, of wanneer de inzet van dure apparatuur noodzakelijk is die dan niet voor de normale werkzaamheden van de organisatie beschikbaar is. Het gevolg van de noodgedwongen beperking van het aantal assessmentopdrachten is een onvoldoende dekking van het assessmentdomein. En de implicatie daarvan is dat de geldigheid van de uitslag zich beperkt tot de aangeboden taken en niet gegeneraliseerd kan worden over alle taken die samen het assessmentdomein vormen. Een conclusie trekken over het verworven zijn van de betreffende competentie is dan onverantwoord.

3.2.3 Gebrekkige scoring

Of complexe onderwijsdoelen (zoals competenties) bereikt zijn, wordt vastgesteld door het handelen van de kandidaat in relevante taaksituaties te beoordelen. Het abstracte karakter van deze complexe onderwijsdoelen dwingt de beoordelaar tot verregaande interpretaties van de gemaakte observaties; competent handelen openbaart zich meestal niet in gemakkelijk observeerbaar en eenduidig te scoren gedrag. Zorgvuldig geconstrueerde scorerubrieken zijn nodig om de informatie te vergaren op basis waarvan een deugdelijke beslissing over de kandidaat genomen kan worden. Een scorerubriek is een scoringsvoorschrift dat bestaat uit prestatiecriteria (de formulering van wat iemand moet kunnen) aan de hand waarvan prestaties van kandidaten beoordeeld kunnen worden (Butler & McMunn, 2006).

Te vaak zijn de scorerubrieken die in onderwijscontexten gebruikt worden van onvoldoende kwaliteit. Daar zijn twee oorzaken voor te noemen.

In de eerste plaats proberen docenten, bewust als ze zijn van de problemen die gepaard gaan met het beoordelen van processen en producten, vooral prestatiecriteria te formuleren die zo min mogelijk geïnterpreteerd hoeven te worden door een examiner. Jammer genoeg beperkt men zich daarmee tot eigenschappen en kenmerken die gemakkelijk waarneembaar zijn en dat zijn lang niet altijd de aspecten die het verschil maken tussen competent en incompetent handelen. Neem bijvoorbeeld de situatie dat de zangprestatie van een conservatoriumstudent beoordeeld moet worden. Met de moderne ICT-middelen waarover we tegenwoordig kunnen beschikken, is het vrij eenvoudig om een gezongen lied op te nemen en dit met gespecialiseerde software te analyseren op kenmerken als toon, pitch, timbre en volume. Met enige inspanning lukt het ook nog wel om voor een specifiek lied per passage de gewenste meetwaarden ten aanzien van die kenmerken te beschrijven waarmee de weg naar een objectieve beoordeling van een creatief, vluchtig product zou zijn vrijgemaakt. Gelukkig willen de meesten niet zover gaan; het staat veel te ver af van waar het bij een lied eigenlijk om moet gaan, namelijk of het mooi klinkt en of je erdoor geraakt wordt. Maar vanwege het subjectieve karakter wordt dit in het geval van een formele beoordeling meestal geen goed idee gevonden. Er zal daarom een tussenweg gevonden moeten worden; een die de te beoordelen competentie niet kapot analyseert en tegelijkertijd zo veel steun biedt aan de examiner dat de beoordeling niet te veel mening en te weinig feitelijk is. Wat dit voorbeeld duidelijk wil maken is dat volledig objectieve beoordeling van competenties geen doel is dat koste wat kost nagestreefd moet worden. Belangrijker is dat dié aspecten beoordeeld worden die kwaliteitsverschillen in processen en producten bewerkstelligen.

In de tweede plaats ontlenu docenten de prestatiecriteria vaak aan de eerder door henzelf of collega's ontwikkelde assessmentopdracht(en). Het risico daarvan is dat er prestatiecriteria worden opgesteld die te veel taakspecifiek zijn en daardoor te weinig informatie kunnen genereren over het uiteindelijke doel van de beoordeling: de vast te stellen competentie (Popham, 2007, pag. 183). Een voorbeeld kan helpen om dit te verduidelijken. Een rijproef op een voor normaal verkeer afgesloten circuit kan nuttig zijn om heel gericht informatie te verzamelen over de rijvaardigheid van personen. Echter, wanneer men voor de beoordeling van zo'n rijproef enkel prestatiecriteria gaat formuleren die zeer nauw aansluiten bij de uit te lokken vaardigheden (bijvoorbeeld: hellingproef, uitwijkmanoeuvre voor een plotseling de weg overstekend kind, voorrangsregels toepassen op rotonde met veel auto's en fietsers) dan is de kans groot dat er te weinig informatie wordt vergaard over de rijvaardigheid in algemene zin. De reden hiervoor is dat deze prestatiecriteria te veel zijn afgestemd op de specifieke assessmenttaken in plaats van op de onderliggende competentie die kandidaten in staat stelt om in 'alle mogelijke assessmenttaken gericht op rijvaardigheid' adequaat te handelen.

Het ontwikkelen van goede scorerubrieken (of voor wie het eenvoudig wil houden: checklists of beoordelingsschalen) vraagt om een systematische aanpak. Nitko en Brookhart geven een uitstekende handleiding (2011, pag. 262 e.v.), maar zie ook hoofdstuk 4.

3.2.4 Ondeskundige examinatoren

Dit obstakel kent bijna iedereen die onderwijs genoten heeft uit eigen ervaring. Het is van alle tijden dat onderwijsdeelnemers bij tijd en wijle ontdekken dat dezelfde prestatie bij docent A een hogere of lagere waardering krijgt dan bij docent B. In vele studies, uitgevoerd in verschillende domeinen en uiteenlopende opleidingsniveaus, is het probleem van de gebrekkige interbeoordelaarsovereenstemming aangetoond (o.a. Van Trijffel et al., 2010; Hollingworth et al., 2006; Goldman, 1994). Een deel van de problematiek kan op het conto van de gebruikte scorerubrieken geschreven worden. Wanneer die ruimte voor interpretatie laten, wat bijna altijd

het geval is, is er een aanzienlijke kans dat twee of meer examinatoren dezelfde prestatie verschillend beoordelen. Beoordelaarsverschillen kunnen echter ook de examiner als bron hebben. Zo is het bijvoorbeeld mogelijk dat de examiner nooit geleerd heeft wat goed beoordelen inhoudt. Beoordelen blijkt een vak apart te zijn, iets dat je niet zo maar kunt omdat je nou eenmaal docent bent. Wie de eisen ziet die aan een gecertificeerd examiner gesteld worden, zal dat onmiddellijk beamen, zie het functieprofiel van examiner van de Nederlandse Vereniging voor Examens (NVE) op www.assessorentraining.org/functieprofiel.pdf. Beoordelaarsverschillen kunnen ook het gevolg zijn van 'personal-bias errors'. Dat zijn 'zwakheden' die de beoordelingen van de examiner systematisch doen afwijken van het gemiddelde oordeel van andere examinatoren. Sommige daarvan komen zo regelmatig voor dat ze in de literatuur beschreven zijn (Haladyna, 1997). In tabel 3.1 worden de meest bekende kort beschreven.

Tabel 3.1: Frequent optredende personal-bias errors

Personal-bias error	Omschrijving
Halo-/Horneffect	Het komt voor dat een te beoordelen proces of product op één of enkele aspecten zeer positief of negatief scoort. Die kwaliteit overstraalt (vandaar de naam van dit effect) de objectieve kwaliteit van de andere aspecten waardoor de examiner deze onbewust te hoog (Halo) of te laag (Horn) waardeert.
Normeringseffect	Als volledig objectieve beoordeling geen haalbare kaart is, krijgen examinatoren ruimte om hun eigen standaarden te hanteren. Sommigen zijn geneigd tot hoge standaarden, anderen tot lage en weer anderen laten zich nooit verleiden tot extreme beoordelingen.
Vergelijkingseffect	Objectief beoordelen wordt extra lastig als kandidaten onmiddellijk na elkaar beoordeeld worden. Examinatoren zijn geneigd om na een zwakke kandidaat een iets beter presterende kandidaat duidelijk hoger te waarderen (of na een sterke kandidaat een minder presterende kandidaat lager te waarderen).
Persoonlijkheidseffect	Door opvoeding en afkomst heeft een examiner persoonlijke opvattingen, attitudes, ervaringen en voorkeuren. Die spelen vaak onbewust mee bij de beoordeling van anderen. Zo is het bijvoorbeeld niet toevallig dat mensen met niet-westerse achternamen minder vaak een oproep krijgen voor een sollicitatiegesprek.

Het optreden van personal-bias errors is te bestrijden door te zorgen voor heldere, eenduidig te interpreteren prestatiecriteria en een procedure die voorschrijft hoe prestaties van kandidaten beoordeeld moeten worden. Bij dit laatste kan gedacht worden aan het scheiden van de opleidings- en beoordelingsfunctie of, als dat niet haalbaar is, aan het anoniem beoordelen van werkstukken. Minstens zo belangrijk is het dat examinatoren berekend zijn op hun taak. Veel docenten in het hoger onderwijs hebben echter geen lerarenopleiding genoten en soms ook geen basiskwalificatie onderwijs gehaald. Docenten met een formele kwalificatie zijn vaak nauwelijks beter geprepareerd aangezien er voor het onderwerp 'toetsing en beoordeling' maar weinig tijd is ingeruimd in de curricula van de lerarenopleidingen. Training van examinatoren is voorlopig de aangewezen weg om examinatoren snel om te leren gaan met prestatiecriteria en beoordelingsprocedures, maar tevens om ze ervan bewust te maken dat hun oordelen vertekend kunnen zijn door het optreden van personal-bias errors. De ervaring leert dat docenten de noodzaak van training pas echt in gaan zien als ze geconfronteerd worden met de verschillende beoordelingen toegekend door examinatoren aan dezelfde assessmentprestaties. Training, zelfs als er een certificering aan gekoppeld is, mag echter niet beschouwd worden als een definitieve oplossing voor het probleem van de gebrekkige interbeoordelaarsovereenstemming. Er is

blijvende aandacht nodig voor de kwaliteit van beoordelingen en van tijd tot tijd zal het nodig zijn om die kwaliteit op empirische wijze vast te stellen, bijvoorbeeld door een steekproef van beoordeelde studentprestaties door andere examinatoren opnieuw te laten beoordelen.

3.2.5 Geen of slechte cesuur

Bij het beoordelen van competentie gaat het erom vast te stellen of het functioneren van de kandidaat voldoet aan de geldende prestatiestandaard. Anders gezegd, of de kandidaat in staat is om adequaat te handelen bij het uitvoeren van de kerntaken van een beroep of functie. Wat adequaat handelen precies inhoudt is, als het goed is, nader omschreven in de scorerubriek waarmee examinatoren de prestaties van kandidaten beoordelen. De scores die een examiner toekent, krijgen echter pas betekenis als ze worden afgezet tegen een kritische score, ook wel cesuur genoemd. De cesuur is niets anders dan de kwantificering van de prestatiestandaard. Dit kwantitatief maken van de prestatiestandaard is een van de lastigste aspecten van toetsconstructie. Vaak wordt er te weinig aandacht aan besteed of kiest men er gemakshalve voor om de cesuur, net als bij kennistoetsen die op een schoolcijferschaal gescoord worden, bij 55% van de maximum score te leggen (zie ook hoofdstuk 2). Een bedenkelijke zaak, want een verkeerd op de scoreschaal gelegde cesuur kan ernstige consequenties hebben. Als de cesuur te laag op de schaal wordt gelegd, zullen relatief veel kandidaten slagen en een deel daarvan ten onrechte. Ten onrechte zakken leidt tot gefrustreerde studenten, slechte rendementscijfers van de opleiding en verspilling van talent in het algemeen. Ten onrechte slagen leidt in eerste instantie tot blijdschap, maar vroeg of laat zullen zowel de betreffende studenten, de onderwijsinstelling als de samenleving geconfronteerd worden met de negatieve gevolgen van onvoldoende gekwalificeerde professionals.

Lezers die meer inzicht willen krijgen in de gevolgen van een verkeerde cesuur treffen in de ToetsSpecial op de website Toetswijzer, <http://toetswijzer.nl/berichten/toetsspecial-toetsen-op-school-hoger-onderwijs>, een simulatie aan die voor elke ingevoerde cesuurscore het aantal terecht en onterecht geslaagde en gezakte kandidaten berekent, gegeven de (eveneens in te stellen) psychometrische kwaliteit van het gebruikte assessmentinstrument. De simulatie maakt duidelijk dat een slecht gekozen cesuurscore ernstige gevolgen kan hebben in termen van ten onrechte geslaagde en gezakte personen, zelfs als de toets aan hoge psychometrische kwaliteitseisen voldoet.

3.2.6 Onvoldoende zicht op authenticiteit van het bewijs

Bij het thema competentiebeoordeling komt vaak de authenticiteit van het bewijs ter sprake. Hier wordt niet altijd hetzelfde onder verstaan. Authenticiteit kan verwijzen naar de kwaliteit van het bewijs in de zin van dat de assessmenttaken en de omstandigheden waaronder ze zijn uitgevoerd sterk lijken op die in de (te voorspellen) criteriumsituatie (Wiggins, 1993). De in paragraaf 3.2.1 genoemde term ‘natuurgetrouwheid’ (fidelity) wordt in dat verband vaak als synoniem gebruikt. Maar het kan ook gaan om het vertrouwen dat men kan stellen in de herkomst van het bewijs. Sinds de opkomst van het populaire beoordelingsportfolio, dat studenten een grote mate van vrijheid geeft om zelf te bepalen hoe ze hun competentie(s) willen aantonen, is de onzekerheid over de herkomst van bewijzen toegenomen. In een recent uitgevoerd onderzoek onder vierdejaars pabo-studenten gaf 5% van de ondervraagde respondenten toe één of meer bewijsstukken in hun portfolio te hebben opgenomen die niet van eigen makelij waren en 54% vermoedde of wist zeker dat anderen zich daar schuldig aan maakten (Straetmans, Roelofs & Peters, 2011).

Het zal duidelijk zijn dat valide beslissingen over het al dan niet verworven zijn van competentie(s) onmogelijk zijn als de herkomst van het bewijs niet te verifiëren is. De oplossing is dat opleidingen weer meer regie gaan voeren over de wijze waarop studenten hun competentie(s) aantonen.

Dit betekent dat niet elke prestatie automatisch als bewijsstuk in een portfolio kan worden geaccepteerd.

3.3 Competentiebeoordeling vraagt om een assessmentprogramma

De beschreven obstakels geven niet alleen aan hoe het mis kan gaan bij competentiebeoordeling, ze geven ook de richting aan waarin de oplossing moet worden gezocht. Die oplossing heet ‘assessmentprogramma’. Kort gezegd houdt dit in dat voor een competentiebeoordeling niet één maar meerdere bewijzen nodig zijn, waaraan bovendien kwaliteitseisen gesteld moeten worden. Het beste bewijs wordt in principe verkregen door observatie van taakuitvoering op de werkvloer, zoals mogelijk is tijdens (eind)stages. Immers, zowel de opdracht als de omstandigheden waaronder die moet worden uitgevoerd, hebben bij stagewerkzaamheden vaak een sterke gelijkenis met de taaksituaties waarmee de beginnend beroepsbeoefenaar in een reële werksituatie (de criteriumsituatie) geconfronteerd kan worden. Jammer genoeg is observatie op de werkvloer om allerlei redenen lastig te realiseren. Tijdgebrek bij examinatoren, het gevaar dat de werkzaamheden of de omstandigheden waaronder ze worden uitgevoerd kunnen opleveren voor de kandidaat en/of diens omgeving, de kosten of organisatorische problemen die de werkzaamheden met zich meebrengen: allemaal redenen waarom er minder vaak geobserveerd wordt op de werkvloer dan vanuit het oogpunt van een goede beoordeling wenselijk zou zijn. Om toch voldoende bewijs te verzamelen is de inzet van andere assessmentvormen – waarbij concessies worden gedaan aan de authenticiteit van de taak, de hulpmiddelen en/of de werkomstandigheden – acceptabel. Een belangrijke vraag die in dit verband opkomt, is of prestaties die het resultaat zijn van verschillende opdrachten en verschillende assessmentvormen, zinvol samengevat kunnen worden om een beslissing te nemen over het al dan niet verworven zijn van de betreffende competentie. Het gevaar dat er ‘appels en peren worden opgeteld’ is in zo’n situatie niet denkbeeldig. Die dreiging kan ingedamd worden door de prestaties die als bewijs voor de verwerving van een bepaalde competentie worden opgevoerd, steeds aan de hand van dezelfde prestatiecriteria te beoordelen. Dit is in beknopt bestek waar het bij een assessmentprogramma om draait. We gaan in deze paragraaf meer in detail in op de specifieke kenmerken van zo’n assessmentprogramma.

3.3.1 Meervoudig bewijs

Eerder definieerden we competentie als de bekwaamheid van de beginnend beroepsbeoefenaar om adequaat te handelen bij de uitvoering van de kerntaken van een beroep of functie. Binnen een kerntaak kan zich een heel scala aan taaksituaties voordoen. Soms zijn de verschillen tussen die taaksituaties zo groot dat er wezenlijk andere aanpakken nodig zijn, die op hun beurt weer geënt kunnen zijn op (deels) andere kennis. Een concreet voorbeeld kan helpen om de problematiek te verduidelijken. In figuur 3.2 worden enkele taaksituaties gespecificeerd waarmee artsen die spoedeisende hulp verlenen geconfronteerd kunnen worden. Het zal ook voor leken duidelijk zijn dat prestaties die in verschillende taaksituaties geleverd zijn niet zomaar uitwisselbaar zijn. Bij het verlenen van spoedeisende hulp aan een patiënt die getroffen is door een hartstilstand (taaksituatie A) wordt een beroep gedaan op heel andere vakinhoudelijke kennis en procedures dan bij een patiënt met een gebroken sleutelbeen die panisch gedrag vertoont (taaksituatie B). Om deugdelijke conclusies te trekken over het competente handelen in spoedeisende hulsituaties is het daarom nodig om te beschikken over prestaties in een reeks van taaksituaties die samen representatief zijn voor alle taaksituaties die zich bij het uitvoeren van de kerntaak kunnen voordoen. Dit betekent dat er bij het aanleggen van een bewijsdossier nauwkeurig gelet moet worden op de representativiteit van het vergaarde bewijs. Richtlijnen vanuit de opleiding over de samenstelling van het bewijsdossier zijn daarbij noodzakelijk.

		Situatie / conditie:						
		In de EHBO-post	Langs de snelweg	In drukke winkelstraat	Agressieve omstanders	Patiënt is panisch	Patiënt is stervende	Enz.
Effectief en efficiënt handelen bij:	Bewusteloosheid							
	Hartstilstand			A				
	Herseninfarct							
	Hersenvloeding							
	Fractuur / trauma					B		
	Bevalling							
	Slagaderlijke bloeding							
	Vergiftiging							
	Enz.							

Figuur 3.2: Taaksituaties die het criteriumdomein beschrijven van de competentie *Spoedeisende Hulp Verlenen*. Het behaalde prestatieniveau in taaksituatie A is niet vanzelfsprekend generaliseerbaar naar andere situaties, zoals taaksituatie B.

3.3.2 Methodemix

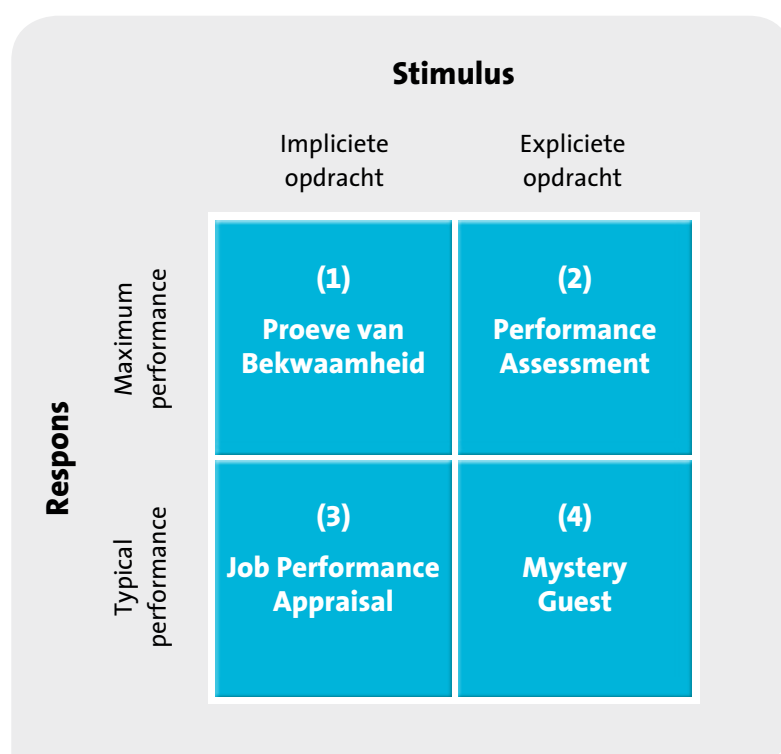
Naarmate een assessmenttaak een hogere natuurgetrouwheid heeft, biedt de score betere mogelijkheden om voorspellingen te doen over het functioneren van de beoordeelde persoon in de criteriumsituatie. Bij competentiebeoordeling zou daarom in principe de keuze moeten vallen op een assessmentvorm die gedrag uitlokt dat zo veel mogelijk lijkt op het gedrag in de criteriumsituatie. Deze principiële keuze betekent echter niet dat andere, minder authentieke of natuurgetrouwe assessmentvormen niet bruikbaar zouden zijn. De ideale assessmentvorm bestaat namelijk niet. 'Basically, you can't win!', zei Kane (1992) daarover. Hij bedoelde daarmee dat de betrouwbaarheid en validiteit van assessmentsscores op gespannen voet staan, waardoor pogingen om de een op te voeren vaak gepaard gaan met reducties van de andere. Zo introduceert het voordeel van de hoge natuurgetrouwheid helaas ook een nadeel, namelijk minder mogelijkheden voor standaardisatie. Daarnaast blijken natuurgetrouwe assessmentvormen vaak meer tijd te kosten waardoor de kandidaat minder verschillende taaksituaties kan uitvoeren. Dit brengt de generaliseerbaarheid over het assessmentdomein in gevaar.

Het principe van de 'methodemix' biedt een uitweg uit dit dilemma (Straetmans & Sanders, 2001). Dit principe houdt in dat heel gericht die assessmentvormen worden ingezet die, gelet op de beschikbare middelen, op de meest efficiënte wijze prestaties (gedrag en producten als resultaat van dat gedrag) uitlokken die scoorbaar zijn aan de hand van de geformuleerde prestatiecriteria.

Welke zijn die andere assessmentvormen? Er worden in de literatuur tientallen methodes genoemd (o.a. Jaspers & Van Zijl, 2011; Van Berkel, Bax & Joosten-ten Brinke, 2014), maar de meeste verschillen niet wezenlijk van elkaar.

Straetmans (2014) beschrijft vier methodieken waarmee het competente handelen van personen beoordeeld kan worden (zie figuur 3.3). De vier methodieken onderscheiden zich van elkaar in de stimulus en de respons. De stimulus bestaat uit de opdracht(en) en de verdere relevante informatie, waaronder de instructies die een kandidaat krijgt bij een assessment. Een stimulus bij een assessment kan expliciet of impliciet zijn. Een stimulus is explicieter naarmate de opdrachten verder zijn uitgewerkt, waardoor het gewenste gedrag en de daaruit voortkomende resultaten beter en efficiënter uit te lokken zijn. De respons betreft de uitvoering van de opdrachten door de kandidaat in de vorm van het beantwoorden van de vragen, of het uitvoeren van de opgedragen taken op basis van de verstrekte informatie. Als de kandidaat zich ervan bewust is dat hij beoordeeld wordt (bijvoorbeeld door de aanwezigheid van een examiner), dan zal die zich normaal gesproken maximaal inspannen om tot een zo goed mogelijke prestatie te komen. Er is dan sprake van 'maximum performance'. Is de kandidaat zich niet bewust van een beoordelingssituatie, dan wordt 'typical performance' waargenomen. De kandidaat vertoont dan het gedrag dat representatief is voor zijn reguliere prestatieniveau.

Door de aard van de prestatie (maximum versus typical performance) af te zetten tegen de aard van de stimulus (impliciet versus expliciet), ontstaat de indeling in figuur 3.3.



Figuur 3.3: Indeling van assessmentvormen op basis van gedrag van de kandidaat en aard van de stimulus

In cel 1 van figuur 3.3 vallen assessmentvormen waarbij de kandidaat weet dat hij tijdens het uitvoeren van de assessmenttaak beoordeeld zal worden. De opdracht kan echter niet tot in detail worden uitgewerkt vanwege de complexe en niet-beheersbare context waarin deze moet plaatsvinden. Dergelijke assessments worden in het onderwijs vaak benoemd als 'proeve van bekwaamheid'. Denk bijvoorbeeld aan de LIO-stage. De assessments in deze categorie worden gekenmerkt door een hoge natuurgetrouwheid, maar ook door een zekere inefficiëntie, omdat de opdracht voor de kandidaat niet gedetailleerd is uitgewerkt. Het gevolg daarvan is dat een examiner soms erg lang moet observeren voordat het gedrag waarnaar de interesse uitgaat zich voordoet.

In cel 2 is het performance assessment ondergebracht. Het betreft hier allerlei assessmentvormen waarbij kandidaten onder meer of minder gecontroleerde omstandigheden werken aan meer of minder authentieke taken. De kandidaat weet dat hij tijdens het reageren op de stimulus beoordeeld zal worden. In tegenstelling tot de proeve van bekwaamheid kan bij deze categorie door middel van gerichte opdrachten het gewenste gedrag (en de daaruit voorkomende resultaten) uitgelokt worden. De meeste sturing valt daarbij te geven in de hands-off variant; de minste bij de hands-on variant (Straetmans, 2006). Van de hands-off variant is bijvoorbeeld sprake bij de 'situational judgment test', waar kandidaten aan moeten geven wat de beste manier van handelen is in een beschreven taaksituatie (zie voor een voorbeeld de COVT-demovideo Medical Emergency, te raadplegen in de ToetsSpecial op de website Toetswijzer, <http://www.toetswijzer.nl/berichten/toetsspecial-toetsen-op-school-hoger-onderwijs>).

Een bekend voorbeeld van de hands-on variant was het praktisch deel van het rijexamen van enkele jaren geleden, waarbij de examiner de kandidaat voortdurend nauwkeurige opdrachten gaf over de te volgen route of uit te voeren verrichtingen. Inmiddels is het rijexamen vernieuwd en moet de kandidaat gedurende een deel van de tijd zelfstandig een route rijden van locatie A naar B.

Een belangrijk nadeel van de expliciete uitlokking van gedrag door middel van opdrachten en instructies is dat daarmee geen antwoord verkregen wordt op de vraag of de kandidaat de wil of bereidheid en het inzicht heeft om in taaksituaties op eigen initiatief tot actie over te gaan.

De assessmentvormen in cel 3 worden vaak gebruikt bij de inservice beoordeling van beroepsbeoefenaren, al dan niet in opleiding. Hierbij kan gedacht worden aan zowel objectieve als subjectieve methodes om zicht te krijgen op het prestatieniveau van functionarissen. Bij objectieve methodes wordt de nadruk gelegd op de opbrengst van een functionaris in termen van geld of producten. Dat klinkt misschien heel vertrouwenwekkend, maar we moeten ons realiseren dat de hoogte van opbrengsten maar zelden volledig kan worden toegeschreven aan de prestaties van een individuele functionaris. Subjectieve methodes maken gebruik van examinatoren die het functioneren van werknemers observeren en beoordelen. Denk hierbij bijvoorbeeld aan systemen voor 360° feedback. Bij de assessmentvormen in cel 3 is de kandidaat er zich niet van bewust dat en/of wanneer hij beoordeeld wordt en ook is vaak niet duidelijk wat exact het gewenste gedrag is. Dit type assessment wordt gekenmerkt door een hoge natuurgetrouwheid, maar helaas ook door matige of zelfs slechte mogelijkheden voor standaardisatie. Verder doen ze een groot beroep op het geheugen van de examiner.

Bij de assessments in cel 4 is wederom sprake van welomschreven opdrachten. Dit keer niet voor de kandidaat, maar voor de persoon die tot taak heeft het gewenste gedrag uit te lokken: de mystery guest. Ook deze vorm van assessment leent zich goed voor toepassing tijdens stages. Meer informatie over de mogelijkheden en ontwerprichtlijnen van mystery guest beoordelingen in het hbo is te vinden op www.saxion.nl/onderwijsinnovatie > Lectoraten > Assessment > Publicaties > Interne publicatie: [Brochure Mystery Guest](#).

De assessmentvorm in cel 1 is in principe te verkiezen boven de altijd wat kunstmatige methodes in cel 2. De methodes in cel 3 en 4 zijn in onderwijssituaties vooral van belang om zicht te krijgen op de attitudinale aspecten van competentie. Ze zijn bij uitstek geschikt om een beeld te krijgen van de werkelijke beroepshouding van beroepsbeoefenaren (in opleiding).

3.3.3 Inzet van meerdere examinatoren

Het belang van de rol die de examiner in een assessmentprogramma vervult, is moeilijk te overschatten. Hij is het die de kwaliteit van de prestaties uitdrukt in scores, op basis waarvan conclusies over de verwerving van de competentie getrokken worden. Een deugdelijke examiner is iemand die zijn vak beheerst, autoriteit geniet bij kandidaten (wat nodig is voor de acceptatie

van het oordeel), zo min mogelijk belang heeft bij de uitslag en voldoende kennis en vaardigheid heeft op het gebied van het beoordelen van gedrag, schriftelijke werkstukken en beroepsproducten.

In de praktijk blijkt het beoordelen van gedrag en producten dermate lastig te zijn dat zelfs de inzet van getrainde examinatoren die in hun vakgebied als expert bekend staan, geen garantie biedt op een juiste beoordeling (Straetmans, 1985). Het beste oordeel is daarom het gemiddelde oordeel van een groep getrainde en deskundige examinatoren. Maar in de onderwijspraktijk stuit de inzet van meerdere, onafhankelijk van elkaar beoordelende examinatoren zowel organisatorisch als financieel op bezwaren. Tot op heden zijn verreweg de meeste prestatiebeoordelingen daarom gebaseerd op het oordeel van één examinator. De verwachting is overigens dat daar snel verandering in gaat komen omdat de commissie Bruijn (2012) de dringende aanbeveling heeft gedaan om bij belangrijke prestaties een tweede beoordelaar van buiten de instelling in te zetten.

De tijd zal leren of dit een haalbare aanbeveling is. Het lijkt verstandig daarnaast andere maatregelen te nemen die de beoordelingskwaliteit kunnen bevorderen. Bijvoorbeeld door te besluiten om competentie te beoordelen op basis van een assessmentprogramma. In een assessmentprogramma levert een kandidaat meerdere prestaties om zijn competentie op een bepaald terrein aan te tonen. Daarbij wordt aanbevolen om zo veel mogelijk prestaties door steeds een andere examinator te laten beoordelen als het niet haalbaar zou zijn om elke prestatie door minimaal twee onafhankelijk van elkaar werkende examinatoren te laten beoordelen. De aan de persoon van een specifieke examinator gebonden beoordelingsfouten beperken zich dan tot één of hooguit enkele prestatiebeoordelingen.

3.3.4 Gestandaardiseerde en transparante scoringsprocedure

Competentiebeoordeling, zo hebben we hiervoor duidelijk gemaakt, is vooral een kwestie van veel informatie verzamelen over de prestaties van kandidaten in relevante taaksituaties. Maar hoe kan die veelal ongelijksoortige informatie (als gevolg van verschillende opdrachten; variatie in de authenticiteit van de taak, omstandigheden, apparatuur en materialen; verschillende examinatoren) zodanig samengevat worden dat er op transparante wijze een eindconclusie aan te verbinden valt over de te beoordelen competentie? Het antwoord op deze vraag luidt: door de verschillende prestaties van de kandidaat steeds met dezelfde scorerubriek te beoordelen. Een scorerubriek bestaat uit een reeks criteria en ondersteunende regels en hulpmiddelen om de prestatie van een kandidaat te beoordelen op een assessmenttaak of open antwoordvraag (AERA, 1999). Een terechte vraag in dit verband is of een scorerubriek die zo algemeen inzetbaar is niet te veel generiek en te weinig taakspecifiek zal zijn. Om dat probleem op te lossen suggereerde Straetmans (2006) om voor elke te beoordelen competentie een scorerubriek te ontwikkelen die zowel generieke als taakspecifieke informatie genereert, zonder dat dit extra werk met zich meebrengt voor examinatoren. Het werkt als volgt: de scorerubriek aan de hand waarvan een competentie beoordeeld wordt, bestaat uit een reeks beoordelingsaspecten, prestatie-indicatoren en beoordelingsschalen.

Beoordelingsaspecten

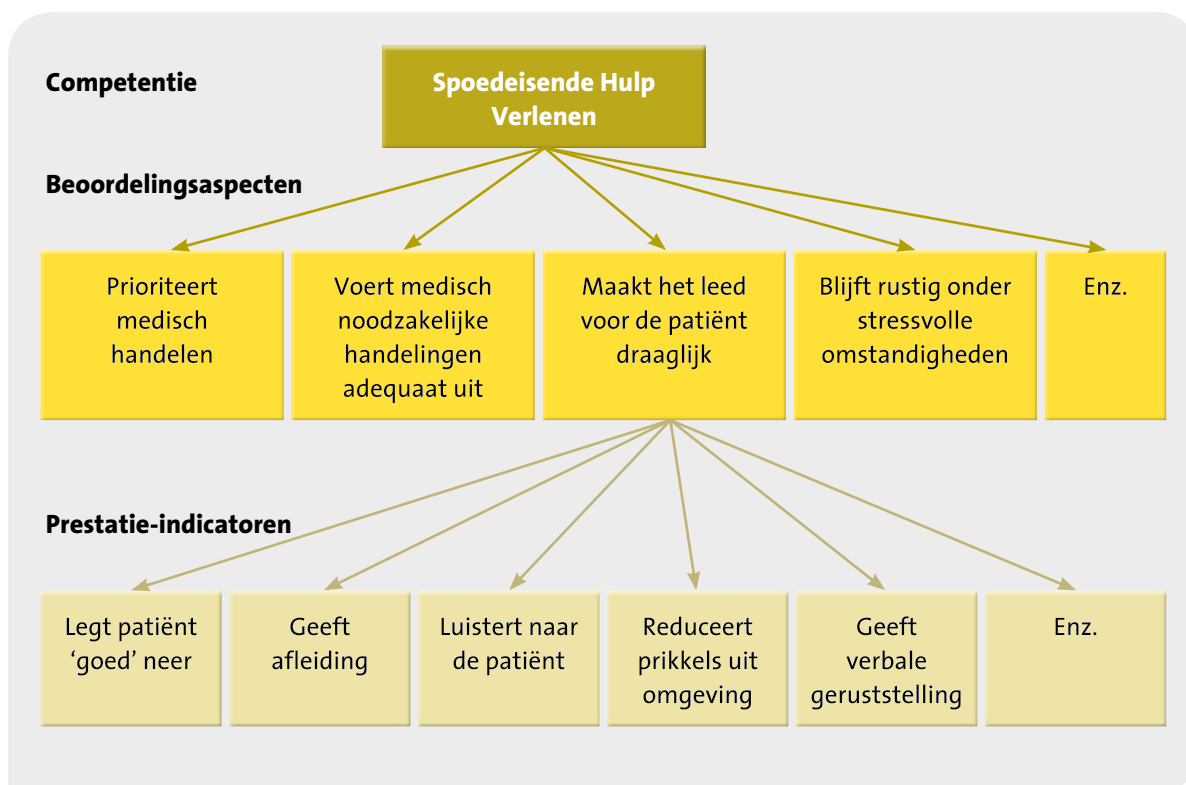
Beoordelingsaspecten zijn de kwaliteiten die in algemene zin kenmerkend zijn voor het handelen van de competente persoon bij het uitvoeren van een kerntaak en/of voor de producten die deze persoon in dat kader vervaardigt. Het formuleren van beoordelingsaspecten is overigens niet alleen van belang voor beoordelingsdoeleinden maar ook voor de opzet van het curriculum. Het gaat immers om een concretisering van de centrale opleidingsdoelstellingen van een competentiegerichte opleiding. Een dergelijke cruciale activiteit hoort een van de eerste stappen te zijn in de ontwikkeling van het curriculum en dient te worden uitgevoerd door een groep van inhoudelijke experts in plaats van door een individuele docent die door het management is belast met toetsontwikkeling.

Prestatie-indicatoren

Prestatie-indicatoren geven aan waar in de taakuitvoering of in het resulterend product aanwijzingen gevonden kunnen worden voor de gewenste kwaliteit die door een bepaald beoordelingsaspect beschreven is. Prestatie-indicatoren worden geformuleerd door een toetsconstructeur die zich daarbij vanuit twee invalshoeken laat inspireren, vanuit de door de curriculumontwikkelingscommissie geformuleerde beoordelingsaspecten en vanuit de assessmentopdracht die hij zelf ontwikkeld heeft:

- De afstemming op de beoordelingsaspecten heeft tot doel dat alleen die elementen beoordeeld worden die werkelijk van belang zijn voor de beheersing van de betreffende competentie.
- De afstemming van de prestatie-indicatoren op de assessmentopdracht biedt twee voordelen. In de eerste plaats voor de kandidaat, die de scores direct in verband kan brengen met goede en minder goede elementen in de taakuitvoering en/of een eventueel vervaardigd product en daarmee handvatten heeft om gericht te werken aan verbetering. In de tweede plaats voor de examiner, die in mindere mate een interpretatie hoeft te geven aan de formuleringen en daardoor een nauwkeuriger oordeel kan geven.

Figuur 3.4 geeft de operationalisatie van een competentie aan de hand van generieke beoordelingsaspecten en taakspecifieke prestatie-indicatoren schematisch weer.



Figuur 3.4: Operationalisatie van de competentie *Spoedeisende Hulp Verlenen* aan de hand van generieke beoordelingsaspecten en taakspecifieke prestatie-indicatoren

Beoordelingsschalen

Het beheersingsniveau van een competentie wordt generiek beschreven aan de hand van beoordelingsaspecten en taakspecifiek uitgewerkt in prestatie-indicatoren. De examiner gebruikt de prestatie-indicatoren om de kwaliteit van het handelen en een eventueel daaruit voortkomend product te beoordelen. Dit gebeurt door een positie op een beoordelingsschaal te markeren. Een belangrijke beslissing bij het ontwikkelen van een beoordelingsschaal betreft het

aantal posities op de scoreschaal. Bij het bepalen hiervan is het goed om zich te laten leiden door de vraag hoeveel prestatieniveaus op betrouwbare wijze onderscheiden kunnen worden. Drie is het minimum, zeven zal voor de meeste toepassingen al te veel zijn. Een ander belangrijk punt gaat over de wenselijkheid van een middenpositie. Als we examinatoren willen dwingen om zich uit te spreken over het al dan niet voldoende zijn van een gedrags- of productkenmerk, dan moet in principe voor een even aantal schaalpunten gekozen worden. Hofstee (1999) noemt dit 'het aannemen van een harde lijn jegens de beoordelaar door die geen vluchtgedrag (in de middenpositie) toe te staan'. Het is echter de vraag of het verstandig is om de examiner te dwingen tot een keuze voor een voldoende of onvoldoende oordeel. Juist in gevallen dat een examiner ernstige twijfels heeft, biedt een neutrale positie op de scoreschaal de mogelijkheid om de meetfout te beperken.

Het lastigste onderdeel van de constructie van een beoordelingsschaal betreft de verwoording van de prestatieniveaus bij de schaalpunten. Steeds opnieuw blijkt het moeilijk om in korte bewoordingen de te onderscheiden prestatieniveaus zo duidelijk te omschrijven dat examinatoren onafhankelijk van elkaar tot dezelfde scores komen bij een beoordeling van dezelfde kandidaten. Als het ook na veel inspanning niet wil lukken om de beschrijvingen helder te krijgen, is het de vraag of het wel zinvol is om de prestaties op het gewenste aantal niveaus te onderscheiden. Indien die vraag bevestigend beantwoord wordt, kan men er eventueel voor kiezen om niet alle schaalpunten te verwoorden. Bij een vijfpuntsschaal zou dit bijvoorbeeld inhouden dat alleen de middelste en de extreme schaalpunten verwoord worden.

Lezers die behoefte hebben aan meer gedetailleerde aanwijzingen over de constructie van scor rubrieken (prestatiecriteria en beoordelingsschalen) verwijzen we naar hoofdstuk 4 van dit boek en naar een toegankelijk geschreven artikel op de website Toetswijzer: 'Beoordelingsschalen in praktijktoetsen: hoe ontwikkel en gebruik je ze?', te raadplegen via <http://www.toetswijzer.nl/berichten/toetsspecial-praktijktoetsen-en-praktische-opdrachten>.

Prestatie-indicatoren zijn vaak taakspecifiek geformuleerd waardoor de kans groot is dat verschillende prestaties, bijvoorbeeld het handelen in taaksituatie I en in taaksituatie II, met (deels) verschillende prestatie-indicatoren gescoord worden. Een directe vergelijking van de behaalde prestatieniveaus in taaksituatie I en II is daarom pas mogelijk na een omzetting van de toegekende scores op de prestatie-indicatoren naar een score op het bovenliggende beoordelingsaspect. In tabel 3.2 wordt deze omzetting geïllustreerd aan de hand van een voorbeeld. De tabel laat zien dat de prestatie van een kandidaat in taaksituatie I met vier van de zeven prestatie-indicatoren (A.1 - A.7) gescoord is en dat deze vier scores via een transformatieregel zijn samengevat in één score op het bovenliggende beoordelingsaspect A. Op deze wijze zijn prestaties die met (deels) verschillende prestatie-indicatoren beoordeeld zijn, toch vergelijkbaar te maken. In de ToetsSpecial op de website Toetswijzer, <http://www.toetswijzer.nl/berichten/toetsspecial-toetsen-op-school-hoger-onderwijs>, is een Excel-rekenblad beschikbaar dat scores transformeert van een schaal met x schaalpunten ($3 \leq x \leq 7$) naar een schaal met y schaalpunten ($3 \leq y \leq 7$) bij een aantal van z gescoorde prestatie-indicatoren ($1 \leq z \leq 12$).

Tabel 3.2: Omzetting van gescoorde prestatie-indicatoren naar een score op het bovenliggende beoordelingsaspect

	Prestatie-indicatoren							Beoordelingsaspect
	A.1	A.2	A.3	A.4	A.5	A.6	A.7	A
Taaksituatie I	4		3	3			1	3
Taaksituatie II		3	1		3			2

3.3.5 Transparante beslissingsprocedure

Toen met de invoering van CGO veel opleidingen overschakelden op een didactiek van zelfverantwoordelijk leren, werd ook gezocht naar een nieuw, daarbij passend beoordelingsconcept. Veel opleidingen omarmden het portfolio. Helaas vond de invoering niet altijd op doordachte wijze plaats. Met name werd er onvoldoende onderscheid gemaakt tussen de verschillende soorten portfolio's. Een beoordelingsportfolio, een dossier met prestaties waarmee de eigenaar zijn competentie aantoonst, stelt andere (zwaardere) eisen aan de daarin opgenomen bewijzen dan een showcase portfolio. Laatstgenoemd portfolio wordt gevuld met het allerbeste werk van de eigenaar, met als doel een toekomstig opdrachtgever te overtuigen van de kwaliteiten van de bezitter. Dit type portfolio dient vooral een commercieel doel. Voor onderwijsdoelen en met name voor beslissingen over de verwerving van competentie, zijn ze niet erg geschikt. Vanwege de grote vrijheid die de bezitter heeft om het showcase portfolio naar eigen inzicht te vullen, moeten dergelijke bewijsdossiers met argwaan bekeken worden. Stel bijvoorbeeld dat een student tien prestaties heeft geleverd om zijn competentie aan te tonen en dat drie ervan van onvoldoende kwaliteit waren, dan zullen laatstgenoemde prestaties uiteraard niet in het portfolio worden opgenomen, als het aan de student ligt. Het zal echter duidelijk zijn dat een dergelijke selectie een vertekend beeld kan geven van de competentie van deze student. De toepassing van het portfolioconcept in het onderwijs kan daarom niet zonder heldere regels, die:

- duidelijk maken wanneer prestaties van studenten kunnen worden geaccepteerd als (deel van het) bewijs voor het verworven zijn van een competentie;
- aangeven volgens welke procedure een beslissing wordt genomen op grond van de verzamelde bewijzen (zie paragraaf 3.3.5).

Duidelijkheid over wanneer prestaties van studenten kunnen worden geaccepteerd als (deel van het) bewijs voor het verworven zijn van een competentie betekent dat de opleiding regels zal moeten opstellen die de authenticiteit, actualiteit, generaliseerbaarheid en extrapolatiebaarheid van de bewijzen waarborgen. In het geval van 'authenticiteit' gaan die regels over voorzorgsmaatregelen die genomen worden om te voorkomen dat studenten bewijs opnemen dat niet het resultaat van eigen inspanning is. Maar ook over het voorkomen van 'meeliftgedrag' bij groepswork. Bij 'actualiteit' gaat het om regels die voorkomen dat iemand zijn diploma krijgt op basis van verouderde prestaties, die mogelijk niet meer representatief zijn voor de huidige bekwaamheid en/of die beoordeeld zijn op basis van prestatiecriteria die inmiddels achterhaald zijn. De regels ten aanzien van 'generaliseerbaarheid' moeten ervoor zorgen dat het vastgestelde prestatieniveau niet taakspecifiek is. En de regels met betrekking tot 'extrapolatiebaarheid' dienen ervoor om te voorkomen dat iemand competent wordt verklaard op grond van prestaties die een magere afspiegeling zijn van het handelen in de reële werksituatie. Kort gezegd komt het erop neer dat opleidingen veel meer de regie gaan voeren bij het aanleggen van bewijsdossiers; iets wat tot voor kort altijd strijdig werd geacht met de portfoliogedachte.

3.4 Een assessmentprogramma leidt tot een verantwoord bewijsdossier

Uitvoering van een assessmentprogramma met de hiervoor besproken kenmerken leidt ertoe dat een student een bewijsdossier opbouwt voor elke (beroeps)competentie die in een opleiding verworven dient te worden. Figuur 3.5 laat zien welke informatie minimaal in zo'n bewijsdossier wordt opgenomen.

Beoordelingsaspecten													
Prestatie	Taak-situatie	Assessment-methode	Datum	Assessor	A	B	C	D	E	F	Totaal-score	Horizontale standaard	Resultaat
1	III	Hands-off, SJT	260511	n.v.t.	4						4	4,8	–
2	II	LOTUS-patiënt	130611	J. Pav.	5	5		5			15	14,3	+
3	IX	Hands-on, afd SEH	200711	H. Bie.	5	5		5	4	6	25	23,7	+
4	XXIV	Hands-on, afd SEH	060911	J. Pav.	5		4	4	5	6	24	23,3	+
5	VII	LOTUS-patiënt rampenoefening	140612	A. v. Ge.	5	4	5	6	5		25	23,2	+
Gemiddelde score per aspect					4,8	4,7	4,5	5,0	4,7	6,0			
Verticale standaard					4,8	4,7	4,3	4,8	4,6	4,8			
Resultaat					+	+	+	+	+	+			

Figuur 3.5: Bewijsdossier als resultaat van de uitvoering van een assessmentprogramma

Het bewijsdossier in figuur 3.5 is van een fictieve student (basisarts) die een specialistische opleiding volgt om als SEH-arts aan het werk te gaan op de afdeling spoedeisende hulp (SEH) van een ziekenhuis. De opleiders hebben de competentie SEH geoperationaliseerd door een beschrijving van een groot aantal taaksituaties (in de figuur slechts aangeduid met Romeinse cijfers) waarmee de beginnende SEH-arts geconfronteerd kan worden en verder door de globale maar essentiële kwaliteiten (beoordelingsaspecten) te benoemen waarin het handelen van de competente SEH-arts zich onderscheidt van zijn nog niet-competente collega.

We lichten het bewijsdossier toe aan de hand van de in paragraaf 3.3 besproken kenmerken:

- **Meervoudig bewijs** | In de eerste plaats valt op dat de kandidaat meer dan één prestatie heeft geleverd om zijn competentie aan te tonen. Ook is duidelijk dat het om steeds andere taaksituaties gaat. Dit vergroot de kans dat de uitgevoerde assessmenttaken samen een voldoende afspiegeling vormen van de taaksituaties waarmee de beginnend SEH-arts geconfronteerd kan worden.
- **Methodemix** | We zien verder dat de prestaties aan de hand van verschillende assessment-methodes zijn uitgelokt en geregistreerd. Met 'geregistreerd' bedoelen we dat er maatregelen zijn genomen om bepaalde taaksituaties te creëren waarin de kandidaat zijn competentie moet demonstreren. Bij observatie op de werkvloer zijn de mogelijkheden om te registreren zeer gering, bij hands-off toetsen is nagenoeg alles (taken, situaties, hulpmiddelen, wijze van responderen) geregistreerd. De regie die de opleiding samen met de student voert over het aanleggen van een bewijsdossier moet ertoe leiden dat op zo efficiënt mogelijke wijze zo veel mogelijk informatie wordt verkregen over alle beoordelingsaspecten van de betreffende competentie.
- **Meerdere examinatoren** | Er zijn verschillende examinatoren ingezet om de prestaties van de kandidaat te beoordelen. In principe heeft het de voorkeur om elke prestatie door minimaal twee onafhankelijk van elkaar werkende examinatoren te laten beoordelen en hun gemiddelde scores te beschouwen als de best haalbare beoordeling van de prestatie. De realiteit is echter dat er meestal onvoldoende middelen zijn voor herhaalde beoordeling van prestaties. Een compromis is dat zo veel mogelijk verschillende examinatoren worden ingezet, waarmee in

ieder geval voorkomen wordt dat een kandidaat systematisch benadeeld of bevoordeeld wordt door de subjectiviteit in het oordeel van een specifieke examinerator.

- *Standaardset van prestatiecriteria* | Ondanks het feit dat de kandidaat handelt in verschillende taaksituaties en er verschillende assessmentmethodes zijn ingezet om dit handelen uit te lokken en te registreren, worden het gedrag en de resultaten daarvan steeds met dezelfde beoordelingsaspecten beoordeeld. Daarbij worden niet noodzakelijk alle beoordelingsaspecten gebruikt. De specifieke assessmenttaak, de omstandigheden waaronder deze taak uitgevoerd wordt en de assessmentmethode die gebruikt wordt om de verlangde kwaliteit zichtbaar te maken, bepalen welke prestatie-indicatoren en dus welke beoordelingsaspecten gescoord kunnen worden.
- *Transparante procedures* | De scores in het bewijsdossier worden volgens heldere, goed beschreven procedures verwerkt tot beslissingen die verantwoord kunnen worden naar alle relevante betrokkenen. Een cruciale vraag bij de inzet van een bewijsdossier is wanneer het verzamelde bewijs de conclusie rechtvaardigt dat de betreffende competentie verworven is. Om willekeur bij beslissingen over studenten te voorkomen, is het noodzakelijk dat de beslissingsprocedure voor alle betrokkenen transparant is. In het geval van het bewijsdossier betekent dit dat duidelijk is hoe de scores geaggregeerd en geëvalueerd worden. Figuur 3.5 laat zien dat er sprake is van een verticale en horizontale evaluatie van de verzamelde scores. Bij de verticale evaluatie wordt gekeken of het gemiddelde prestatieniveau op elk beoordelingsaspect afzonderlijk voldoet aan de verticale standaard. De horizontale evaluatie geeft aan of de kandidaat in een bepaalde taaksituatie adequaat heeft gehandeld. Hiervan is sprake als de score groter of gelijk is aan de horizontale standaard. Zowel de verticale als de horizontale standaard worden ontleend aan de zogeheten grenswaarden. Dit zijn de kritische waarden op de schaal van de beoordelingsaspecten waarover een groep inhoudelijk deskundigen consensus heeft bereikt. Een grenswaarde markeert het punt op de schaal van een beoordelingsaspect dat competente van incompetent kandidaten scheidt. Verticale standaarden gelden per beoordelingsaspect en zijn daarom identiek aan de grenswaarde van een beoordelingsaspect. Horizontale standaarden gelden per prestatie en zijn gelijk aan de som van de grenswaarden van alle bij de beoordeling van de betreffende prestatie gebruikte beoordelingsaspecten. Het is aan de opleiding om te bepalen welke combinatie van horizontale en verticale evaluatieresultaten moet leiden tot de conclusie dat de betreffende competentie verworven is.

In de ToetsSpecial op de website Toetswijzer, <http://www.toetswijzer.nl/berichten/toetsspecial-toetsen-op-school-hoger-onderwijs>, biedt een Excel-spreadsheet administratieve ondersteuning bij de beoordeling van competenties op basis van assessmentprogramma's. De spreadsheet is niet bedoeld voor serieuze administratieve toepassing, maar uitsluitend om in praktische zin kennis te nemen van de principes van assessmentprogramma's en de aanleg van bewijsdossiers. Bij de spreadsheet is een handleiding beschikbaar, ook te vinden in de ToetsSpecial op de website Toetswijzer, <http://www.toetswijzer.nl/berichten/toetsspecial-toetsen-op-school-hoger-onderwijs>.

3.5 Ten slotte

Vaststellen of studenten de opleidingsdoelen bereikt hebben, is er met de invoering van competentiegericht opleiden niet makkelijker op geworden. Duidelijk mag zijn dat een reeks voldoende gemaakte kennistoetsen en een als voldoende beoordeelde eindschrijving niet langer volstaan om een diploma af te geven. Er hoort een deugdelijk bewijs te worden overgelegd waaruit blijkt dat de (beroeps)competenties verworven zijn. Dit hoofdstuk heeft hopelijk duidelijk gemaakt dat een dergelijk bewijs alleen binnen bereik ligt als de opleiding weer de regie van de bewijsvoering in handen neemt en zo ervoor zorgt dat er betrouwbare, valide en transparante beslissingen worden genomen over het al dan niet toekennen van diploma's.

4 Het beoordelen van afstudeeropdrachten

4 Het beoordelen van afstudeeropdrachten

*Brenda Zandsteeg, Marieke van der Schaaf,
Casper Hulshof en Frans Prins*

Bij dit hoofdstuk over het beoordelen van afstudeeropdrachten plaatsen we vooraf de kanttekening dat beoordelen nooit losstaat van het leren. Beoordelen is meer is dan het afsluiten van een opleiding. Een opleiding vereist samenhang tussen leren en beoordelen. Beoordeling is één van de componenten in een leeromgeving naast de nagestreefde leerdoelen, instructie en begeleiding, en opdrachten. Dit vraagt van docenten dat zij helder zijn over de doelen die ze nastreven en dat zij de afzonderlijke componenten goed op elkaar laten aansluiten (Biggs, 1996). Er zijn aanwijzingen dat onderwijs waarin het onderwijsprogramma en de beoordelingsmethode consistent zijn met elkaar, leidt tot betere leerresultaten (Cohen, 1987; Porter, Polikoff, Barghaus & Yang, 2013).

Dit hoofdstuk start met een schets van de context van afstudeeropdrachten in het hoger onderwijs: wat zijn kenmerken van afstudeeropdrachten, wat wordt beoordeeld en wat zijn dilemma's daarbij. We staan stil bij de verschillende functies van beoordelen en de eisen waaraan het beoordelen van afstudeeropdrachten moet voldoen. We besteden aandacht aan het formuleren van criteria en standaarden en aan de beschikbare beoordelingsmodellen. Aan het eind gaan we in op individueel beoordelen of groepsbeoordeling, het gebruik van rubrics, peer beoordeling, zelfbeoordeling, de vermenging van beoordelen en begeleiden en tot slot portfoliobeoordeling.

4.1 Onderzoeken en professioneel handelen

De afstudeerfase in het hoger beroepsonderwijs (hbo) en wetenschappelijk onderwijs (wo) bestaat gewoonlijk uit het succesvol afronden van een aantal verschillende projecten. Het product van afstuderen aan een hogeschool of universiteit is in zekere zin de kroon op het werk. Het is het moment om de opleiding in de praktijk te brengen, dan wel in wetenschappelijk onderzoek dan wel in de beroepspraktijk.

In de afstudeerfase in het hbo staat het beoordelen van beroepsmatig handelen centraal. Onderwerp van beoordeling is zowel het beroepsproduct zelf als de legitimering ervan vanuit onderzoek. In het kader van innovatief werken hebben onderzoekend werken, werken op basis van de best beschikbare informatie (evidence based werken) en uitvoeren van onderzoek in de afstudeerfase van hbo-opleidingen steeds meer aandacht gekregen. Het betreft veelal praktijkonderzoek en soms praktijkgericht wetenschappelijk onderzoek. De te beoordelen afstudeerproducten in het hbo zijn dus veelal beroepsproducten, presentaties, reflectiegesprekken en soms onderzoeksverslagen. Beroepsproducten hebben verschillende vormen, afhankelijk van de discipline. Denk aan beeldende kunst, zorg en welzijn, bouwkunde of horeca. Afstudeeropdrachten zijn vaak groepsopdrachten. Vergeleken met andere landen legt het Nederlandse hoger beroepsonderwijs relatief veel nadruk op groepsopdrachten en stages (HBO-Monitor 2006, Vereniging Hogescholen).

In de afstudeerfase in het wo is het zelfstandig voorbereiden, uitvoeren en rapporteren van een wetenschappelijk onderzoek de meest voorkomende eindopdracht. Dit wordt beoordeeld aan de

hand van een onderzoeksverslag, al dan niet in de vorm van een wetenschappelijk artikel aangevuld met een presentatie of verdediging. Daarnaast is professioneel handelen in het toekomstig werkdomein een aspect van de eindbeoordeling in het wo. De beoordeling hiervan vindt plaats in de stage, door het beoordelen van de legitimering van het professioneel handelen en beslisgedrag en ook de kwaliteit van het stageproduct. Niet alle wo-opleidingen kennen een verplichte stage. De te beoordelen producten zijn dus in de eerste plaats onderzoeksverslagen, presentaties van onderzoek (verdediging) en stageverslagen (legitimeringsverslagen). Afstudeeropdrachten worden in het wo vrijwel altijd individueel uitgevoerd en beoordeeld. Samengevat kunnen we zeggen dat in de afstudeerfase in het hoger onderwijs zowel professioneel handelen als het doen van onderzoek worden beoordeeld.

4.2 Kenmerken van afstudeeropdrachten

Afstudeeropdrachten vragen door hun open vorm om een andere wijze van beoordelen dan traditionele toetsen (Stokking & Van der Schaaf, 1999; Van der Schaaf, Goris & Stokking, 2000). Kenmerken van de open vorm van afstudeeropdrachten zijn:

- Openheid en authenticiteit
- Complexe taken
- Langere tijdsperiode
- Integratie van beoordeling met instructie en begeleiding
- Beoordeling van product en proces
- Diversiteit

Openheid en authenticiteit

Een afstudeeropdracht kan op meerdere manieren goed gemaakt worden. Studenten doen een echt onderzoek of voeren een praktijkopdracht uit. De afstudeeropdracht wordt in de beroepspraktijk zelf uitgevoerd. Studenten moeten dus samenwerken met collega's. Dit biedt hen ook de mogelijkheid om feedback en begeleiding te vragen.

Complexe taken

Een afstudeeropdracht doet een beroep op zowel kennis als verschillende vaardigheden. Studenten moeten onder meer voldoende vakinhoudelijke kennis inzetten, de opdracht uitvoeren, de planning in de gaten houden en bij voorkomende obstakels oplossingen bedenken om bij te sturen.

Langere tijdsperiode

De afstudeeropdracht wordt uitgevoerd over een langere tijdsperiode.

Integratie van beoordeling met instructie en begeleiding

In plaats van instructie en begeleiding te scheiden van beoordeling, bijvoorbeeld door na het geven van een les studenten een toets te laten maken, is bij afstudeeropdrachten sprake van integratie van beide. Dit is bijvoorbeeld aan de orde bij tussentijdse beoordelingen of feedbackmomenten. Soms krijgen studenten halverwege hun afstudeeropdracht een beoordeling die meetelt voor het eindcijfer. Ze kunnen dan profiteren van de feedback die ze bij een dergelijke beoordeling krijgen. Ook de feedback van de docent in begeleidingsmomenten draagt eraan bij dat studenten hun proces en product nog kunnen bijstellen voordat het eindoordeel wordt gegeven.

Beoordeling van product en proces

Het beoordelen van afstudeeropdrachten kan betrekking hebben op het product en het proces. Bij het beoordelen van het product wordt een uitspraak gedaan over de kwaliteit van het

beroepsproduct of het onderzoeksverslag. Het is echter de vraag of deze uitspraak voldoende representatief is voor het competentieniveau van de student. Het valt soms te beargumenteren om ook het proces mee te nemen in de beoordelingscriteria. Denk daarbij aan de hoeveelheid begeleiding, het werken volgens een planning, de nauwkeurigheid van werken, de hoeveelheid geïnvesteerde tijd, de werkhouding of inzet van de student, de complexiteit van de opdracht, de individuele bijdrage aan groepswork of de samenwerking van de opdrachtgroep. De beoordeling van het proces kan plaatsvinden via een reflectieverslag, peer feedback, een criteriumgericht interview of door het oordeel van de praktijkbegeleider mee te laten wegen.

Diversiteit

Afstudeeropdrachten in het hoger onderwijs laten een grote diversiteit zien. Denk daarbij aan de complexiteit van de opdracht, de context waarin de opdracht wordt uitgevoerd en de mate van zelfstandigheid die wordt verwacht van de student. Zo kan het voorkomen dat er expliciet een maximale tijdsperiode gesteld is voor de uitvoering van het afstuderen, dat er strikte fasen zijn met op te leveren tussenproducten en dat er een maximum aan begeleidingstijd is gedefinieerd. Maar dat is niet altijd het geval. Het kan ook zijn dat de student zelf een onderwerp moet bedenken en daarbij een begeleider of opdrachtgever zoekt, of dat de student participeert in een lopend project.

Niet alleen tussen opleidingen maar ook binnen opleidingen zijn soms verschillen aanwezig. De beoordelingsinstrumenten en de daarin gebruikte criteria kunnen verschillen in de mate van specificiteit en uitgebreidheid. Ook de inhoud van de criteria, bijvoorbeeld de mate waarin taal en stijl meewegen bij de beoordeling, kan verschillen. De diversiteit in de aard van de opdracht en de beoordelingswijze kan problematisch zijn als het doel is om een uitspraak te doen over het competentieniveau van de student. Het vergelijken van afstudeerbeoordelingen binnen en tussen opleidingen wordt hierdoor niet eenvoudiger en het bewaken van het eindniveau wordt hierdoor bemoeilijkt.

Om tot een betere vergelijkingsmogelijkheid tussen eindbeoordelingen te komen, kunnen oplossingen worden gezocht in de afstemming tussen opleidingen over de bandbreedte waarbinnen afstudeeropdrachten zich kunnen bevinden. Daarbij worden eisen aan de afstudeeropdracht gesteld met betrekking tot bijvoorbeeld de complexiteit van de opdracht, de context waarin de opdracht moet worden uitgevoerd en de mate van zelfstandigheid. Omdat deze criteria altijd een zekere mate van subjectiviteit hebben, is daarnaast ook afstemming over de beoordeling wenselijk. Benchmarks van het eindniveau, bijvoorbeeld door het steekproefsgewijs onderling beoordelen van afstudeerproducten en hierover consensus te bereiken, kan bijdragen aan de uniformiteit van het eindniveau van opleidingen. Het mag duidelijk zijn dat dit tijd en geld kost en dat competitie tussen opleidingen hieraan niet bijdraagt.

4.3 Dilemma's bij de beoordeling van afstudeeropdrachten

Afstudeeropdrachten vragen dus, door hun open vorm, om een andere manier van beoordelen dan traditionele toetsen. Daarnaast is er een aantal dilemma's bij het beoordelen van afstudeeropdrachten, zoals:

- het formuleren van beoordelingscriteria;
- het vaststellen van standaarden (cesuur);
- het kiezen van een beoordelingsmodel;
- individueel beoordelen of groepsbeoordeling en het betrekken van het proces bij de beoordeling.

Alvorens we deze dilemma's bespreken (paragraaf 4.6 t/m 4.9), gaan we eerst in op de functies van beoordelen en de eisen waaraan de beoordeling van afstudeeropdrachten moet voldoen.

4.4 Functies van beoordelen bij afstudeeropdrachten

Bij afstudeeropdrachten kan naast summatieve beoordeling ook sprake zijn van formatieve beoordeling. We spreken van formatief beoordelen wanneer een student tussentijds feedback of verbeteruggesties krijgt. Bij summatieve beoordeling wordt de student op de uiteindelijke kwaliteit van het afstudeerwerkstuk afgerekend met bepaalde consequenties zoals het wel of niet slagen voor een vak. Bij afstudeeropdrachten is meestal sprake van beide functies. Tijdens het werken aan de afstudeeropdracht is een student gebaat bij formatieve beoordelingen, terwijl onderdelen of het totale eindwerkstuk summatief worden beoordeeld (Shavelson et al., 1993; Straetmans, 1998).

4.5 Eisen waaraan de beoordeling van afstudeeropdrachten moet voldoen

Beoordelingen van praktijkopdrachten moeten voldoen aan de volgende eisen (Stokking & Van der Schaaf, 1999; Stokking et al., 2004):

- Betrouwbaarheid
- Validiteit
- Praktische bruikbaarheid
- Aanvaardbaarheid

Betrouwbaarheid

Als kandidaten direct na elkaar twee vergelijkbare examens zouden afleggen dan zou het zo moeten zijn dat de kandidaten die een hoge score op het eerste examen behalen ook een hoge score op het tweede examen behalen, en dat de kandidaten die een lage score behalen ook een lage score op het tweede examen behalen. Is dat niet het geval, dan kunnen we geen vertrouwen hebben in de betrouwbaarheid van het examen. Voor meer informatie over betrouwbaarheid verwijzen we naar hoofdstuk 3 van Toetsen op School, www.toetsenopschool.nl.

Validiteit

De beoordeling moet betrekking hebben op dié kennis, vaardigheden en attitudes die de opleiding met het afstudeerproduct wil meten. Ook de aansluiting van de moeilijkheidsgraad op het beoogde eindniveau is hierbij van belang. Zie voor meer informatie over validiteit hoofdstuk 4 van Toetsen op School, www.toetsenopschool.nl.

Praktische bruikbaarheid

Docenten moeten de beoordeling kunnen uitvoeren (qua bekwaamheid en beschikbare tijd) en de beoordeling moet:

- inzicht geven in hoever de studenten zijn in hun ontwikkeling van onderzoekscompetenties en wat er nog aan schort;
- geschikt zijn voor het meetellen bij de vaststelling van cijfers;
- geschikt zijn voor de communicatie met studenten.

Aanvaardbaarheid

De beoordeling moet aanvaardbaar zijn, dat wil zeggen:

- de beoordelingsprocedure en de beoordelingscriteria moeten duidelijk en inzichtelijk zijn en kunnen worden uitgelegd aan anderen;
- de procedure en criteria moeten objectief en tegenover anderen te verdedigen zijn;
- studenten moeten weten waar ze aan toe zijn;
- studenten moeten gelijkwaardig worden behandeld en gelijkwaardige kansen krijgen (o.a. qua beschikbare tijd, hulpmiddelen en begeleiding).

4.6 Formuleren van beoordelingscriteria

Competenties die in een afstudeeropdracht worden beoordeeld bestaan uit geïntegreerde kennis, vaardigheden en attitudes die nodig zijn om de opdracht uit te voeren. We kunnen deze competenties niet direct waarnemen, maar ze afleiden van de voortgang in het afstudeerproces en tussentijdse producten, zoals een conceptproduct, onderzoeksplan, tussentijdse reflectie, presentatie, etc. Bij deze onderdelen kunnen dan ook criteria en standaarden worden geformuleerd die de student houvast kunnen bieden bij het uitvoeren van de opdracht en die voor de begeleider als leidraad kunnen dienen voor het geven van feedback en beoordelingen. Criteria geven aan waar het afstudeerwerk aan moet voldoen. De standaarden beschrijven hoe goed aan de criteria moet worden voldaan voor een bepaalde beoordeling (bijvoorbeeld voldoende/onvoldoende). Omwille van de transparantie van beoordelingen en om het proces van beoordelen te structureren, is het gewenst om de criteria en bijbehorende standaarden te expliciteren. Zie voor meer informatie ook hoofdstuk 3.

Voor het formuleren van criteria en standaarden kunnen we op verschillende manieren te werk gaan. Grofweg kan een onderscheid worden gemaakt tussen constructgericht of taakgericht beoordelen.

Bij constructgericht beoordelen gaat het om het beoordelen van kennis, vaardigheden, competenties of constructen. Denk daarbij aan bijvoorbeeld schrijfvaardigheid, ondernemerschap of kennis van onderzoeksmethoden.

Bij de taakgerichte benadering richt de beoordeling zich op de stappen van de opdracht. Bij het doen van praktijkonderzoek kunnen bijvoorbeeld de volgende stappen worden benoemd (Stokking & Van der Schaaf, 2000):

- 1 Formuleren van het probleem (hanteren van de vakspecifieke begrippen; herkennen en formuleren van een probleem in vaktermen)
- 2 Formuleren van de onderzoeksvraag met eventuele deelvragen en hypothesen
- 3 Maken en bewaken van het onderzoeksplan bestaande uit een onderzoeksopzet en een werkplanning
- 4 Verzamelen en selecteren van gegevens
- 5 Verwerken van de gegevens
- 6 Analyseren van de gegevens
- 7 Beschrijven van de resultaten
- 8 Interpreteren van de resultaten in conclusies
- 9 Bediscussiëren en evalueren van het onderzoek
- 10 Rapporteren van het onderzoek: schriftelijk (verslag of ontwerp) en/of mondeling (presentatie)

We verwijzen hier ook naar de publicatie van Andriessen (2012) over kwaliteitscriteria voor afstudeeronderzoek voor hbo-bacheloropleidingen. De publicatie bestaat uit 8 delen. Deel 1 gaat in op de ontvankelijkheidseisen die aan het afstudeerwerkstuk gesteld kunnen worden. Hierbij moeten we denken aan criteria zoals de volledigheid van het onderzoeksverslag, de logische opbouw van het verslag en de uiterlijke verzorging. Indicatoren voor het laatste criterium betreffen spelling en interpunctie, redactionele vormgeving, literatuurlijst en bronvermelding. Indien aan één of meerdere ontvankelijkheidseisen niet voldaan wordt, wordt de scriptie niet in behandeling genomen. Deel 2 tot en met 8 beschrijven kwaliteitscriteria voor het afstudeerwerkstuk, geclusterd naar de zeven fasen in het onderzoeksproces:

- 1 Maken van doelstelling, probleemstelling en deelvragen (de bijbehorende kwaliteitscriteria betreffen: inhoudelijke verankering, relevantie, afbakening, precisie en functionaliteit)
- 2 Ontwikkelen van methodologie
- 3 Uitvoeren van literatuuronderzoek en ontwikkelen van theoretisch en conceptueel kader
- 4 Uitvoeren van onderzoek

- 5 Trekken van conclusies
- 6 Maken van eventuele aanbevelingen en/of beroepsproducten
- 7 Terugblik/reflectie over het onderzoekstraject

In de ToetsSpecial op de website Toetswijzer, <http://www.toetswijzer.nl/berichten/toetsspecial-toetsen-op-school-hoger-onderwijs>, is een voorbeeld te vinden van onderzoeksvaardigheden voor rechtenstudenten. Het voorbeeld beschrijft de zes fasen in het juridisch onderzoekproces (Curry-Sumner & Van der Schaaf, 2011).

4.7 Formuleren van beoordelingsstandaarden

Wanneer de criteria zijn geformuleerd waaraan afstudeeropdrachten moeten voldoen, is vervolgens de vraag aan hoeveel criteria studenten moeten voldoen voor het behalen van een voldoende beoordeling, oftewel de standaard of cesuur van een examen of afstudeeropdracht. Voor de ontwikkeling van standaarden zijn verschillende methoden beschikbaar (Berk, 1996). Deze methoden kunnen normgericht of criteriumgericht zijn.

Bij een normgerichte methode wordt de score van een student bij een afstudeerwerkstuk vergeleken met de scores die de andere studenten in de normgroep hebben behaald. Vervolgens wordt bepaald met welke scores studenten zullen slagen of zakken (Hambleton & Pitoniak, 2006). Bij een criteriumgerichte methode wordt het afstudeerwerkstuk vergeleken met criteria die vooraf zijn beschreven (absolute standaard). Criteriumgerichte methoden voor het stellen van standaarden kunnen meer toetsgecentreerd zijn of meer studentgecentreerd.

Bij de toetsgecentreerde methode is de standaard gebaseerd op de inschattingen van experts over fictieve voorbeelden van afstudeeropdrachten van studenten. De meest gebruikte toetsgecentreerde methode is de methode van Angoff. Daarbij maakt een groep experts een inschatting van de score op de criteria die minimaal nodig is om voor de (afstudeer)opdracht te slagen (Hambleton et al., 2000).

Bij een studentgecentreerde benadering worden standaarden vastgesteld op basis van daadwerkelijk ingeleverde afstudeeropdrachten door studenten. Een voorbeeld is de 'grensgeval methode' waarbij een panel van experts een veelheid van afstudeeropdrachten beoordeelt die zich net op de grens van voldoende of onvoldoende bevinden. Vervolgens wordt aan de hand van de mediaan van beoordelingen in deze groep bepaald waar de grens tussen een voldoende en onvoldoende komt te liggen (Hambleton & Pitoniak, 2006).

Voor meer informatie over het bepalen van standaarden en cesuur verwijzen we naar hoofdstuk 2 en hoofdstuk 9 van Toetsen op School, www.toetsenopschool.nl.

4.8 Beoordelingsmodellen voor het combineren van criteria

Doorgaans worden afstudeeropdrachten beoordeeld op meerdere (deel)criteria die de opdracht representeren, bijvoorbeeld: analyse, ontwerp en product; probleemstelling, dataverzameling en data-analyse. Voor de beoordeling moeten de scores op de criteria gecombineerd worden tot een eindoordeel. Twee modellen die hiervoor vaak worden gebruikt zijn een compenserend en een conjunctief beoordelingsmodel (zie ook hoofdstuk 2). Daarnaast is het mogelijk om een disjunctief model te hanteren. In het compenserende model kunnen deelscores op criteria elkaar compenseren. Het conjunctieve model vereist dat de student op elk criterium een bepaalde vastgestelde minimum score behaalt. In een disjunctief model zijn slechts voor bepaalde criteria minimale scores verplicht.

Een keuze voor een van deze modellen kan gebaseerd zijn op ten minste twee overwegingen. Ten eerste de opvatting over het doen van onderzoek. Vereist het doen van onderzoek competenties van studenten waarin ze sterke en minder sterke kanten kunnen hebben die elkaar kunnen

compenseren? Ten tweede, valt te verwachten dat de beoordelingscriteria sterk met elkaar verwant zijn of op elkaar voortbouwen, waardoor compensatie voor de hand ligt? Meestal wordt gekozen voor een compenserend model, omdat het reëel is te veronderstellen dat studenten zwakke en sterke kanten hebben en dat er nog ruimte is voor een verdere ontwikkeling.

4.9 Individueel beoordelen of groepsbeoordeling

Afstudeeropdrachten worden soms individueel uitgevoerd maar vaak ook in duo's of in groepen van drie of zelfs vier studenten. Het voordeel van het laatste is dat er door de studenten meer werk verzet kan worden en daardoor ook een grotere diepgang en betere kwaliteit kan worden bereikt. Werken in een groep bevordert ook het overleg tussen studenten. En groepswork sluit in veel gevallen goed aan bij de beroepspraktijk waar de student na het afstuderen in terecht komt, want in de beroepspraktijk wordt veel in (multidisciplinaire) teams gewerkt.

Groepswork heeft ook economische voordelen want de begeleiding van een groep van drie studenten kost minder tijd dan het begeleiden van drie studenten afzonderlijk. In veel instellingen in het hoger onderwijs is het individueel begeleiden van afstudeeropdrachten bijna niet meer te betalen. Dat komt voornamelijk door de enorme toename van het aantal studenten in het hoger onderwijs de afgelopen decennia. Het ministerie van OCW vermeldt een toename van 15% van de bevolking in de leeftijd tussen 18 en 25 jaar in 1975 naar bijna 40% in 2010 (bron: www.trendsinebeeldocw.nl). Het aantal medewerkers aan hogescholen en universiteiten is echter niet in dezelfde mate meegegroeid.

Er is echter ook een keerzijde van groepswork, met name de beoordeling van individuele studenten. Bij groepswork is het voor docenten vaak niet helder wie welk aandeel en welke kwaliteit heeft geleverd bij de totstandkoming van het afstudeerproduct. Als studenten een onevenredig klein aandeel leveren dan spreken we van meeliftgedrag. Dit betekent dat bij groepswork een student kan slagen voor het afstudeerwerk terwijl hij/zij niet de inspanning heeft geleverd en niet heeft laten zien dat hij/zij de te toetsen competenties bezit.

Er zijn verschillende manieren om meeliftgedrag te ondervangen:

- In de eerste plaats kunnen docenten onderdelen van de afstudeeropdracht individueel beoordelen of aan de afstudeeropdracht een individuele opdracht koppelen. Een goed voorbeeld daarvan is om bij een afstudeeronderzoek het onderzoeksplan en de dataverzameling door een groep te laten uitvoeren (waardoor het voordeel om veel werk te verzetten behouden blijft) en de resultaten- en/of de discussiesectie van een afstudeerscriptie individueel uit te laten voeren (en te beoordelen). Een ander voorbeeld is om studenten naast de groepsopdracht een individuele opdracht te geven, zoals het peer reviewen van een conceptproduct van een andere groep studenten. Deze opdracht past vaak goed bij kwalificaties die studenten moeten verwerven bij het afstuderen en het levert de medestudenten extra feedback op. Daarnaast is een review een redelijk compacte opdracht die de docenten niet veel extra nakijkwerk oplevert.
- Op de tweede plaats zou meeliftgedrag ondervangen kunnen worden als de docent tijdens de begeleiding zicht krijgt op het groeps- en samenwerkingsproces. Dat kan door de groep studenten in een gezamenlijke tussentijdse reflectie expliciet te laten rapporteren over wie welk aandeel in het afstudeerwerk heeft gehad of tijdens begeleidingsbijeenkomsten expliciet te vragen naar het aandeel van de groepsleden. Wat voorkomen moet worden, is dat pas aan het eind van het afstudeerwerk blijkt dat één van de groepsleden meeliftgedrag heeft vertoond, want dan is er weinig meer bij te sturen.
- Ten slotte kan de kwaliteit van de samenwerking, eventueel getoetst door een peer beoordeling, onderdeel zijn van de eindbeoordeling, maar natuurlijk alleen als het

samenwerken onderdeel is van de te verwerven kwalificaties tijdens het afstudeerwerk. In de ToetsSpecial op de website Toetswijzer, <http://www.toetswijzer.nl/berichten/toetsspecial-toetsen-op-school-hoger-onderwijs>, is een voorbeeld beschikbaar van een peer beoordelingsformulier voor het beoordelen van het samenwerkingsproces bij een afstudeerproject.

4.10 Gebruik van rubrics, peer beoordeling en zelfbeoordeling

Rubrics

Vandaag de dag worden beslissingen over afstudeeropdrachten ook vaak gebaseerd op rubrics. Rubrics bevatten criteria waarbij voor elk criterium ook de onderscheiden niveaus worden beschreven. Anders gezegd, rubrics beschrijven aan welke aspecten de kwaliteit van het werk moet voldoen voor een bepaald niveau, bijvoorbeeld voldoende, goed of excellent (Arter & Chappuis, 2007; Hafner & Hafner, 2003). De beoordelaar bepaalt welk niveau het best overeenkomt met de kwaliteit van het beoordeelde werk (Sadler, 2009).

Rubrics kunnen ook gebruikt worden voor het geven van peer feedback, voor peer beoordeling en voor het interpreteren van docent feedback. Een rubric kan dus niet alleen summatief, maar ook formatief worden ingezet. Het formatief inzetten van een rubric start aan het begin van het afstudeertraject. Doordat de rubric een beschrijving geeft van de verschillende kwaliteitsniveaus op de criteria van een afstudeeropdracht, is het voor studenten duidelijk wat er van hen verwacht wordt.

Sadler betoogt dat voor het begrijpen van beoordelingscriteria en standaarden een student het liefst over de schouder van een expert-beoordelaar mee zou moeten kijken bij het beoordelen (Sadler, 1989 en 2010). Dat gebeurt helaas meestal niet. Een rubric geeft de student echter al veel meer inzicht in de beoordeling dan traditionele beoordelingscriteria. Een rubric biedt studenten dus houvast bij het werken aan een afstudeeropdracht. Een recente review van Panadero en Jonsson (2013) laat zien dat het gebruik van rubrics de transparantie van het beoordelingsproces verhoogt, testangst reduceert en het zelfvertrouwen en de zelfregulatie van het leren van studenten verhoogt.

Aan de Universiteit Utrecht is een rubric ontwikkeld voor het beoordelen van afstudeeronderzoeksartikelen. De rubric beschrijft vier kwaliteitsniveaus voor de verschillende onderdelen van een onderzoeksartikel, te weten een theoretische introductie, een methode-sectie, een resultaten-sectie en een discussie (De Kleijn, Prins & Van Tartwijk, 2013). Ook is er in deze rubric plaats ingeruimd om de kwaliteit van het onderzoeksproces, met name de rol van de student daarin, te beoordelen. Bekijk de rubric in de ToetsSpecial op de website Toetswijzer, <http://www.toetswijzer.nl/berichten/toetsspecial-toetsen-op-school-hoger-onderwijs>.

Peer beoordeling

Als studenten, al of niet met een rubric, het werk van medestudenten beoordelen, heet dat peer beoordeling (peer assessment). Voor een effectief gebruik van peer beoordeling tijdens afstudeerwerk dienen studenten allereerst uitleg te krijgen over het gebruik van de rubric. Alle begrippen in een rubric moeten voor de student helder zijn. Vervolgens moeten peer beoordelingsactiviteiten in de procedure en planning worden opgenomen. Met andere woorden: de peer beoordeling moet worden georganiseerd. Dat betekent dat onder meer duidelijk moet zijn voor studenten wanneer de peer beoordeling plaatsvindt, wiens afstudeerwerk er wordt beoordeeld, hoeveel feedback er moet worden gegeven, aan welke kwaliteitseisen de peer feedback moet voldoen en hoe de peer feedback gecommuniceerd wordt naar de medestudenten (Sluijsmans, 2002; Sluijsmans & Prins, 2006). Het resultaat van een juist ingevoerde peer beoordelingsprocedure is dat studenten beter weten wat er bij een taak, zoals het afstudeerwerk, van hen wordt verwacht en dat zij meer feedback ontvangen (zie bijvoorbeeld Gielen, Peeters, Dochy, Onghena & Struyven, 2010).

In de ToetsSpecial op de website Toetswijzer, <http://www.toetswijzer.nl/berichten/toetsspecial->

[toetsen-op-school-hoger-onderwijs](#), is een voorbeeldformulier peer beoordeling beschikbaar.

Zelfbeoordeling

Ook voor het gebruik van zelfbeoordeling (self assessment) geldt dat het deel uit moet maken van de procedure en dat het ondersteund moet worden met een instructie of een beoordelingsinstrument, wil het effectief zijn bij afstudeeropdrachten. Orsmond en Merry (2013) betogen dat zelfbeoordeling studenten helpt om feedback van de docent beter te begrijpen.

In de ToetsSpecial op de website Toetswijzer, <http://www.toetswijzer.nl/berichten/toetsspecial-toetsen-op-school-hoger-onderwijs>, is een voorbeeldformulier zelfbeoordeling te vinden.

4.11 Beoordelen en begeleiden bij afstudeeropdrachten

Bij afstudeeropdrachten staat de zelfstandige aanpak door de student centraal. Dat wil niet zeggen dat er geen enkele begeleiding plaatsvindt. Sterker, vaak is meer dan één begeleider op verschillende momenten bij de uitvoering en begeleiding betrokken.

Afstudeerstage

Bij een afstudeerstage is sprake van een begeleider op de stageplek en een begeleidend docent vanuit de opleiding. Beide spelen op andere tijdstippen verschillende rollen: de stagebegeleider gedurende het hele traject, de docent vanuit de opleiding alleen aan het begin en het eind van de stage. De stagebegeleider kan zelf ook weer aangestuurd zijn door een externe opdrachtgever. Die opdrachtgever is dan soms op de achtergrond raadgever, maar wil ook tevreden zijn met het opgeleverde product. De indruk die de docent aan het eind van de stage heeft van de student wordt dus beïnvloed door het oordeel van de stagebegeleider en de opdrachtgever en eventueel nog andere betrokkenen in de stageorganisatie. Bij stages is het eerder gewoonte dan uitzondering dat er meerdere partijen betrokken zijn bij het vormen van een oordeel over het succes ervan. Desondanks is uiteindelijk de docent verantwoordelijk voor de beoordeling van de stage. De docent draagt daarmee verantwoordelijkheid tegenover de opleiding, de student en de stageorganisatie. Om de organisatie als stageplaats te behouden en een eerlijk afgewogen oordeel over de stage te geven waar alle partijen mee kunnen leven, is het van belang dat de begeleidend docent – die slechts op afstand betrokken is bij de stage – aan het begin van de stage heldere afspraken maakt met alle partijen. Met name over op te leveren producten is het verstandig om zwart-op-wit te zetten wat alle partijen aan het einde van de stage kunnen verwachten. De belangen van de betrokken partijen kunnen uiteen lopen: de opdrachtgever wil een product, de stagebegeleider wil een intern verslag van werkzaamheden en de docent wil een (theoretisch onderbouwde) verantwoording van de student. De student die alleen het laatste oplevert, kan een stage succesvol afronden (er is tenslotte aan de voorwaarde van de opleiding voldaan) maar de verhoudingen kunnen daardoor verstoord raken. Het is daarom ook belangrijk om de stage af te ronden met een persoonlijk gesprek waarbij alle partijen betrokken zijn en waarin alle producten naast elkaar worden gelegd. De stagebegeleider kan dan een eigen oordeel geven over de stage dat meetelt in de eindbeoordeling.

Afstudeeronderzoek

Bij een afstudeeronderzoek is de begeleiding vanuit de instelling vaak intensiever. De begeleider kan in principe de drie rollen van opdrachtgever, dagelijks begeleider en beoordelaar in één persoon verenigen. Dit maakt het beoordelen van het eindproduct complex. De hoeveelheid begeleiding die de student nodig heeft om tot een succesvol eindproduct te komen verschilt per student en per begeleider. Studenten verschillen in achtergrondkennis en manier van aanpak, en dus zal de hoeveelheid steun en behoefte aan feedback in elk individueel geval verschillen.

Het ligt voor de hand dat dit van invloed is op de uiteindelijke beoordeling. Immers, de hoeveelheid en aard van de begeleiding gedurende het afstudeertraject staat in verband met de door de student getoonde zelfstandigheid tijdens de uitvoering. De toevoeging van een rubriek 'mate van getoonde zelfstandigheid' aan een beoordelingsformulier is hiervoor niet voldoende. Het risico bestaat dat het inzicht van de student en de begeleider hierover ver uiteenlopen. Een reden daarvoor is dat gedurende veel afstudeertrajecten de begeleiding in het begin intensiever zal zijn dan tegen het einde. Studenten kunnen het belang van een goede onderzoeksvraag als schakel tussen theorie en methode onderschatten, en juist daarom is de eerste fase van onderzoek van cruciaal belang voor het hele traject. Echter, tegen het einde van het traject resteert slechts de verslaglegging van het onderzoek, een fase die vooral wordt gekenmerkt door tekstuele feedback.

Bij het oordeel over zelfstandigheid kan de intensieve eerste fase uit het oog zijn verloren, wat tot een geflatteerde score kan leiden. Twee manieren om op dit punt tot een meer valide oordeel te komen, zijn het bijhouden van de verstrekte begeleidingstijd en het notuleren (en later archiveren) van begeleidingsgesprekken. De eerste manier geeft een kwantitatieve indicatie van de hoeveelheid begeleiding die is gebruikt en daarmee een ruwe indicatie van zelfstandigheid. De tweede manier is zorgvuldiger omdat deze ook een inzicht geeft in de inhoudelijke beslissingen die zijn genomen, en wie er verantwoordelijk voor is geweest (student, begeleider, of in samenspraak met elkaar). In beide gevallen kan het nuttig zijn om van elke afstuderende student in een bestand bij te houden op welke wijze het contact is verlopen. Ook 'virtueel' contact (e-mail, telefoongesprekken, etc.) kan daarin worden bijgehouden. Een overzicht van wijze van contact, duur van het gesprek (als dat van toepassing is) en de aard van de genomen beslissingen kunnen een heldere indicatie geven van de door studenten getoonde zelfstandigheid.

4.12 Portfoliobeoordeling

Ook portfolio's kunnen voor formatieve en summatieve beoordelingsdoeleinden worden gebruikt. Een portfolio bestaat uit een verzameling van het werk van een student over een langer tijdsbestek waarin de ontwikkeling op bepaalde vaardigheden en attitudes wordt getoond. Portfolio's kunnen verschillen in de mate waarin ze zijn gestructureerd. Doorgaans bestaan ze uit producties (stukken die speciaal gemaakt zijn voor het portfolio), reproducties (registraties van het werk van een student) en oordelen van anderen over het werk van de student, bijvoorbeeld peer feedback of oordelen van de supervisor (Collins, 1991). Vandaag de dag wordt vaak gewerkt met elektronische portfolio's, wat doorgaans inhoudt dat een student een metaportfolio opstelt met al het materiaal over zijn ontwikkeling waaruit hij kan selecteren welke materialen hij voor welke beoordelingsmomenten en beoordelingsdoelen zichtbaar maakt aan de beoordelaar.

Aangezien portfolio's een rijke en authentieke afspiegeling vormen van het werk van een student wordt een portfolio vaak als een valide instrument voor de beoordeling van vaardigheden en competenties gezien. Echter, de gegevens in een portfolio zijn vaak beschrijvend, contextgebonden, uniek en persoonlijk van aard en vragen om veel interpretatie voordat ze te scoren zijn (Moss, 1994). Dit kan afbreuk doen aan de betrouwbaarheid en validiteit van de beoordeling (zie ook hoofdstuk 3). Shapley en Bush (1999) en Reckase (1995) beweren dat interbeoordelaarsbetrouwbaarheid van portfoliobeoordelingen kan toenemen wanneer de inhoud van het portfolio wordt gespecificeerd en normen en scoringsregels van te voren duidelijk zijn vastgesteld.

5 De examencommissie in het hoger onderwijs

5 De examencommissie in het hoger onderwijs

Jan Adema

Een tweetal wijzigingen in de Wet op het hoger onderwijs en wetenschappelijk onderzoek (WHW) heeft de examencommissie de afgelopen jaren nadrukkelijk de taak gegeven om de kwaliteit van de examinering en het eindniveau van de opleiding zeker te stellen. De onafhankelijkheid en de borgende rol van de examencommissie zijn versterkt doordat het management van de opleiding is uitgesloten van deelname en doordat een extern lid moet worden aangesteld.

De wet geeft echter maar weinig aanwijzingen over hoe het borgen van kwaliteit in zijn werk zou moeten gaan en hoe de commissieleden hun rol in de organisatie moeten invullen. Hoe kan de examencommissie de kwaliteit zodanig garanderen dat met vertrouwen onder elk diploma een handtekening gezet kan worden? En hoe verhoudt de examencommissie zich tot de andere actoren in de onderwijs- en examenorganisatie?

Veel examencommissies hebben nieuwe leden gekregen doordat er binnen de nieuwe regelgeving geen plaats meer is voor management in examencommissies en omdat externe leden moeten worden aangesteld. Deze nieuwe leden hebben behoefte aan ondersteuning om zich in te werken in hun taak. Hiermee is er behoefte ontstaan aan informatie en handreikingen over de invulling van de rol van de examencommissie.

Dit hoofdstuk is bedoeld om examencommissieleden te informeren over hun rol in de organisatie en een beeld te schetsen van het werk van de examencommissie. Daarnaast is het ook bedoeld voor leden van de toetscommissie, examinatoren en opleidingsdirectie zodat deze functionarissen weten wat ze van de examencommissie kunnen of zouden mogen verwachten.

Het belangrijkste uitgangspunt voor dit hoofdstuk is de wettekst waarin staat dat de examencommissie bepaalt of de kandidaat voldoet aan de eisen volgens de onderwijs- en examenregeling (OER) en dus een diploma kan uitreiken. Deze laatste stap in de keten van het examineringsproces kan de examencommissie alleen zetten als ze zich van de kwaliteit van alle voorgaande stappen (opstellen van het examenprogramma en de examens, afname, beoordeling, cesuur en registratie van resultaten) op de hoogte heeft gesteld.

Een situatie waarbij pas bij de beslissing over diplomering van een kandidaat tekortkomingen aan het licht komen over de kwaliteit van de examinering, waardoor een diploma niet toegekend kan worden, is hoogst onwenselijk en moet voorkomen worden. Door een proactieve, onderzoekende houding, en waar nodig het geven van adviezen ter verbetering van de kwaliteit, borgt de examencommissie het eindniveau van de opleiding.

Met examens worden in dit hoofdstuk bedoeld: tentamens, toetsen, scripties, eindwerkstukken en andere opdrachten die een beoordeling opleveren die meetelt voor het behalen van het getuigschrift. Daar waar teksten uit de WHW geciteerd worden, is de term 'tentamens' gehandhaafd.

5.1 Examencommissie in ontwikkeling

In 2009 verscheen het onderzoeksverslag 'Boekhouder of wakend oog' van de Inspectie van het Onderwijs. Na eerdere signalen dat de examencommissies in het hoger onderwijs niet zouden kunnen instaan voor het eindniveau en de kwaliteit van de toetsing besloot de Onderwijsinspectie tot een uitgebreid en representatief onderzoek onder de examencommissies in het hoger onderwijs. Uit dit onderzoek bleek dat de borging van het eindniveau tekortschoot. De examencommissie stond onvoldoende garant voor de betrouwbaarheid en validiteit van de toetsen en de handhaving van regels rond beoordeling en normering. De examencommissies hadden bovendien te weinig kennis op het gebied van wetgeving en toetsing.

In deze periode kwam een hogeschool in opspraak doordat bij een speciaal traject voor langstudeerders concessies waren gedaan aan het te behalen eindniveau. Daarop volgden meer berichten over de gebrekkige kwaliteit van het onderwijs en de examinering bij diverse instellingen in het hoger onderwijs.

Intussen scherpste de Nederlands-Vlaamse Accreditatieorganisatie (NVAO) de regels aan, onder andere met de eis dat toetsing en examinering per se van voldoende kwaliteit moeten zijn om de accreditatie te kunnen verkrijgen. Maar met een accreditatie die eens in de zes jaar plaatsvindt, vond de wetgever de waarde van elk individueel diploma nog onvoldoende gegarandeerd. De aanbeveling uit 'Boekhouder of wakend oog', dat een betere wettelijke verankering van de rol van de examencommissie nodig is, lag op dat moment al klaar en in 2010 werd het wetsvoorstel Versterking besturing (hoger onderwijs) aangenomen in de Eerste Kamer. Belangrijke punten in deze wet zijn:

- het bestuur zorgt voor een onafhankelijke en deskundige examencommissie;
- de examencommissie borgt de kwaliteit van de tentamens en examens;
- de examencommissie bepaalt of een student voldoet aan de voorwaarden die de onderwijs- en examenregeling stelt voor het behalen van het diploma;
- de examencommissie stelt een jaarverslag op.

In de publicatie 'Geslaagd! : handreiking examencommissies' (Vereniging Hogescholen, 2011) hebben de hogescholen afspraken gemaakt over naleving van de wet en een aantal gedragsregels uitgewerkt. Op het punt van de onafhankelijkheid van de examencommissie zijn deze gedragsregels explicieter dan de wettekst, zie pag. 9:

'Wil de examencommissie haar werk goed kunnen doen, dan moet deze onafhankelijk kunnen opereren. Dat wil zeggen: zonder druk van het management. De onderliggende veronderstelling daarbij is dat indien personen met een managementfunctie lid zouden zijn van een examencommissie, zij een financieel belang of een ander niet-onderwijskundig belang kunnen laten prevaleren.'

Hiermee spreken de hbo-instellingen af dat het management geen zitting heeft in de examencommissie. Ook de niet-hiërarchische relatie van de examencommissie ten opzichte van het bestuur wordt nader uitgewerkt, zie pag. 11:

'(...) de instelling moet zorgen dat de examencommissies binnen de instelling onafhankelijk hun werk kunnen verrichten. Dat betekent bijvoorbeeld dat het college van bestuur geen verplichtingen kan opleggen aan de examencommissie inzake de beoordeling van studenten. Het college blijft wel eindverantwoordelijk voor de kwaliteit van het onderwijs en de graadverlening; een examencommissie moet handelen binnen de grenzen van de onderwijs- en examenregeling.'

In de bijlage van 'Geslaagd! : handreiking examencommissies' is een handig overzicht te vinden van relevante artikelen uit de WHW over de examencommissie.

In 2012 verscheen het rapport van de commissie Bruijn getiteld 'Vreemde ogen dwingen'. In opdracht van de Vereniging Hogescholen doet de commissie aanbevelingen om het vertrouwen in de examenkwaliteit te vergroten door 'vreemde ogen' toe te laten bij de examens, bijvoorbeeld door:

- samen te werken bij het construeren van instellingsoverstijgende examens;
- externe leden te benoemen in de examencommissies;
- externe beoordelaars te betrekken bij het beoordelen van de eindwerkstukken;
- de deskundigheid van examinatoren en examencommissies te vergroten.

Een andere aanbeveling in dit rapport is om een certificering voor examinatoren in te stellen: een basiskwalificatie examinering en een seniorkwalificatie examinering (BKE/SKE). De BKE is gericht op de examinerator en de SKE op de deskundigheid van onder andere examencommissie-leden en beleidsmedewerkers op het gebied van het examenprogramma als geheel.

In het wo is dezelfde wetgeving van kracht als in het hbo. De krachtdadige aanpak van de Vereniging Hogescholen heeft ervoor gezorgd dat het hbo in de uitvoering van de nieuwe regelgeving een voorsprong heeft op het wo.

Gelijktijdig met de ontwikkelingen in het hbo, zoals hiervoor beschreven, vond in het middelbaar beroepsonderwijs (mbo) een soortgelijke ontwikkeling plaats waarbij de examencommissie nadrukkelijk de rol kreeg om de kwaliteit van de examinering te borgen. In het 'Toezichtkader bve 2012' heeft de Onderwijsinspectie een concrete aanpak beschreven hoe het borgen in zijn werk gaat, zie pag. 51: 'De examencommissie vergewist zich met gepaste grondigheid en frequentie van de kwaliteit van het exameninstrumentarium, de afname, de beoordeling, de besluitvorming betreffende de diplomering en van de deskundigheid van bij de examinering betrokken personen. Om vast te stellen of de beoordeling adequaat verloopt, wordt gebruikgemaakt van bijvoorbeeld resultatenanalyses, studenttevredenheidsevaluaties en gegevens uit (steekproefsgewijze) dubbele beoordelingen.' Verder heeft de commissie volgens het toezichtkader als taak om oorzaken van tekortkomingen te analyseren, aanbevelingen te doen en te volgen of deze effect hebben. In het mbo ligt de nadruk niet op een onafhankelijke examencommissie. De schooldirecteur als voorzitter van de examencommissie is in het mbo (nog) sterk geworteld in de organisatie.

Overstijgend over de verschillende onderwijssectoren heeft de Nederlandse Vereniging voor Examens (NVE) in 2011 functieprofielen voor examenfunctionarissen en een bijbehorend organogram voor de examenorganisatie ontwikkeld. In 2013 en 2015 is een herziene versie gepubliceerd. Hier wordt de examenorganisatie los van het onderwijs beschouwd. Dit sluit aan bij de gedachte dat examinering een proces is dat eigen regels kent en andere kennis en vaardigheden vraagt dan het verzorgen van onderwijs. In het model van de NVE is de manager examinering buiten de examencommissie geplaatst en heeft de examencommissie de rol van verantwoordelijke voor borging van de examenkwaliteit. Het management voert de examinering uit zoals deze is vastgelegd in de OER en de examencommissie houdt toezicht op de examenkwaliteit en het behaalde eindniveau. Paragraaf 5.2.1 gaat nader in op het NVE-functieprofiel.

In de Wet versterking kwaliteitswaarborgen hoger onderwijs, die in 2014 van kracht is geworden in het hbo en in 2015 in het wo, wordt de lijn van eerdere wetgeving en adviezen doorgezet. In deze wet is bepaald dat leden van de examencommissie geen financiële verantwoordelijkheid mogen dragen voor de opleiding. Tevens is de verplichting van een extern lid in de examencommissie vastgelegd. In het voorafgaande wetsvoorstel werd het onderzoek aangehaald van ResearchNed (2012) naar de ervaringen van hogescholen en universiteiten met externen in de examencommissie. Hieruit blijkt dat in alle onderzochte casussen in het hbo en wo de voordelen van een extern lid opwegen tegen de kosten. In deze casussen wordt de volgende toegevoegde waarde van het externe lid onderscheiden:

- inbreng kennis (met name toetsdeskundigheid, juridische kennis, werkveld);
- frisse, kritische blik en daardoor bijdrage aan een zakelijke, professionele en open houding in de commissie zelf en uiteindelijk ook aan de kwaliteit van de commissie;
- grotere standvastigheid en onafhankelijkheid en meer gezag/legitimiteit van de examencommissie ten opzichte van zowel bestuur, management als docenten.

In opdracht van de Vereniging Hogescholen bracht de Expertgroep Protocol in 2014 onder de titel 'Beoordelen is mensenwerk' advies uit over de wenselijkheid en mogelijkheid om een gezamenlijk protocol voor het afstuderen in te voeren. Volgens de Expertgroep is een gezamenlijk protocol voor alle hbo-opleidingen niet wenselijk. Wel doet de Expertgroep de aanbeveling om samen te werken bij de borging van de kwaliteit van afstudeerprogramma's. Verder geeft de Expertgroep aanbevelingen hoe onderzoekend vermogen beoordeeld kan worden binnen het afstuderen. De Vereniging Hogescholen heeft de aanbevelingen overgenomen als inspiratiebron, niet als verplichting. De rapportage van de Expertgroep geeft examencommissies handvaten om de kwaliteit van het afstudeerprogramma te bevragen. Kijk voor meer informatie op www.vereniginghogescholen.nl > Kennisbank > Thema's > Onderwijs: Pilot 'Protocol afstuderen'.

In 2015 verscheen 'Verdere versterking', het tweede onderzoeksrapport van de Inspectie van het Onderwijs naar het functioneren van examencommissies in het hoger onderwijs, als vervolg op 'Boekhouder of wakend oog' uit 2009. En de Vereniging Hogescholen publiceerde een nieuwe editie van de 'Handreiking examencommissies' (2015). Beide documenten bieden concrete aanknopingspunten voor examencommissies om hun werkzaamheden vorm te geven.

5.2 Rol van de examencommissie in de organisatie

In de WHW staan meerdere artikelen die voor de examencommissie van belang zijn. Deze kunnen voor iemand die nieuw begint als lid van een examencommissie nogal fragmentarisch overkomen en ze geven samen niet automatisch een helder beeld van de rol die de commissie speelt in de organisatie. Het is nodig om dat samenhangend beeld te schetsen om het begrip van de rol van de commissie voor nieuwe en zittende leden en andere geïnteresseerden te vergroten. Een examencommissielid wil niet alleen doen wat er in de wet staat, maar wil dit ook begrijpen en als zinvol ervaren. Om dit te kunnen realiseren, is een duidelijk beeld nodig van wat de examencommissie doet, zodat bij de verschillende betrokkenen in de organisatie dezelfde interpretaties en verwachtingen van deze rol ontstaan. Doel van deze paragraaf is om de rol van de examencommissie samenhangend te schetsen en de relatie met andere actoren in de organisatie te beschrijven.

5.2.1 Examencommissie volgens het NVE-functieprofiel

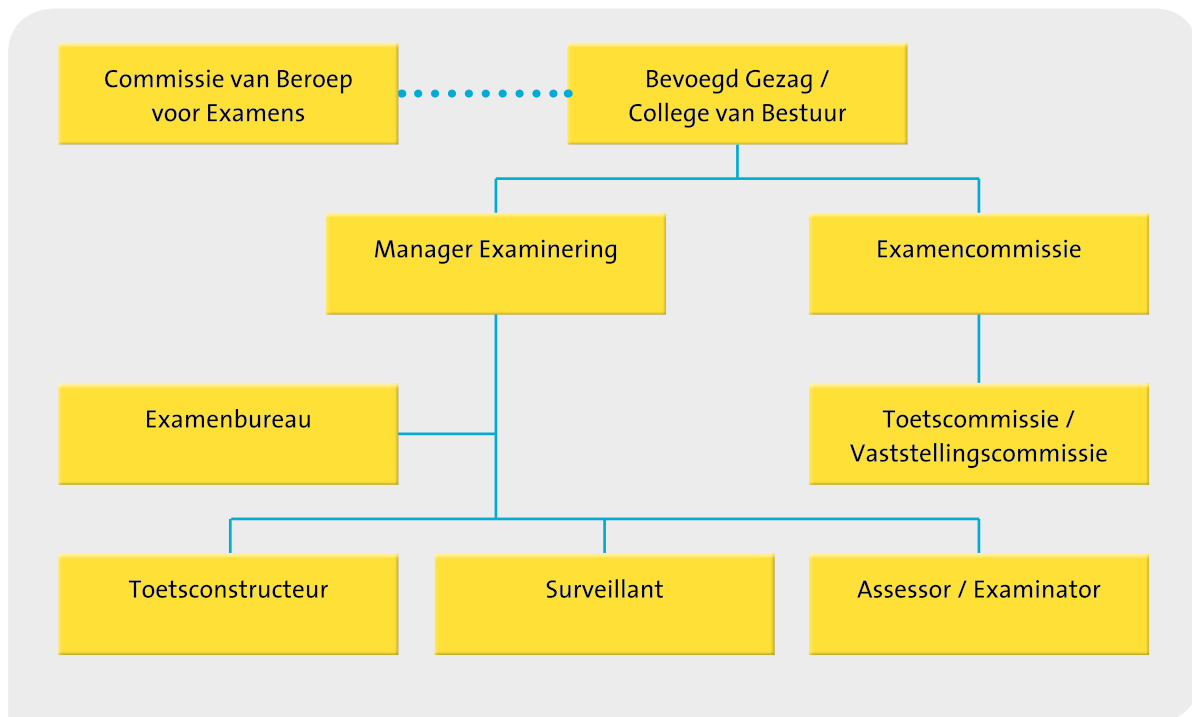
In het NVE-functieprofiel staat het doel van de examencommissie, gestructureerd in twee deeltaken, beschreven (zie ook www.nvexamens.nl > Functieprofielen). De examencommissie borgt dat gediplomeerden voldoen aan de exameneisen voor het verkrijgen van een diploma of certificaat. De examencommissie geeft aan deze verantwoordelijkheid inhoud door:

- 1 Het bewaken van de kwaliteit van het examenproces:
 - het evalueren van het examenproces;
 - het signaleren van risico's die de kwaliteit van de examinering bedreigen;
 - het doen van voorstellen aan het bevoegd gezag om die risico's te beperken;
 - het (mede) opstellen van het examenreglement;
 - het aanwijzen van assessoren en examinatoren op basis van hun deskundigheid;
 - het volgen van ontwikkelingen binnen en buiten de organisatie die de kwaliteit van de examinering kunnen beïnvloeden.
- 2 Het nemen van beslissingen waarbij het voldoen aan de exameneisen in het geding is, zoals:
 - het vaststellen van de kwaliteit van het examen;
 - het vaststellen of een examendeelnemer voldoet aan de eisen voor het verkrijgen van een diploma;

- het verlenen van vrijstellingen;
- het verlenen van toestemming voor het afleggen van een afwijkend examen;
- het beslissen over aanpassingen van de cesuur na afname van een examen;
- het beslissen over klachten van kandidaten over het examen of over de beoordeling.

Het NVE-functieprofiel bevat de meeste taken die in de wet genoemd staan. In paragraaf 5.3 worden de wettelijke taken en de overige genoemde taken besproken.

In het bijbehorende NVE-organogram van de examenorganisatie, zie figuur 5.1, vinden we de examencommissie onafhankelijk van het management gepositioneerd.



Figuur 5.1: NVE-organogram van de examenorganisatie (bron: www.nvexamens.nl)

5.2.2 Relatie met andere actoren in de organisatie

De lijnen in het NVE-organogram in figuur 5.1 geven de hiërarchische structuur aan in de organisatie. In de dagelijkse praktijk van het onderwijs lopen er veel functionele en informele lijnen tussen de actoren en heeft de examencommissie met verschillende personen en partijen uit de examenorganisatie en de onderwijsorganisatie te maken. We lichten de belangrijkste relaties toe.

De examencommissie en het bestuur van de instelling

Na benoeming door het bestuur of de decaan staat de examencommissie niet in een hiërarchische relatie met het bestuur. Dit wil zeggen dat het bestuur de leden geen verplichtingen op kan leggen. Wel is het bestuur rechtstreeks verantwoordelijk voor de facilitering van de examencommissie. Dit biedt de examencommissie de mogelijkheid om een onafhankelijke rol te nemen en de kwaliteit te bewaken. Een beslissing van de examencommissie heeft hierdoor een absolute waarde en kan niet genegeerd worden. De examencommissie lijkt daarmee een fremdkörper in de organisatie en, hoewel ze wel gebonden is aan de regels van de instelling zoals deze bij de benoeming zijn afgesproken, lijkt ze oncontroleerbaar. Dit is een merkwaardige situatie die mogelijk tot problemen zou kunnen leiden als een examencommissie een geheel eigen lijn zou volgen bij een belangrijke beslissing. Jurisprudentie is er op dit punt nog niet. Examencommissie en bestuur zijn, zoals het ernaar uitziet, gebonden om er in overleg samen

uit te komen. Een examencommissie(lid) uit zijn functie zetten, is een maatregel die alleen onder extreme omstandigheden zou mogen voorkomen.

De examencommissie en het management

In het hbo en wo is in het algemeen de directeur van de opleiding of cluster van opleidingen ook verantwoordelijk voor het uitvoeren van de examinering. De directeur van de opleiding heeft geen zitting in de examencommissie. De kracht van de examencommissie moet juist zijn dat ze de kwaliteit borgt en hierbij beslissingen neemt en adviezen geeft zonder financiële of rendementsmotieven. De examencommissie heeft echter geen budget en daarmee is de uitvoering van adviezen en beslissingen alleen mogelijk na een akkoord van het management. Hier kunnen kwaliteit en kosten tot tegengestelde belangen leiden. Als de examencommissie bijvoorbeeld adviseert om een extra surveillant in te zetten, is de uitvoering van deze maatregel alleen te realiseren door inzet van mensen en middelen en dit is een beslissing van het management.

De onafhankelijkheid van de examencommissie werkt naar twee kanten. De examencommissie houdt zich daarom niet bezig met het (onderwijs- en) toetsbeleid anders dan hierover te adviseren binnen haar wettelijke taak. Als een opleiding ervoor kiest om in het eerste jaar vooral meerkeuzetoetsen af te nemen, kan een examencommissie met een andere voorkeur dit niet tegenhouden. Overleg tussen examencommissie en management is daarmee een voorwaarde om tot een gezamenlijk onderschreven examenkwaliteit te komen. Daar waar examencommissie en management geen overeenstemming kunnen bereiken, loopt de lijn van overleg via het bestuur. Voor de uitvoering van structurele wijzigingen in het systeem van toetsing – zoals de voorstellen uit het rapport ‘Vreemde ogen dwingen’ – zullen management en examencommissie samen op moeten trekken. Dit samen optrekken kan vorm krijgen door het management bij de vergaderingen als toehoorder/adviseur uit te nodigen of door periodiek overleg tussen de voorzitter en het management. Dit overleg vindt plaats met behoud van de onafhankelijkheid en verantwoordelijkheid van beide rollen.

De examencommissie en de toetscommissie

In het hoger onderwijs is vaak een toetscommissie aangesteld om de kwaliteit van de tentamens te onderzoeken en adviezen ter verbetering te geven. Het ligt voor de hand dat de toetscommissie in deze rol inhoudelijk aangestuurd wordt door de examencommissie. De examencommissie heeft immers als wettelijke taak om de kwaliteit van tentamens te borgen. De examencommissie bepaalt de methode van onderzoek naar de kwaliteit van examens en de prioriteiten in de werkzaamheden van de toetscommissie, en maakt hierbij gebruik van de specifieke expertise van de leden van de toetscommissie. De toetscommissie rapporteert aan de examencommissie over de voortgang van de werkzaamheden en resultaten. Indien een commissie is aangesteld voor het organiseren en plannen van examens is aansturing door het management voor de hand liggend. Deze zou dan voor de helderheid niet de naam toetscommissie moeten hebben. De examencommissie kan, behalve van het werk van de toetscommissie, ook gebruikmaken van onderzoeken die door de afdeling kwaliteitszorg gedaan worden. Studenttevredenheid over examinering bijvoorbeeld wordt meestal door de afdeling kwaliteitszorg onderzocht.

De examencommissie en de examinerator

De examinatoren zijn degenen die de prestaties van een student beoordelen. De docent is meestal ook de examinerator van zijn vak. Dat wil zeggen dat hij het onderwijs geeft, de tentamenvragen maakt en het werk van de studenten beoordeelt. Voor een objectieve beoordeling van studenten zou het goed zijn als de examinering in zekere mate los staat van het onderwijs. Het rapport ‘Vreemde ogen dwingen’ doet ten aanzien van dit punt aanbevelingen om tot meer externe validering van de kwaliteit te komen waardoor een objectiever oordeel mogelijk wordt. Dit kan onder andere door samenwerking bij het ontwikkelen van instellingsoverstijgende toetsen, waarbij meerdere docenten van verschillende

instellingen de vragen maken, of door uitwisseling van beoordelaars tussen instellingen. We zien dat voor de beoordeling van het eindwerkstuk bij de meeste opleidingen een tweede, onafhankelijke beoordelaar wordt ingezet. De examencommissie kan afhankelijk van de specifieke situatie per opleiding voorstellen doen om de objectiviteit te verhogen. De examencommissie ziet erop toe dat de examinerator de procedures volgt. Als de examencommissie vragen heeft aan de examinerator over zijn beoordelingen is de examinerator verplicht de examencommissie van informatie te voorzien (WHW, artikel 7.12c lid 2). De examencommissie kan de examinerator echter niet passeren en een ander cijfer toekennen. Wel kan de examencommissie de examinerator vragen het oordeel te heroverwegen of nader toe te lichten. Het zal niet veel voorkomen maar vanuit de bevoegdheid tot borgen van de kwaliteit kan de examencommissie op procedurele gronden een examen ongeldig verklaren (WHW, artikel 7.12b lid 1a). Een 'second opinion' die bij klachten wordt ingezet en waarbij het tweede oordeel als definitief geldt, kent geen wettelijke grondslag als hier de beoordeling van de eerste examinerator zonder duidelijke reden ongeldig wordt verklaard. Als de examencommissie wil dat twee examineratoren het werk beoordelen, moet dit vooraf in een procedure worden vastgelegd, inclusief de wijze waarop beide examineratoren tot een gezamenlijk oordeel komen. Als een docent niet aan de eisen voldoet die de OER stelt of niet goed functioneert als examinerator, dan kan de examencommissie besluiten deze docent niet (meer) als examinerator aan te wijzen. Dit heeft mogelijk veel impact op de docent en op de examenorganisatie en voordat deze beslissing wordt genomen moet dan ook in overleg met de leidinggevende en de docent samen naar oplossingen gezocht worden.

De examencommissie en andere examencommissies

Binnen de WHW is er niet expliciet sprake van samenwerking en afstemming tussen examencommissies; elke examencommissie heeft een zelfstandige beslisbevoegdheid. Afstemming tussen examencommissies binnen een universiteit of hogeschool is echter wel degelijk wenselijk zodat studenten binnen dezelfde instelling volgens dezelfde regels behandeld worden. Het kan toch voorkomen dat verschillende examencommissies binnen de instelling tot een andere interpretatie van een regel en daarmee tot verschillende beslissingen komen. Hoewel afstemming nodig en wenselijk is, is het aan te bevelen dat niet voor elk uitzonderlijk geval binnen een opleiding een voor de gehele instelling geldende regel wordt opgesteld. Sommige uitzonderingen komen maar één keer voor, of kunnen zonder een wijziging van de reglementen opgelost worden. Het overleg tussen examencommissies kan veel tijd in beslag nemen die niet meer beschikbaar is voor de overige taken. Bij instellingsoverstijgende toetsing zijn vaak meerdere examencommissies betrokken. Er moeten dan afspraken gemaakt worden om ervoor te zorgen dat er geen verschillen ontstaan in de afnamesituatie, de toegepaste cesuur of het aantal herkansingen waarmee de beoogde standaardisatie juist weer tenietgedaan wordt. De wet kent per opleiding maar één examencommissie. Een examencommissie mag wel voor meerdere opleidingen aangesteld worden (WHW, artikel 7.12 lid 1). Een gelaagde structuur, bestaande uit een overkoepelende examencommissie met daaronder subexamencommissies, is in strijd hiermee. Het wordt in een dergelijke constructie ondoorzichtig waar de verantwoordelijkheid en beslissingsbevoegdheid ligt. Als deze structuur met een gelaagde examencommissie toch voorkomt, is het meestal de subexamencommissie die diplomeert en de kwaliteit borgt, en daarmee de belangrijkste wettelijke taken uitvoert.

5.3 Taken van de examencommissie

Het werk van de examencommissie bestaat uit wettelijke en niet-wettelijke taken.

5.3.1 Wettelijke taken

De taken zoals in de wet vermeld, worden binnen de instelling geïnterpreteerd en vertaald in regels in de OER of in andere reglementen. De examencommissie gebruikt deze regels als leidraad bij het handelen en beslissen. Bij uitzonderlijke verzoeken is het goed om behalve de OER ook de wettelijke tekst te raadplegen om te kijken of de student in zijn recht staat.

Hieronder staan de artikelen uit de WHW (geldend in 2015), <http://wetten.overheid.nl/BWBR0005682>, waarin de examencommissie als actor vermeld wordt, voorzien van een korte toelichting.

Het op objectieve en deskundige wijze vaststellen of een student voldoet aan de voorwaarden die de onderwijs- en examenregeling stelt ten aanzien van kennis, inzicht en vaardigheden die nodig zijn voor het verkrijgen van een graad (WHW, artikel 7.12 lid 2)

De laatste stap bij de beslissing om het getuigschrift te verlenen, is het controleren of de student alle studiepunten heeft behaald en aan alle voorwaarden voldoet. Hiervoor moet een degelijke procedure ingericht worden waarbij gegevens zodanig worden bijgehouden dat de beslissing voor elke student te verantwoorden is:

- Voor elke beoordeling is een onderliggend bewijs in het dossier aanwezig.
- De beoordelingen zijn correct overgenomen in het digitale systeem.
- De beslisprocedure voor diplomering is correct geprogrammeerd in het digitale systeem.

Het borgen van de kwaliteit van tentamens en examens (WHW, artikel 7.12b lid 1a)

Deze taak wordt verderop in dit hoofdstuk besproken bij borging van de kwaliteit (paragraaf 5.4).

Het vaststellen van richtlijnen en aanwijzingen binnen het kader van de OER om de uitslag van de tentamens en examens te beoordelen en vast te stellen (WHW, artikel 7.12b lid 1b)

De examencommissie is gebonden aan de centraal vastgestelde regels zoals deze in de OER vermeld zijn. Als uitwerking van deze algemene regels kan de examencommissie voor een specifiek examen regels opstellen, bijvoorbeeld voor een beoordeling van een stage in het buitenland. De examencommissie kan ook een afwijkende examenvorm toestaan, zoals een mondeling examen voor iemand met dyslexie.

Het verlenen van vrijstellingen voor het afleggen van één of meer tentamens (WHW, artikel 7.12b lid 1d)

Gronden voor vrijstelling worden vastgelegd in de OER. Elke verleende vrijstelling dient met de argumentatie te worden vastgelegd. De examencommissie legt een lijst aan van beslissingen met argumentatie zodat bij toekomstige aanvragen een gelijke beslissing genomen wordt.

Het treffen van maatregelen in geval van fraude (WHW, artikel 7.12b lid 2)

Bij een vermoeden van fraude onderzoekt de examencommissie of er werkelijk sprake is van fraude en legt vervolgens een sanctie op. De sancties zijn vastgelegd in de OER en bestaan meestal uit uitsluiting van deelname aan examens voor een periode van maximaal een jaar waardoor de student een flinke studievertraging oploopt. Verwijderen van een fraudeur van de opleiding is een sanctie die de examencommissie niet zelfstandig kan nemen. Hiervoor moet ook de inschrijving bij de instelling beëindigd worden. De examencommissie heeft in het kader van de borging van de kwaliteit ook een rol bij het voorkomen van fraude door na te gaan of de uitvoering van de preventieve maatregelen voldoende is. Bij een vermoeden van fraude dient de examinerator dit te melden aan de examencommissie en (nog) geen oordeel te geven voor het gemaakte werk.

Het aanwijzen van examinatoren voor het afnemen van tentamens en het vaststellen van de uitslag daarvan (WHW, artikel 7.12c)

Bij het aanwijzen van examinatoren moet de examencommissie de deskundigheid van de examinatoren beoordelen. In de basiskwalificatie onderwijs (BKO) is een deel over examinering opgenomen. Hoewel er inmiddels voor het hbo een voorstel voor een aparte basiskwalificatie examinering (BKE) is uitgewerkt, is het nu nog zo dat examinatoren worden aangewezen op basis van hun BKO. Binnen veel opleidingen worden automatisch alle docenten deskundig en geschikt geacht als examinator voor hun vak. Een praktische gang van zaken is dan ook dat de examencommissie op basis van een lijst van docenten de examinatoren aanwijst. Deze praktijk blijkt niet altijd een garantie voor kwalitatief goede examens.

De examencommissie kan dan ook nadere eisen stellen aan examinatoren, bijvoorbeeld door te eisen dat voor het beoordelen van eindwerkstukken in de masterfase de examinator gepromoveerd is. Een andere voorwaarde voor aanwijzing kan zijn dat een examinator niet de begeleidend docent is bij een stage of eindwerkstuk. Welke rol een BKE-certificaat gaat spelen bij het aanwijzen, is op dit moment nog niet geheel helder. Instellingen moeten eerst de gelegenheid krijgen beleid en scholing op dit gebied vorm te geven.

In het wo geven ook assistenten in opleiding vaak een deel van het onderwijs en zij zijn dan ook bij de beoordeling betrokken. Dit gebeurt dan onder verantwoordelijkheid van een docent die als examinator is aangewezen. De examinator blijft in dit geval verantwoordelijk voor een goede werkwijze bij de beoordeling en kan hier door de examencommissie op aangesproken worden. Dit geldt ook als studenten elkaar beoordelen bij een groepsopdracht. De examinator zorgt dan door goede instructie en duidelijkheid over de criteria voor een eerlijke werkwijze. De wenselijkheid van beoordelen door assistenten of door studenten is ter beoordeling van de opleiding zelf.

Het beheer van een handtekeningenregister en het bijhouden wanneer nieuwe aanwijzingen nodig zijn, kan een taak zijn van de ambtelijk secretaris van de examencommissie. Een dergelijk register maakt het voor de administratie mogelijk om te zorgen dat gemaakt werk of beoordelingen door een bevoegd examinator worden aangeleverd en dat examenmateriaal alleen aan bevoegde personen wordt uitgeleverd.

Het uitreiken van het getuigschrift, met daaraan toegevoegd het diplomasupplement, ten bewijze dat het examen met goed gevolg is afgelegd (WHW, artikel 7.11 lid 2 en lid 4)

De student kan zijn getuigschrift aanvragen, of de examencommissie kan zelf bijhouden wanneer een student aan alle verplichtingen voldaan heeft. Een lid van de examencommissie tekent namens de examencommissie het getuigschrift. Een tweede handtekening is door of namens het bestuur.

Het verlenen van toestemming aan een student om een vrij onderwijsprogramma te volgen waarvan het examen leidt tot het verkrijgen van een graad (WHW, artikel 7.12b lid 1c)

Het beslissen over de toegestane keuze van minoren valt binnen deze regel. De examencommissie bekijkt of het programma als geheel aan het niveau en de inhoud van de opleiding voldoet. Het staat de student vrij om ook andere voorstellen te doen omtrent zijn onderwijsprogramma, zoals een buitenlandse stage die niet standaard in het programma staat. De overweging van de examencommissie om dit toe te staan betreft, naast de inhoud en het juiste niveau, ook of de examinering op een goede manier kan plaatsvinden. Een speciaal programma zal mogelijk extra kosten met zich meebrengen en hiermee zal de opleidingsdirecteur akkoord moeten zijn.

Als het volgen van een individueel, speciaal programma is goedgekeurd door de examencommissie moet de examencommissie vervolgens extra inspanning leveren bij het toezien op de juiste uitvoering en toetsing van het programma. In het verleden zijn opleidingen door onvoldoende toezicht op speciale programma's negatief in het nieuws gekomen.

Het uitreiken van een verklaring van behaalde tentamens aan diegene die meer dan één tentamen met goed gevolg heeft afgelegd maar aan wie niet een getuigschrift kan worden uitgereikt (WHW, artikel 7.11 lid 5)

Als een student wil overstappen naar een andere opleiding kan het voor hem van belang zijn om een verklaring van reeds behaalde examens te ontvangen om daarmee vrijstellingen te kunnen krijgen bij de nieuwe opleiding.

Het jaarlijks opstellen van een verslag van de werkzaamheden (WHW, artikel 7.12b lid 5)

Voor het jaarverslag stelt de instelling meestal een format beschikbaar. Een vuistregel is dat de notulen van alle vergaderingen samengevoegd het jaarverslag vormen. Er kunnen in het jaarverslag geen belangrijke voorvallen ongenoemd blijven en dat wat niet is besproken in vergaderingen kan geen deel uitmaken van het jaarverslag. Het jaarverslag bevat in ieder geval:

- Gegevens over de samenstelling en werkwijze van de commissie en een evaluatie van de werkwijze.
- Een verslag van de wettelijke taken:
 - een cijfermatig overzicht van vrijstellingen, fraude, (toename van) klachten, bijzondere verzoeken, speciale programma's met hierbij een toelichting of evaluatie;
 - een cijfermatig overzicht van getuigschriften (let wel dat het rendement van de opleiding geen onderdeel is van het jaarverslag van de examencommissie omdat de examencommissie hier nadrukkelijk niet verantwoordelijk voor is);
 - (een samenvatting van) de regels die de examencommissie heeft opgesteld voor de afname van examens;
 - een verslag van het onderzoek en daaruit voortvloeiende conclusies en adviezen over de kwaliteit van de examens.

Handelen naar aanleiding van uitspraken van het college van beroep (WHW, artikel 7.61 lid 5)

Het college van beroep voor de examens kan een examencommissie opdracht geven om een beslissing te heroverwegen of opnieuw te onderbouwen. De examencommissie is verplicht om opdrachten van het college van beroep uit te voeren.

5.3.2 Niet-wettelijke taken

Tot de niet-wettelijke taken behoren:

- Het behandelen van klachten (van studenten en docenten)
- Het beslissen over aanpassen van de cesuur
- Het geven van bindende studieadviezen

Het behandelen van klachten

Studenten worden steeds assertiever en gebruiken hun recht om een klacht in te dienen of in beroep te gaan tegen beslissingen van examinatoren en examencommissies. Strikt genomen is het behandelen van klachten geen wettelijke taak, de hogeschool of universiteit mag dit ook anders regelen en bijvoorbeeld een aparte klachtencommissie instellen. In de praktijk worden echter bij de meeste instellingen klachten over toetsing wel door de examencommissie behandeld. De student mag formeel met zijn klacht de examencommissie passeren en direct naar het college van beroep stappen. In dat geval zal het college de student altijd eerst terugverwijzen naar de examencommissie om daar tot een oplossing te komen. De examencommissie op haar beurt zal een student die nog niet in overleg is geweest met de examinator(en) eerst daarnaar terugverwijzen.

Het behandelen van klachten en verzoeken is een valkuil voor de examencommissie. Het komt voor dat een commissie tekortschiet in het uitoefenen van de overige taken doordat onevenredig veel tijd wordt besteed aan het behandelen van klachten en verzoeken die in een niet-aflatende stroom van e-mails binnenkomen. Bij de behandeling van de klachten en verzoeken is het goed

om als commissie uit de rol van docent te stappen en ook andere aspecten van de zaak te bekijken, zoals de consequenties voor de organisatie. Aan een besluit om een extra herkansing toe te staan, kleef bijvoorbeeld mogelijk de consequentie dat een examiner nieuwe vragen moet maken. De vraag is of dat altijd haalbaar is.

Maatregelen om de tijd aan klachten en verzoeken te beperken:

- Plan een gesprek van de student met een docent of studieloopbaanbegeleider voordat de klacht bij de examencommissie komt. Dit kan veel klachten oplossen voor ze zijn ingediend. Bij een klacht speelt vaak meer mee dan alleen ontevredenheid over een examen.
- Gebruik een digitaal invulformulier waarbij de student gestructureerd alle informatie zelf aanlevert met betrekking tot de klacht. Dit spaart veel tijd bij de behandeling.
- Mandateer het afhandelen van klachten en verzoeken aan een dagelijks bestuur waardoor niet de gehele commissie ermee bezig is.
- Wees streng en rechtvaardig hoewel dat moeilijk is voor leden die in de docentrol zitten.
- Voer consequent kwaliteitsbeleid uit zoals in paragraaf 5.4 wordt behandeld bij 'Borgen van de kwaliteit van examinering'. Dit vermindert het aantal klachten en maakt reageren op klachten gemakkelijker omdat er gegevens beschikbaar zijn betreffende de kwaliteit.
- Kwaliteit van toetsing berust voor een groot deel op standaardisering. Bij uitzonderingen is de kwaliteit vaak te weinig gegarandeerd. Dit maakt het afwijzend reageren op verzoeken beter uitlegbaar.

Het komt ook voor dat docenten klagen bij de examencommissie over een examen of een beslissing van de examencommissie. Hiervoor is wettelijk niets geregeld, een afhandeling van een dergelijke klacht kent daardoor geen formeel karakter. Een docent kan voor een formele klacht bij de directie terecht.

Het beslissen over aanpassen van de cesuur

Een bijzondere beslissing binnen de examenorganisatie is het vaststellen van de definitieve cesuur. Wie bepaalt dat het cesuurvoorstel zoals dat is geformuleerd bij het opstellen van het examen blijft staan na afname of dat de cesuur wordt gewijzigd? De WHW geeft de examiner/assessor de rol om het tentamen af te nemen en de uitslag te bepalen (WHW, artikel 7.12c). Aan de andere kant heeft de examencommissie de mogelijkheid om richtlijnen en aanwijzingen te geven voor het bepalen van de uitslag. Ook vanuit de verantwoordelijkheid voor het eindniveau is de examencommissie betrokken bij wijzigingen in de cesuur na de afname. Hoe kan de examencommissie anders garant staan voor het eindniveau?

Een werkwijze die aan beide verantwoordelijkheden recht doet is de volgende. De examiner overlegt een voorstel tot aanpassing met inhoudelijke onderbouwing aan de examencommissie. De examencommissie kijkt of het examen ook na de aanpassing aan de kwaliteitseisen voldoet. De wijziging wordt geaccordeerd en door de examencommissie vastgelegd in de notulen. Indien vaak aanpassingen van de cesuur voorkomen, is het verstandig om nader te onderzoeken of hier redenen voor aan te wijzen zijn.

Bij een digitaal examen dat is samengesteld uit een opgavenbank waarvan de opgaven door verschillende constructeurs gemaakt zijn, is er geen examiner die het examen afneemt maar doet de computer dit. In deze situatie lijkt met name de examencommissie aangewezen om te beslissen over aanpassingen van de cesuur of het laten vervallen van vragen.

De verantwoordelijkheden voor de gang van zaken bij dit type examens kunnen worden vastgelegd in de OER.

Het geven van bindend studieadvies

Het geven van bindende studieadvies is geen wettelijke taak. Als de examencommissie binnen de instelling is aangewezen om deze adviezen te geven dan volgt de examencommissie hier de regels van de instelling. In de praktijk blijkt het geven van een bindend studieadvies, vooral als het negatief is, regelmatig tot bezwaren te leiden. De onderbouwing van dit advies

zal vaak op de proef gesteld worden en dat is begrijpelijk, want de persoonlijke gevolgen voor de student zijn groot.

5.4 Borgen van de kwaliteit van examinering

Hoe borgt een examencommissie de kwaliteit van de tentamens en examens? Al eerder kwam ter sprake dat er een omslag heeft plaatsgevonden van een reactieve rol, waarbij de examencommissie pas in actie kwam als de tentamens waren afgerond, naar een proactieve rol waarbij de examencommissie zich vanaf het opstellen van de OER met de kwaliteit bezighoudt. De examencommissie zal dus proactief risico's moeten signaleren die de diplomering in gevaar kunnen brengen. De examencommissie maakt deze risico's kenbaar bij de opleiding en komt met analyse en verbetervoorstellen. Als bijvoorbeeld blijkt dat een aantal examens niet het beheersingsniveau toetst dat volgens het curriculum behaald moet worden, kan dit aanleiding zijn om op dit punt voor te stellen om met een toetsmatrijs gebaseerd op taxonomie-indeling te gaan werken. De rol van de examencommissie als adviseur is vrij algemeen geaccepteerd, maar niet geheel onomstreden. Door te adviseren zou de examencommissie zich aan het beleid committeren en daarmee haar onafhankelijke rol verliezen.

Een proactieve houding van de examencommissie kan de volgende stappen inhouden:

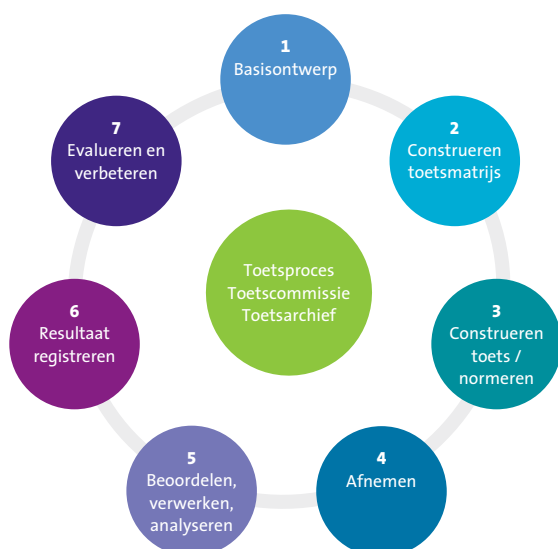
- 1 Operationaliseren van de kwaliteit van examinering
- 2 Risicoanalyse
- 3 Onderzoeken en analyseren
- 4 Adviseren van het bevoegd gezag en management

5.4.1 Operationaliseren van de kwaliteit van examinering

Om het eindniveau en de kwaliteit van examinering te kunnen borgen, is een model nodig dat examenkwaliteit operationaliseert. We zouden ook kunnen zeggen dat de examencommissie een visie op kwaliteit nodig heeft. Door een model te hanteren wordt het begrip kwaliteit afgebakend en worden hoofd- en bijzaken onderscheiden.

Van de modellen die in het hoger onderwijs gebruikt worden, lijkt een aantal heel geschikt als kader voor het operationaliseren van kwaliteit:

- De Expertgroep BKE/SKE (2013) gebruikt het model van de toetscyclus om de kwaliteit van toetsen en de competenties van de examinatoren en examencommissie in beeld te brengen. Dit model brengt de stappen van de toetscyclus in beeld en geeft zicht op het toetsproces van constructie tot evaluatie (zie figuur 5.2).
- De kwaliteitspiramide van Sluijsmans et al. (2012) benoemt als aspecten van kwaliteit: toetsitems, toetsen, examenprogramma, toetsbeleid, examenorganisatie en deskundigheid.
- Specifiek voor het afstudeerprogramma geeft de Expertgroep Protocol (2014) een lijst van twaalf vragen die de opleiding moet kunnen beantwoorden om de kwaliteit van het afstudeerprogramma te kunnen verantwoorden.



Figuur 5.2: Stappen in de toetscyclus (bron: Expertgroep BKE/SKE)

Door voor elke stap van de toetscyclus criteria op te stellen en deze te verwerken in een checklist met een scoreschaal wordt de kwaliteit per stap meetbaar. Een norm toevoegen aan de scoreschaal per stap maakt het mogelijk de kwaliteit van examens te borgen binnen vooraf gestelde grenzen. Op dit moment zijn er binnen het hoger onderwijs geen algemeen geaccepteerde criteria, scoreschalen en normen om de kwaliteit van een examen per stap van de toetscyclus te bepalen en te borgen. Er zijn wel globaal gedeelde uitgangspunten die het mogelijk maken om de kwaliteit op een zinvolle manier te onderzoeken en aanbevelingen te doen voor verbetering. Aangezien het borgen van de kwaliteit van examinering een wettelijke taak is, zal de examencommissie toch een uitspraak moeten doen over kwaliteit. Op dit moment zien we nog dat veel examencommissies impliciet een uitspraak doen met een omgekeerde bewijsvoering: als er geen signalen zijn dat er iets aan de hand is, dan is de kwaliteit blijkbaar in orde. Vanuit het gegeven dat de examencommissies niet naar een algemene standaard kunnen verwijzen, is dit begrijpelijk.

We zouden kunnen stellen dat als elk afzonderlijk examenonderdeel voldoet aan de eisen, een diploma met vertrouwen kan worden uitgereikt als de student aan alle voorwaarden voldoet. Borgen zou zich vanuit deze visie kunnen beperken tot de toetscyclus. Toch blijkt het juist vanuit een proactieve houding zinvol dat de examencommissie ook andere aspecten van kwaliteit in het examenkwaliteitsmodel betreft. Deze aspecten zijn bijvoorbeeld: het examenprogramma en het examenreglement (beide opgenomen in de OER), de examenorganisatie, het toetsbeleid en de toetsdeskundigheid. Verbeteringen op deze gebieden die door de examencommissie bewerkstelligd kunnen worden, hebben een groter effect dan een advies voor een specifiek examen.

We lichten het belang van deze aspecten voor de taak van de examencommissie hierna nader toe.

Examenprogramma

De WHW bepaalt onder meer dat de OER informatie moet bevatten over de vorm van de tentamens, de inhoud, de examenmomenten en de volgorde (WHW, artikel 7.13). De naamgeving verschilt per instelling, maar hier noemen we het overzicht van alle tentamens, papers, essays, stages, scripties, projecten, portfolio's en overige opdrachten die studiepunten opleveren het examenprogramma. Ter informatie: de Expertgroep Protocol (2014) spreekt van een afstudeerprogramma als het geheel van toetsonderdelen waarin de student laat zien te beschikken over de eindkwalificaties van de opleiding.

Het examenprogramma geeft aan welke examens met welke toetsvorm en welke inhoud wanneer gedaan worden. Het examenprogramma is idealiter het basisplan op basis waarvan de examinatoren verder de individuele toetsen uitwerken. In de praktijk kan het omgekeerde het geval zijn: dat wat examinatoren in de praktijk doen, wordt genoteerd in het examenprogramma. Deze laatste werkwijze geeft geen garantie voor een voldoende samenhangend programma. Het geheel aan examens moet inhoudelijk de opleidingsdoelen dekken en ook voldoen aan het te bereiken niveau volgens de Dublin-descriptoren, het EQF (European Qualifications Framework) en het NLKK (Nederlands kwalificatiekader). Het onderzoekend vermogen wordt in het advies van de Expertgroep Protocol apart genoemd als aandachtspunt van het eindniveau.

Het is belangrijk dat de juiste toetsvormen bij de leerdoelen worden gekozen om recht te doen aan de diversiteit van de vaardigheden in het toekomstige beroep. Een mix van toetsvormen (zie ook hoofdstuk 3) helpt ook voorkomen dat studenten benadeeld worden omdat ze minder goed een bepaalde toetsvorm kunnen hanteren terwijl ze mogelijk wel aan de opleidingseisen voldoen. De kwaliteit van de individuele examens is dus alleen te beoordelen in samenhang met het examenprogramma.

De examencommissie stelt het examenprogramma niet zelf op. Ze bekijkt wel alle bovengenoemde aspecten omdat deze direct raken aan het eindniveau en aan de kwaliteit van examinering. Als dit niet voldoet dan adviseert ze de opleiding het examenprogramma op onderdelen aan te passen.

Het examenprogramma heeft daarmee de volgende functies die van belang zijn voor de examencommissie:

- het examenprogramma is een samenhangend geheel van afzonderlijke examens dat het behalen van het eindniveau door studenten garandeert;
- het examenprogramma biedt een kader aan de examinatoren om de examens vorm te geven;
- het examenprogramma biedt een kader voor de toetscommissie bij de beoordeling van de kwaliteit van afzonderlijke examens.

Examenreglement

Het geheel aan regels rond de organisatie en afname van examens waar studenten en examinatoren zich aan moeten houden, wordt in het algemeen het examenreglement genoemd. Het examenreglement is een apart document of een onderdeel van de OER. De WHW bepaalt dat de OER regels aangeeft omtrent de afname en beoordeling van de tentamens (artikel 7.13). In de WHW staat dat de examencommissie moet werken binnen de regels van de OER. Daar waar nadere uitwerking van de regels in de OER nodig is voor de afname en beoordeling van examens kan de examencommissie deze regels vaststellen (artikel 7.12b). Daar waar het examenreglement niet voldoet, kan ze verbeteringen voorstellen. Herziening van de OER gebeurt in het algemeen eens per jaar door de opleidingsdirectie waarbij de examencommissie adviseert over gewenste en noodzakelijke aanpassingen.

De examencommissie handelt binnen de regels van de OER en is dus gebaat bij heldere en volledige regelgeving. Omdat de examencommissie door haar rol van beslissers veel met de OER werkt, is zij bij uitstek een goede adviseur bij aanpassingen van de OER.

Examenorganisatie

Is de examenorganisatie zodanig ingericht dat taken en verantwoordelijkheden duidelijk zijn? De inrichting van de examenorganisatie is niet de verantwoordelijkheid van de examencommissie. Als er taken en verantwoordelijkheden onduidelijk zijn, kan dit wel de kwaliteit van de examinering raken. Door het proactief onderzoeken van de kwaliteit kunnen lacunes in de processen, taken en verantwoordelijkheden duidelijk worden. Hierover kan de examencommissie gericht adviseren. Als bijvoorbeeld uit een inventarisatie door de examencommissie blijkt dat er maar bij weinig meerkeuzetoetsen afnamegegevens

geanalyseerd worden (leidend tot aanpassing van de vragen) dan kan de examencommissie adviseren dat op dit punt het beleid wordt aangepast. Of als uit een steekproef blijkt dat dossiers niet volledig zijn, kan de examencommissie aanbevelen om afspraken rond aanleveren van het gemaakte werk te herzien.

Toetsbeleid

Het toetsbeleid, dat wordt opgesteld door het instellingsbestuur of de opleidingsdirectie, bevat de visie op toetsing en examinering en het op deze visie gebaseerde beleid met richtlijnen voor de uitvoering door alle betrokkenen. Het toetsbeleid moet passen binnen de wetgeving, de NVAO-richtlijnen en de afspraken die gemaakt zijn met andere instellingen, zoals voor het hbo de afspraken over externe validering. Vanuit haar deskundigheid is de examencommissie in staat om knelpunten die in het toetsbeleid kunnen ontstaan rond examinering te signaleren en hierover te adviseren.

Toetsdeskundigheid

Bij het benoemen van de examinatoren stelt de examencommissie zich op de hoogte of de benodigde deskundigheid aanwezig is. Uit het onderzoek van de toetscommissie naar de kwaliteit van de examens kan een algemeen beeld naar voren komen waarmee gerichte adviezen over deskundigheidsbevordering gegeven kunnen worden.

In het hbo zullen de basiskwalificatie examinering en seniorkwalificatie examinering (BKE/SKE) in de behoefte aan toetsdeskundigheid en toetsbekwaamheid moeten gaan voorzien.

5.4.2 Risicoanalyse

Als bij elke stap in de toetscyclus kwaliteitscriteria, een scoringssysteem en een norm zijn geformuleerd in de vorm van een checklist, kan de examencommissie de kwaliteit van de examinering in de eigen opleiding onderzoeken. Een manier om te bepalen waar de aandacht van de examencommissie primair naar uit moet gaan, is om risico's voor de kwaliteit van examinering in kaart te brengen en te ordenen. De risico's met de grootste impact op de kwaliteit en de grootste waarschijnlijkheid dat er ook daadwerkelijk iets mis zal zijn, komen bovenaan de lijst. Op deze punten onderzoekt de examencommissie de stand van zaken als eerste. Het volgen van ontwikkelingen binnen en buiten de organisatie die de kwaliteit van de examinering kunnen beïnvloeden, is een belangrijk hulpmiddel om een goede risicoanalyse te maken.

In tabel 5.1 en 5.2 zijn risico's geformuleerd bij alle genoemde kwaliteitsaspecten. Een examencommissie kan deze concreter formuleren voor de eigen situatie.

Tabel 5.1: Kwaliteitsaspecten, risico's en onderzoeksvragen bij de toetscyclus

Kwaliteitsaspect	Risico	Onderzoeksvraag
1 Basisontwerp	De leerdoelen in het curriculum sluiten onvoldoende aan bij het einddoel van de opleiding. Het eindniveau wordt niet gehaald of onvoldoende onderbouwd.	Komen alle doelen van de opleiding op het juiste niveau in opleiding en examinering voor?
2 Construeren toetsmatrijs	De opgaven en opdrachten in het examen zijn niet representatief voor de leerdoelen.	Gegeven de leerdoelen, zijn vakgenoten het eens dat de opgaven de leerdoelen dekken en op niveau zijn? Wordt een matrijs gebruikt?
3 Construeren toets en het vaststellen van de cesuur	De antwoorden (of het gedrag) van de kandidaat zijn onvoldoende objectief te beoordelen gegeven het beschreven antwoordmodel en de beoordelingsinstructie. De scores/cijfers zijn niet betrouwbaar vast te stellen. De cesuur is niet aanwezig, niet transparant of correspondeert niet met het juiste niveau.	Voldoen de opgaven en antwoordmodellen aan de vuistregels voor constructie? Is de cesuur helder? Is er een verantwoording van de cesuur?
4 Afname van de toets	De afname geeft niet alle kandidaten gelijke kansen voor het goed beantwoorden van de vragen of opdrachten. Er is gelegenheid tot fraude, plagiaat of meeliften.	Zijn de afnamecondities voor alle kandidaten gelijk? Worden regels tegen fraude, plagiaat of meeliften gehandhaafd?
5 Beoordelen, verwerken, analyseren	Beoordelaars zijn onvoldoende objectief. Opgaven hebben onvoldoende onderscheidend vermogen.	Komen beoordelaars tot gelijke beoordeling bij dezelfde prestatie? Is er afstemming geweest tussen beoordelaars? Blijkt uit een toets- en itemanalyse dat de toets voldoende betrouwbaar is?
6 Resultaat registreren	De geregistreerde resultaten zijn niet correct of niet controleerbaar.	Zijn bij een steekproefsgewijze controle de geregistreerde resultaten correct en de dossiers volledig?
7 Evalueren en verbeteren	Er is geen systematiek voor evaluatie en kansen op verbetering worden niet benut. Kwaliteit is niet geborgd.	Wordt het proces geëvalueerd door examinatoren? Worden verbeterplannen doorgevoerd?

Tabel 5.2: Kwaliteitsaspecten, risico's en onderzoeksvragen bij de algemene kwaliteitsaspecten

Kwaliteitsaspect	Risico	Onderzoeksvraag
Examenprogramma	Het eindniveau wordt onvoldoende behaald of onvoldoende onderbouwd.	Zijn de prestaties van afgestudeerden conform de leerdoelen?
	Studenten worden op verkeerde gronden afgewezen.	Functioneren alumni op het gewenste niveau? Zijn de doelen correct verwerkt in curriculum en examenprogramma?
Examenreglement	De OER geeft onvoldoende duidelijkheid om beslissingen te nemen.	Zijn actuele ontwikkelingen verwerkt in de OER? Zijn er individuele gevallen waarbij de OER tekortschiet?
Examenorganisatie	Taken en verantwoordelijkheden zijn niet duidelijk toebedeeld.	Zijn taken, verantwoordelijkheden en processen beschreven?
Toetsbeleid	Taken worden niet of niet volgens de afgesproken procedure uitgevoerd.	Werken betrokkenen volgens de beschreven procedures?
	Betrokkenen voelen zich niet verantwoordelijk om de gewenste kwaliteit te leveren.	Voelen betrokkenen zich verantwoordelijk?
Toetsdeskundigheid	Het werk van examinatoren of andere betrokkenen bij het toetsproces voldoet niet aan de kwaliteitseisen.	Voldoen betrokkenen aan BKE/SKE-eisen?

5.4.3 Onderzoeken en analyseren

In tabel 5.1 en 5.2 is per kwaliteitsaspect en risico al een algemene onderzoeksvraag geformuleerd. Een examencommissie kan, afhankelijk van de eigen situatie, deze algemene vragen nader specificeren en op basis van de belangrijkste risico's een onderzoeksplan opstellen. Budget en tijd zijn belangrijke factoren die de reikwijdte en diepgang van het onderzoek bepalen. Het is aan te bevelen om in een jaarcyclus te werken of zelfs over meerdere jaren te plannen. Dit geeft tijd om grondig onderzoek naar de kwaliteit te doen. Het onderzoeksplan en de facilitering wordt met het bestuur en management afgestemd. Met instemming van het bestuur en management staat de examencommissie sterker in de uitvoering van het onderzoeksplan.

Het onderzoek kan door de toetscommissie worden uitgevoerd, door examencommissieleden zelf, door de afdeling kwaliteitszorg of door externe partijen. Bij het onderzoeken van de kwaliteit zal de examencommissie behoefte hebben aan gegevens over de uitvoering van de examinering. Veel examencommissies hebben deze gegevens niet bij de hand, opzoeken kost veel tijd en is soms achteraf niet mogelijk. Het examenbureau lijkt de plaats waar gegevens kunnen worden verzameld en opgeslagen. De examencommissie kan overleggen met de directie om te bepalen welke gegevens verzameld moeten worden. Bij de te verzamelen gegevens kunnen we denken aan:

- Is een toetsmatrijs aanwezig?
- Is het examen bij alle studenten door twee examinatoren beoordeeld?
- Is een toets- en itemanalyse gemaakt?

- Is het proces door examinatoren geëvalueerd?
- Is er een klacht over het tentamen geweest?
- Welke controles zijn uitgevoerd bij de registratie van de cijfers?

De onderzoeken kunnen een projectmatige of thematische aanpak of een meer steekproefsgewijze vorm hebben. We geven een voorbeeld van een onderzoeksplan voor twee jaar.

Eerste jaar van het onderzoeksplan

- Een aselecte steekproef van 20% van de examens uit elk leerjaar screenen op de kwaliteit van constructie.
- Screenen van twee examens die laag scoorden in het studenttevredenheidsonderzoek.
- Onderzoek naar de risico's van fraude bij digitale afname van examens.
- Inventarisatie van individuele en alternatieve trajecten.
- Inventarisatie van de eigen deskundigheid in de examencommissie.
- Onderzoek naar de oorzaken van extreem lage en hoge scores op examens.

Tweede jaar van het onderzoeksplan

- Een nieuwe aselecte steekproef van 20% van de examens uit elk leerjaar screenen.
- Kwalitatief onderzoek naar de maatregelen voor objectiviteit van stagebeoordelingen.
- Onderzoek naar beoordelaarsovereenstemming bij de beoordeling van eindwerkstukken.
- Inventarisatie van het gebruik van toets- en itemanalyse bij theorie-examens.
- Vervolgonderzoek naar aanleiding van voorgaand jaar, bijvoorbeeld onderzoek naar de effectiviteit van de maatregelen om fraude bij digitale afname te voorkomen.

5.4.4 Adviseren van het bevoegd gezag en management

Op basis van de resultaten van onderzoek analyseert de examencommissie wat de oorzaken zijn van de sterke en zwakke punten. Verbetervoorstellen worden geformuleerd en besproken met bestuur en management. Door de voorgestelde werkwijze waarbij de belangrijkste risico's onderzocht worden en een grondig onderzoek ten grondslag ligt aan de adviezen van de examencommissie, zullen de voorgestelde maatregelen de kwaliteit van de examinering gunstig beïnvloeden. Als de kwaliteit van de examinering volgens de onderzoeken als geheel voldoende naar voren komt, kan de examencommissie met gefundeerd vertrouwen de getuigschriften ondertekenen.

5.5 Samenstelling van de examencommissie

De examencommissie moet volgens de WHW onafhankelijk zijn. De j82 heeft op basis hiervan afgesproken, zoals eerder opgemerkt, dat het management geen deel uitmaakt van de examencommissie. Dit is inmiddels ook in de wetgeving vastgelegd. De wet zegt verder dat ten minste één docent van de opleiding zitting heeft in de examencommissie en dat er een extern lid moet zijn.

Behalve deze regels zijn er geen harde eisen aan de samenstelling van de examencommissie. Wel kunnen we redenerend vanuit de taak als borger en beslisser nadenken over de samenstelling.

Vanuit de rol als borger van kwaliteit is betrokkenheid en bekendheid met de opleiding even belangrijk als afstand tot de dagelijkse gang van zaken in het onderwijs. Binnen de organisatie kunnen personen van verschillende geledingen benoemd worden om verschillende deskundigheid binnen te halen, aangevuld met één of meer externe leden. Een externe kan iemand zijn uit een andere examencommissie van dezelfde of van een andere instelling. Ook kan een extern lid met juridische expertise of toetsdeskundigheid benoemd worden.

Het aantal leden in de commissie is in de praktijk wisselend en afhankelijk van het aantal studenten en opleidingen dat de commissie onder haar hoede heeft. Vanuit het oogpunt van besluitvorming is een kleinere commissie efficiënter, maar bij te weinig draagvlak van beslissingen kan dit juist een negatief effect hebben. Een te groot aantal leden draagt niet meer bij aan essentiële deskundigheid en onafhankelijkheid. De te grote commissie loopt het risico om een log orgaan te worden zonder voeling met de praktijk van het werk van de examencommissie. Beslissingen worden dan door een klein dagelijks bestuur of in de wandelgangen genomen. Bij een goede organisatie kunnen brede examencommissies ook voordelen hebben en hoeft niet iedere opleiding het wiel opnieuw uit te vinden.

De examencommissie heeft verschillende soorten deskundigheid nodig bij de werkzaamheden. Het heeft dan ook de voorkeur als de leden verschillende deskundigheid inbrengen: kennis van de opleiding en het te behalen eindniveau, kennis van toetsing, een SKE-kwalificatie, juridische expertise en kennis van kwaliteitszorg. Maar ook ervaring als examencommissielid of bestuurlijke ervaring kan de commissie ondersteunen. De commissie dient als geheel voldoende deskundig te zijn, en waar zij de deskundigheid niet zelf in huis heeft, kan ze zich ontwikkelen en scholen. Voor specifieke expertise kunnen adviseurs uitgenodigd worden.

Een aantal van vier tot zeven leden blijkt in de praktijk heel werkbaar, ook al zullen er vast commissies zijn die met drie of acht leden even goed werk leveren. Een aardig gedachte-experiment is het volgende: stel bij elk lid van de commissie voor dat hij/zij een aantal maanden met verlof gaat. Moet het commissielid dan vervangen worden? Bij elk lid dat niet vervangen hoeft te worden, kan de commissie zich afvragen of het effectief is dat hij of zij lid is. Aan de andere kant kunnen we ook de vraag stellen: is er buiten de commissie een andere, waardevolle kijk op de kwaliteit van examinering? Zo ja, hoe kan deze visie of persoon dan betrokken worden bij de commissie? Als lid, adviseur of verzorger van een cursus?

Bij het benoemen van nieuwe leden vraagt het bestuur de zittende leden om advies of ten minste hun mening (WHW, artikel 7.12a). Het bestuur kan van een nieuw lid vragen dat hij/zij de uitgedragen visie van de instelling op het gebied van onderwijs en examinering ondersteunt. De zittende leden hebben behoefte aan deskundigheid en een goede mix van afstand tot en betrokkenheid bij de opleiding. Een benoeming van een nieuw lid vindt meestal plaats voor een periode van drie jaar. Aangezien het enkele jaren duurt om deskundigheid en ervaring op te bouwen, lijkt dat een minimum. Het is niet wenselijk dat de commissie veel van samenstelling wisselt. Aan de andere kant is het vanuit het oogpunt van onafhankelijkheid ook goed dat er wel wisselingen zijn.

5.6 Regeling examencommissie en huishoudelijk reglement

Een centraal opgesteld reglement voor de examencommissies geeft binnen de instelling de kaders aan voor de examencommissies. Dit bevordert dat examencommissies in gelijke gevallen gelijk zullen handelen.

De examencommissie stelt zelf ook regels op over de uitvoering van de eigen taken en invulling van bevoegdheden. Dit wordt meestal het huishoudelijk reglement genoemd. In het huishoudelijk reglement vinden we onder andere afspraken over:

- Mandaten
- Dagelijks bestuur
- Rol van de ambtelijk secretaris
- Werkwijze bij brieven aan studenten naar aanleiding van klachten, verzoeken of fraude
- Jaarplan examencommissie

Mandaten

Het vastleggen van mandaten voor commissieleden om beslissingen te nemen namens de hele commissie verhoogt de slagvaardigheid en effectiviteit van de commissie. Hierbij valt te denken

aan mandaten voor het verlenen van vrijstellingen, het aanwijzen van examinatoren, het tekenen van getuigschriften en het geven van bindende studieadviezen. Voor alle gemandateerde acties geldt dat er een vastgelegde procedure moet zijn en dat het mandaat alleen geldt binnen deze afgesproken procedure. Het mandaat wordt bij voorkeur aan twee leden verleend die alleen unaniem mogen beslissen. Alle besluiten dienen aan de examencommissie gerapporteerd te worden en komen als bijlage bij de notulen. Hiermee zijn alle gemandateerde besluiten duidelijk als commissiebesluiten traceerbaar en heeft de commissie zicht op hoe het mandaat wordt gebruikt en kan zij waar nodig toelichting vragen. Een mandaat kan ook worden gegeven om te handelen in onvoorziene situaties die kunnen optreden tijdens examenweken. Met de huidige communicatiemiddelen is het bijna altijd mogelijk om ook in acute situaties medecommissieleden te raadplegen om ten minste met twee leden, of met zo breed mogelijke steun, een beslissing te nemen.

Dagelijks bestuur

Indien een tweetal leden meerdere mandaten heeft om veel voorkomende besluiten te nemen, zullen ze als dagelijks bestuur functioneren. Dit past geheel binnen een visie van slagvaardigheid. Een risico hierbij kan zijn dat de andere leden van de examencommissie zich weinig actief betrokken voelen bij de gang van zaken en besluitvorming. Als het dagelijks bestuur dan ook met het management regelmatig overleg voert, wordt de commissie gepasseerd en komt onafhankelijke besluitvorming in het geding.

Rol van de ambtelijk secretaris

Meestal is de ambtelijk secretaris niet benoemd als lid van de examencommissie in de zin van meebeslissend lid. Vaak maakt de secretaris wel onderdeel uit van het dagelijks bestuur. Dit kan echter per instelling verschillen en er zijn geen vaste regels voor. Een ambtelijk secretaris die meerdere commissies ondersteunt, heeft een sterke invloed op het goed volgen van afgesproken procedures en gelijke werkwijze binnen de commissies. Het doorvoeren van nieuwe regelgeving wordt daardoor vergemakkelijkt. De ambtelijk secretaris is dan ook vaak een specialist op gebied van de OER en wetgeving. Andere zaken aangaande de besluitvorming die in het huishoudelijk reglement genoemd kunnen worden, zijn het quorum voor besluitvorming en de wijze van besluitvorming.

Werkwijze bij brieven aan studenten

Bij meerdere taken van de examencommissie staat het nemen van een beslissing over de student centraal. De beslissingen van de examencommissie raken de toekomstplannen van de student en ook zijn financiën als hij bijvoorbeeld studiefinanciering zou mislopen. Een zorgvuldige beslissing door de examencommissie is vaak van groot belang voor de betrokkenen. De beslissing zal meestal schriftelijk gecommuniceerd worden. Het is verstandig om hierbij standaard formats te hanteren die de besluitvorming kunnen ondersteunen. Elk schrijven naar aanleiding van een verzoek, klacht of beslissing bij fraude zou de volgende onderdelen moeten bevatten:

- Het verzoek of de klacht van de student, of de omstandigheden en waarnemingen die tot een vermoeden van fraude hebben geleid.
- Een toelichting op basis van welke regel uit de OER of van de WHW het behandelen of beslissen van deze zaak onder verantwoordelijkheid van de examencommissie valt.
- De informatie die de commissie heeft gebruikt bij de behandeling. Welke documenten zijn ingezien, met wie is gesproken? Hierbij wordt tevens de chronologische volgorde aangegeven: de datum dat de klacht is binnengekomen, data van gesprekken en data van correspondentie rond de behandeling. Zonder volledige informatie kan geen goed besluit genomen worden en dat wrekt zich bij het formuleren van argumenten bij een beslissing. Hoor en wederhoor toepassen en daadwerkelijk spreken met de student is bij ernstige klachten en verdenking van fraude aan te bevelen in deze onderzoeksfase.

- De afweging die de commissie heeft gemaakt. Argumenten worden opgesomd en belangen van de betrokken partijen gewogen. Wetsartikelen of regels uit de OER worden genoemd. Het weerleggen van argumenten kan hier deel van uitmaken. Hieruit moet duidelijk worden op basis van welk argument of belang de beslissing is genomen. Dit blijkt vaak een lastig te formuleren punt omdat hier kleur bekend moet worden. Een voorbeeld van een formulering waarin niet echt stelling wordt genomen: 'We hebben jammer genoeg moeten besluiten om (...).' Beter is: 'Het belang om iedere student onder gelijke omstandigheden examen te laten doen en daarmee het niveau en de kwaliteit te kunnen borgen, weegt voor de commissie zwaarder dan uw argument voor een aangepast examen.'
- Als toepassing van de reglementen tot onaanvaardbare consequenties leidt, kan de examencommissie anders besluiten, rekening houdend met het algemene rechtsbeginsel van redelijkheid en billijkheid.
- De uitspraak van de commissie. De uitspraak sluit aan op de klacht of het verzoek en is een rechtstreeks antwoord hierop. Naast de directe reactie op het verzoek kunnen aanvullende uitspraken gedaan worden. De klacht over de beoordeling wordt bijvoorbeeld niet gehonoreerd maar de student wordt de mogelijkheid geboden een extra herkansing te doen.
- Opvolging van de uitspraak. Hoe wordt de uitspraak geëffectueerd? Wie doet wat en binnen welke termijn?
- Beroep tegen de uitspraak. De mogelijkheid aangeven om in beroep te gaan tegen de beslissing bij het college van bestuur. Bij wie en binnen welke termijn?

Het voorgaande klinkt voor sommige verzoeken wellicht formeel en juridisch zwaar aangezet. Elke commissie kan hierin wel een eigen toon en stijl vinden, maar gezien de wettelijke taken is de examencommissie misschien wel genooddaakt formeel en op afstand van de student te blijven en juist vanuit die rol de student te dienen door te waken over de examens en procedures. Bij moeilijke beslissingen kan de examencommissie de juridische afdeling of een jurist om advies vragen bij de formulering of onderbouwing van een besluit. Om de schijn van willekeur bij besluiten te voorkomen, is het een regel dat deze altijd door twee personen genomen worden en dat de besluiten vastgelegd worden in de notulen.

Voor wie verder wil lezen: het tijdschrift 'Examens' van de NVE bespreekt in elk nummer een casus van bezwaar, verzoek of fraude. Ga naar www.nvexamens.nl voor het digitale archief van het tijdschrift 'Examens'.

Jaarplan examencommissie

Gezien het aantal taken en de complexiteit van de taken is het nodig dat de examencommissie een planning maakt, bijvoorbeeld per jaar, van haar activiteiten. Het eerder genoemde onderzoeksplan maakt daar deel van uit. De inhoud en prioriteiten van het jaarplan kan met bestuur en management worden afgestemd. Het faciliteren van de uitvoering van het jaarplan kan dan ook meteen aan bod komen. De effectiviteit van de commissie wordt vergroot als de verantwoordelijkheid voor het uitvoeren en/of monitoren van de voortgang van onderdelen van het jaarplan onder de commissieleden verdeeld wordt. Uit het jaarplan kunnen de vergaderfrequentie en agendapunten worden afgeleid.

5.7 Ten slotte

De ontwikkelingen van de laatste jaren sluiten aan bij de steeds dringender geuite wens om de kwaliteit van examens binnen de onderwijsinstellingen in het hoger onderwijs en de eindkwalificaties van hun afgestudeerden adequaat te borgen. De wetgeving en het accreditatiekader zijn aangepast en veel examencommissies voelen zich steeds meer thuis in hun nieuwe rol. In dit hoofdstuk hebben we de examencommissie handreikingen gegeven om deze borging gestalte te geven.

6 De kwaliteit van toetsen en examens

6 De kwaliteit van toetsen en examens

Piet Sanders

In Nederland en Vlaanderen wordt de kwaliteit van opleidingen in het hoger onderwijs beoordeeld door de Nederlands-Vlaamse Accreditatieorganisatie (NVAO). We besteden hier aandacht aan het accreditatiestelsel van de NVAO omdat een van de standaarden van het beoordelingskader om de basiskwaliteit van opleidingen te beoordelen, de kwaliteit van toetsen en examens is. Paragraaf 6.1 gaat nader in op de beoordelingskaders van het NVAO-accreditatiestelsel van het hoger onderwijs. De bespreking maakt op verschillende plaatsen gebruik van teksten die letterlijk ontleend zijn aan de publicatie 'Beoordelingskaders accreditatiestelsel hoger onderwijs' (NVAO, 2014).

Voor de beoordeling van de kwaliteit van toetsen en examens wordt gebruikgemaakt van beoordelingssystemen. In Nederland zijn de twee meest gebruikte beoordelingssystemen het 'Beoordelingssysteem voor de kwaliteit van tests' van de Commissie Testaangelegenheden Nederland (COTAN) en de 'Regeling standaarden examenkwaliteit mbo 2012' die de Inspectie van het Onderwijs gebruikt om de examens in het middelbaar beroepsonderwijs te beoordelen. Voor een beschrijving van deze beoordelingssystemen verwijzen we naar hoofdstuk 10 van Toetsen op School (2013), zie www.toetsenopschool.nl. Aangezien geen van de twee beoordelingssystemen geschikt is om toetsen en examens uit het voortgezet en hoger onderwijs te beoordelen, heeft het Research Center voor Examinering en Certificering (RCEC) een nieuw beoordelingssysteem ontwikkeld dat in paragraaf 6.2 van dit hoofdstuk gepresenteerd wordt.

Tot slot gaan we in paragraaf 6.3 in op wat de relatie tussen het NVAO-beoordelingskader en het RCEC-beoordelingssysteem voor de kwaliteit van toetsen zou kunnen zijn. De conclusie is dat het RCEC-beoordelingssysteem zeer goed ingezet kan worden bij standaard 3 van het NVAO-beoordelingskader die betrekking heeft op toetsing.

6.1 NVAO-accreditatie van opleidingen in het hoger onderwijs

Nadat in 1999 meer dan 30 Europese landen besloten om in het hoger onderwijs de bachelor-masterstructuur in te voeren, hebben Nederland en Vlaanderen in 2002 een accreditatiestelsel ingevoerd dat beschreven wordt in de NVAO-publicatie 'Beoordelingskaders accreditatiestelsel hoger onderwijs' (2014). De NVAO is de organisatie die de basiskwaliteit van bestaande opleidingen beoordeelt en nieuwe opleidingen in het hoger onderwijs toetst. Indien een opleiding niet aan de kwaliteitseisen voldoet dan heeft dat gevolgen voor zowel de onderwijsinstelling als de studenten. Die gevolgen betreffen de financiering van de opleiding door de overheid, het verlenen van wettelijk erkende diploma's en het recht van studenten op studiefinanciering. Onderwijsinstellingen voor hoger onderwijs dienen zelf een aanvraag in te dienen bij de NVAO voor accreditatie en toetsing. Niet alleen afzonderlijke opleidingen maar ook verwante opleidingen van dezelfde onderwijsinstelling of verschillende onderwijsinstellingen kunnen gezamenlijk geaccrediteerd worden (clustervisitatie). Geaccrediteerde opleidingen staan vermeld in het Centraal Register Hoger Onderwijs (CROHO) en de accreditatie geldt voor zes jaar.

Beoordelingskaders NVAO-accreditatie

Het accreditatiestelsel van de NVAO bestaat uit zes beoordelingskaders:

- 1 Instellingstoets kwaliteitszorg.
- 2 Beperkte opleidingsbeoordeling (BOB) voor opleidingen die positief beoordeeld zijn op de instellingstoets kwaliteitszorg.
- 3 Uitgebreide opleidingsbeoordeling (UOB) voor opleidingen die negatief beoordeeld zijn op de instellingstoets kwaliteitszorg.
- 4 Beperkte toets nieuwe opleidingen (TNO) voor opleidingen die positief beoordeeld zijn op de instellingstoets kwaliteitszorg.
- 5 Uitgebreide toets nieuwe opleidingen (TNO) voor opleidingen die negatief beoordeeld zijn op de instellingstoets kwaliteitszorg.
- 6 Bijzonder kenmerk.

In dit hoofdstuk bespreken we het beoordelingskader van (1) de instellingstoets kwaliteitszorg en (2) de beperkte opleidingsbeoordeling. Voor informatie over de andere beoordelingskaders verwijzen we naar de publicatie ‘Beoordelingskaders accreditatiestelsel hoger onderwijs’ (NVAO, 2014).

6.1.1 Beoordelingskader instellingstoets kwaliteitszorg

Bij de instellingstoets kwaliteitszorg gaat het om vijf vragen:

- 1 Wat is de visie van de instelling op de kwaliteit van haar onderwijs?
- 2 Hoe wil de instelling deze visie realiseren?
- 3 Hoe meet de instelling in hoeverre de visie wordt gerealiseerd?
- 4 Hoe werkt de instelling aan verbetering?
- 5 Wie is waarvoor verantwoordelijk?

De vijf vragen zijn vertaald in vijf standaarden. Een auditcommissie, bestaande uit ten minste vier leden, geeft over elk van de vijf vragen een gewogen en gemotiveerd oordeel op een driepuntsschaal: voldoet wel, voldoet niet of voldoet ten dele aan de standaard. Vervolgens geeft de auditcommissie een gemotiveerd eindoordeel over de vraag of een instelling ‘in control’ is over de kwaliteit van haar opleidingen. Ook dat oordeel vindt plaats op basis van een driepuntsschaal: positief, negatief of positief onder voorwaarden. Voor meer informatie over dit beoordelingskader verwijzen we naar ‘Beoordelingskaders accreditatiestelsel hoger onderwijs’ (NVAO, 2014).

6.1.2 Beoordelingskader beperkte opleidingsbeoordeling

Bij de beperkte opleidingsbeoordeling gaat het om drie vragen:

- 1 Wat beoogt de opleiding?
- 2 Hoe realiseert de opleiding dit?
- 3 Worden de doelstellingen bereikt?

Standaarden beperkte opleidingsbeoordeling

De drie vragen bij het beoordelingskader zijn vertaald in vier standaarden:

Standaard 1 | De beoogde eindkwalificaties van de opleiding zijn wat betreft inhoud, niveau en oriëntatie geconcretiseerd en voldoen aan internationale eisen.

Standaard 2 | Het programma, het personeel en de opleidingsspecifieke voorzieningen maken het voor de instromende studenten mogelijk de beoogde eindkwalificaties te realiseren.

Standaard 3 | De opleiding beschikt over een adequaat systeem van toetsing.

Toelichting: De toetsen en de beoordeling zijn valide, betrouwbaar en voor studenten inzichtelijk. De examencommissie van de opleiding borgt de kwaliteit van de tentaminering en examinering.

Standaard 4 | De opleiding toont aan dat de beoogde eindkwalificaties worden gerealiseerd.

Toelichting: Het gerealiseerde niveau blijkt uit de resultaten van tussentijdse en afsluitende toetsen, de eindwerken en de wijze waarop afgestudeerden in de praktijk of in een vervolgopleiding functioneren.

Over deze vier standaarden geeft een visitatiepanel een gewogen en gemotiveerd oordeel op een vierpuntsschaal: onvoldoende, voldoende, goed of excellent. Vervolgens geeft dit panel een gemotiveerd eindoordeel over de kwaliteit van de opleiding als geheel, ook op een vierpuntsschaal. De kwaliteit van de opleiding is 'onvoldoende' indien niet voldaan wordt aan de gangbare basiskwaliteit en er op meerdere vlakken ernstige tekortkomingen zijn. De opleiding is 'voldoende' indien voldaan wordt aan de gangbare basiskwaliteit en over de volle breedte een acceptabel niveau wordt vertoond. Een opleiding is 'goed' indien de opleiding systematisch en over de volle breedte uitsteekt boven de gangbare basiskwaliteit. Een opleiding is 'excellent' indien de opleiding systematisch en over de volle breedte uitsteekt boven de gangbare basiskwaliteit en als een (inter)nationaal voorbeeld geldt.

Standaard 3 en 4 van het beoordelingskader beperkte opleidingsbeoordeling hebben betrekking op toetsing. Met name standaard 3 gaat over de kwaliteit van toetsing en wordt hierna uitvoerig besproken. Het kan niet anders dan dat hierbij een aantal passages uit de NVAO-publicatie 'Beoordelingskaders accreditatiestelsel hoger onderwijs' letterlijk overgenomen moet worden. Parafraseren van bedoelde teksten zou namelijk tot verwarring kunnen leiden.

We bespreken het beoordelingskader aan de hand van de volgende onderwerpen:

- Samenstelling van het visitatiepanel
- Clustergewijze visitatie
- Beoordelingsproces
- NVAO-besluitvorming
- Verplichte documentatie
- Beslisregels

Samenstelling van het visitatiepanel

De onderwijsinstelling stelt het visitatiepanel samen, wijst een secretaris aan en draagt het panel ter instemming voor aan de NVAO. De instelling levert, op een door de NVAO vastgestelde wijze, gegevens aan over de deskundigheden en onafhankelijkheid van de panelleden en de secretaris. De instelling kan ook opdracht geven aan een extern evaluatiebureau om het panel samen te stellen en ook in dat geval moet het panel ter instemming worden voorgedragen.

De volgende evaluatiebureaus voeren visitaties uit:

- AeQui: evaluatiebureau voor het hoger onderwijs
- ASIIN: Akkreditierungsagentur für Studiengänge der Ingenieurwissenschaften, der Informatik, der Naturwissenschaften und der Mathematik
- Certiked VBI
- Lloyd's Register Nederland
- EAPAA: European Association for Public Administration Accreditation
- FIBAA: Internationale Stiftung für Qualität bei Bachelor und Masterstudiengängen / Foundation for International Business Administration Accreditation
- Hobéon: Hobéon Certificering
- NQA: Netherlands Quality Agency
- QANU: Quality Assurance Netherlands Universities

Aan het visitatiepanel worden de volgende eisen gesteld:

- Het panel bestaat uit ten minste vier leden, onder wie ten minste twee gezaghebbende domeindeskundigen en een student.

- Het panel beschikt in zijn geheel over de volgende deskundigheden:
 - deskundigheid ten aanzien van de ontwikkelingen in het vakgebied;
 - internationale deskundigheid;
 - werkvelddeskundigheid in het voor de opleiding relevante beroepenveld (indien van toepassing);
 - ervaring met het geven en ontwikkelen van onderwijs op het desbetreffende opleidingsniveau en deskundigheid ten aanzien van de door de opleiding gehanteerde onderwijsvormen (bijvoorbeeld afstandsonderwijs, projectgestuurd onderwijs, competentiegericht onderwijs);
 - studentgebonden deskundigheid;
 - visitatie- of auditdeskundigheid.
- Het panel is onafhankelijk. De leden hebben ten minste vijf jaar geen banden gehad met de instelling waartoe de opleiding behoort.
- De voorzitter van het panel is door de NVAO getraind.
- Het panel wordt bijgestaan door een onafhankelijke, externe secretaris die getraind en gecertificeerd is door de NVAO. De secretaris is geen lid van het panel.

Clustergewijze visitatie

Met ingang van 2015 worden alle opleidingen clustergewijs gevisiteerd. Door opleidingen clustergewijs te visiteren worden oordelen van dezelfde opleidingen betrouwbaarder en beter vergelijkbaar. De invoering van clustergewijze visitatie betekent de introductie van het begrip ‘visitatiegroep’ wat een groep opleidingen is die gezamenlijk door één commissie van deskundigen beoordeeld wordt.

Beoordelingsproces

We bespreken de volgende onderdelen van het beoordelingsproces:

- Kritische reflectie
- Locatiebezoek
- Beoordelingsproces binnen het visitatiepanel
- Beoordelingsrapport

Kritische reflectie

Centraal in het accreditatieproces staat de kritische reflectie die door de opleiding ten behoeve van de beoordeling door het visitatiepanel opgesteld wordt. De kritische reflectie dient de standaarden van het kader voor de beperkte opleidingsbeoordeling te volgen, waarbij de sterke en de zwakke punten van de opleiding worden beschreven. Tevens wordt aangegeven welke verbetermaatregelen zijn genomen naar aanleiding van de vorige visitatie. De kritische reflectie is een op zichzelf staand document, dat zelfstandig te lezen is. De omvang van de kritische reflectie is maximaal 15 pagina's, exclusief bijlagen.

Locatiebezoek

Het verplichte locatiebezoek ten behoeve van de beperkte opleidingsbeoordeling duurt in beginsel een dag. Indien vergelijkbare opleidingen van een instelling tezamen worden beoordeeld, kan er een relatieve reductie van de bezoektijd plaatsvinden. Teneinde te kunnen beoordelen of de beoogde eindkwalificaties worden behaald, bestudeert het panel een selectie van eindwerken overeenkomstig de NVAO-richtlijn ‘Beoordeling eindwerken’, te vinden op www.nvao.net. In veel gevallen vormt de eindschrijving de eindtoets van de opleiding. Naast, of soms in plaats van, de eindschrijving komen ook andere eindwerken voor zoals een portfolio, een beroepsproduct, een of meerdere tentamens, een artikel, een artistieke prestatie, etc. De opleiding beschrijft hoe zij de realisatie van het eindniveau toetst. Het panel bepaalt uiteindelijk welke producten of welk ‘palet van eindprestaties’ als eindwerken van de opleiding worden beoordeeld.

De opleiding stelt het panel een complete en geanonimiseerde overzichtslijst ter beschikking van alle afstudeerders van de laatste twee studiejaar. In ieder geval is daarop vermeld: studentnummer, titel van het eindwerk, afstudeerdatum, variant en locatie van de opleiding en de behaalde resultaten in de afstudeerfase (bijvoorbeeld: cijfer beoordeling eindwerk(en), cijfer mondelinge verdediging, eindcijfer afstuderen). Het panel selecteert minimaal vijftien eindwerken uit deze lijst. Het panel beoordeelt de eindwerken voorafgaand aan het locatiebezoek. Tijdens het locatiebezoek wordt gesproken met beoordelaars/examinatoren van de opleiding om een goed inzicht te krijgen in de wijze waarop de beoordeling tot stand is gekomen.

Beoordelingsproces binnen het visitatiepanel

Het visitatiepanel geeft een oordeel over alle in het beoordelingskader opgenomen standaarden. Dat oordeel wordt gemotiveerd door een weging van de positieve en kritische elementen uit de bevindingen van het panel. Het oordeel kan zijn: onvoldoende, voldoende, goed of excellent. Vervolgens formuleert het panel een algemeen, gewogen en gemotiveerd, oordeel over de kwaliteit van de opleiding. Ook dat oordeel wordt gegeven op een vierpuntsschaal, variërend van onvoldoende tot excellent.

Beoordelingsrapport

De secretaris van het visitatiepanel stelt een beoordelingsrapport op van ongeveer 15 pagina's. Het rapport heeft als voornaamste inhoud de oordelen van het panel op de standaarden. Indien er verbeteracties mogelijk zijn, dan neemt het panel in een aparte paragraaf aanbevelingen op. De secretaris van het visitatiepanel stuurt het rapport naar het bestuur van de instelling nadat alle panelleden hebben ingestemd met de inhoud van het rapport.

NVAO-besluitvorming

Het instellingsbestuur vraagt op basis van het beoordelingsrapport accreditatie aan bij de NVAO. De NVAO kan besluiten de opleiding te accrediteren, niet te accrediteren of een herstelperiode toe te kennen.

Verplichte documentatie

Tijdens het locatiebezoek zijn de volgende documenten ter inzage: het jaarverslag van de examencommissie en de verslagen van de opleidingscommissie (indien een opleidingscommissie vereist is), toetsopgaven met bijbehorende beoordelingscriteria en normering (antwoordmodellen) en een representatieve selectie van handboeken en overig studiemateriaal. Het gaat hierbij steeds om reeds bestaande documenten.

Beslisregels

Voor de beperkte opleidingsbeoordeling geldt dat het eindoordeel over een opleiding in elk geval 'onvoldoende' is indien standaard 1, 3 of 4 als 'onvoldoende' beoordeeld wordt. Een onvoldoende bij standaard 1 kan niet leiden tot het toekennen van een herstelperiode door de NVAO. Het eindoordeel over een opleiding kan alleen 'goed' zijn indien ten minste twee standaarden als 'goed' worden beoordeeld waaronder in elk geval standaard 4. Het eindoordeel over een opleiding kan alleen 'excellent' zijn indien ten minste twee standaarden als 'excellent' worden beoordeeld waaronder in elk geval standaard 4.

Ten aanzien van de beoordelingen dient opgemerkt te worden dat verschillende evaluatiebureaus de standaarden verschillend kunnen operationaliseren wat betekent dat ook de beoordelingen kunnen verschillen. Meer uniformiteit ten aanzien van de operationalisaties zou dan ook gewenst zijn (zie ook paragraaf 6.1.3).

6.1.3 NVAO-databank beoordeelde opleidingen

De resultaten van de accreditaties zijn beschikbaar in een databank op de website van de NVAO, search.nvao.net. In deze paragraaf geven we twee voorbeelden van beoordelingen uit deze databank.

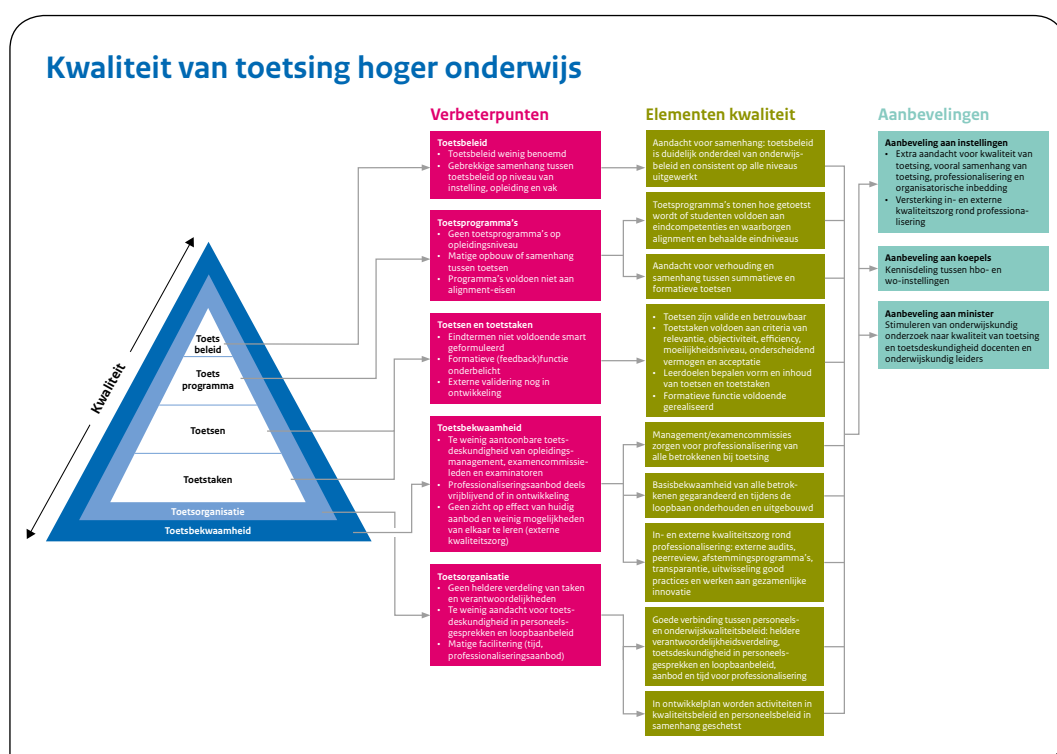
Het eerste voorbeeld betreft de beoordelingen van de bacheloropleiding Psychologie en de masteropleiding Psychologie van de Radboud Universiteit Nijmegen die binnen de clusterbeoordeling psychologie vallen waarvoor in 2012 in totaal 27 opleidingen beoordeeld werden. De QANU-visitatiecommissie beoordeelt standaard 3 voor de bacheloropleiding en de masteropleiding als voldoende. De argumentatie voor dit oordeel luidt als volgt: 'Er wordt binnen de opleidingen adequaat getoetst of de studenten de leerdoelen van programma-onderdelen hebben gerealiseerd. De opleidingen maken relatief veel gebruik van meerkeuze-toetsen, waarbij de commissie de opleidingen adviseert de kwaliteit daarvan in het oog te houden. Voor een groeiend deel van de toetsen zijn toetsmatrices beschikbaar waaruit blijkt hoe de toetsing aansluit op de leerdoelen. Het is duidelijk hoe docenten tot een beoordeling komen. De opleidingen zetten de laatste tijd sterk in op (borging van) toetskwaliteit. De examencommissie is relatief laat doch voortvarend gestart met een verscherpt toezicht op de toetsing. Op basis van de beoordeling van bachelor- en mastertheses stelt de commissie vast dat het niveau van zowel bachelorstudenten als masterstudenten overeenkomt met de verwachtingen die de commissie heeft van studenten van deze opleidingen. De oordelen van de commissie komen in het algemeen overeen met de door de opleidingen toegekende beoordelingen. De commissie stelt wel vast dat het niveau van enkele theses te wensen overliet, zij had de indruk dat in deze gevallen niet tijdig was bijgestuurd door de begeleider. Het gemiddelde niveau van de wetenschappelijke eindwerken bevestigt de commissie echter in haar opvatting dat de toetsing en beoordeling van wetenschappelijke vorming in de bachelor- en masteropleiding over het algemeen adequaat is. De commissie heeft door de structuur van de beoordelingsformulieren en door de wijze van invullen geen volledig zicht op hoe het eindoordeel van theses tot stand is gekomen. De commissie adviseert de opleiding hier aandacht aan te besteden. Aangezien de drie mastervarianten recent zijn gestart, is er nog niet goed zicht op de mate waarin de voorbereiding op de arbeidsmarkt adequaat is. De eerste signalen wijzen erop dat de masteropleiding goed aansluit op de praktijk, stelt de commissie vast.'

Het tweede voorbeeld betreft de beoordeling in 2013 van de bacheloropleiding Commerciële economie van Stenden Hogeschool. Standaard 3 wordt door de NQA-visitatiecommissie als voldoende beoordeeld. De argumentatie voor dit oordeel luidt als volgt: 'De opleiding maakt een zinvol onderscheid in de beoordeling van competenties (eindkwalificaties) in de integrale bouwstenen en van de beoordeling van kennis en vaardigheden van de student in de ondersteunende bouwstenen. Het toetsbeleid sluit daarmee goed aan bij het gekozen onderwijsconcept. Door het gebruik van toetsmatrijzen wordt zorgvuldig bewaakt dat alle competenties gedurende de opleiding op het gewenste niveau worden beoordeeld. Er is sprake van een zinvolle variatie van toetsvormen en van voldoende inhoudelijke feedback. De criteria waarop de studenten beoordeeld worden, zijn relevant en inzichtelijk. Bij de beoordeling van groepsproducten wordt de inbreng van individuele groepsleden voldoende meegenomen. Alle afzonderlijke onderwijseenheden worden individueel getoetst. De kwaliteit van de toetsing en beoordeling wordt actief bewaakt door de toetscommissie. De scriptiecommissie bewaakt zowel de kwaliteit van de afstudeervoorstellen als van het eindresultaat. In voorkomende gevallen treedt zij op als derde beoordelaar. Om de intersubjectiviteit van de beoordeling te bevorderen worden intervisiebijeenkomsten van docenten georganiseerd en wordt samengewerkt met andere opleidingen, binnen en buiten de hogeschool. Het panel is van mening dat de opleiding de beoogde eindkwalificaties en het hbo-bachelorniveau in voldoende mate realiseert. Het panel ziet een verhoging van het niveau van de afstudeerprojecten van het afgelopen studiejaar ten opzichte van voorgaande jaren. Het panel stelt vast dat de opleiding in dit opzicht een gericht beleid voert. Het heeft een nieuw beoordelingsformulier in gebruik genomen en heeft initiatieven genomen om de kwaliteit van de beoordeling te verhogen. Het panel heeft hier waardering voor en spoort de opleiding aan deze lijn verder door te zetten.

Uit onderzoek blijkt dat het werkveld en de afgestudeerden tevreden zijn over het niveau van de opleiding.’

Voorgaande twee voorbeelden zijn illustratief voor de beoordelingen van alle andere opleidingen uit het hoger onderwijs. Wat bij die beoordelingen opvalt is dat hoewel sommige opleidingen beter beoordeeld worden dan andere opleidingen, het oordeel bijna altijd hetzelfde, namelijk voldoende, is. Wat ook opvalt is de essentiële rol die de examencommissie speelt of geacht wordt te spelen. De rol van de examencommissie komt uitgebreid aan de orde in hoofdstuk 5 van dit boek.

In de NVAO-databank komen negatieve beoordelingen slechts bij weinig opleidingen voor. Dat zou de indruk kunnen wekken dat er niets mis is met toetsing en examinering in het hoger onderwijs. In het boek ‘Toetsen in het hoger onderwijs’ (2014) van Van Berkel, Bax en Joosten-ten Brinke staat echter in het voorwoord dat een van de redenen voor het boek is dat er nog steeds veel misgaat bij toetsing en examinering. Ook de reden voor het instellen van de commissie Bruijn was dat er in het hbo terechte twijfels waren over de kwaliteit van toetsing en examinering. Met betrekking tot standaard 3 is een van de aanbevelingen in het eindrapport van de Expertgroep Protocol (2014): ‘Bepleit bij de NVAO, de evaluatiebureaus en de zelfstandige secretarissen een uniforme wijze van werken bij het beoordelen van standaard 3 waarin er (a) gebruik gemaakt wordt van de beoordelingsmodellen van de opleiding, tenzij die van onvoldoende kwaliteit zijn, en (b) bij de visitatie gekeken wordt naar al het bewijs dat in het afstudeerprogramma verzameld wordt en niet alleen naar de eindwerkstukken.’ In een begin 2016 verschenen rapport, over de kwaliteit van de toetsing in het hoger onderwijs, noemt de Inspectie van het Onderwijs aan de hand van de kwaliteitspiramide (Joosten-ten Brinke, 2011; Sluijsmans e.a., 2015) de zaken die verbetering behoeven. In figuur 6.1, ontleend aan het inspectierapport, worden de verbeterpunten genoemd en aanbevelingen gegeven.



Figuur 6.1: Kwaliteit van toetsing in hoger onderwijs (bron: Inspectie van het Onderwijs)

6.2 RCEC-beoordelingssysteem voor de kwaliteit van toetsen en examens

Het RCEC heeft een systeem ontwikkeld voor de beoordeling van de kwaliteit van toetsen en examens die in het onderwijs gebruikt worden. Het beoordelingssysteem hanteert zes criteria:

- 1 Doel en gebruik
- 2 Kwaliteit van toets- en examenmateriaal
- 3 Representativiteit
- 4 Betrouwbaarheid
- 5 Standaardbepaling en normhandhaving
- 6 Afname en beveiliging

Elk criterium wordt beoordeeld door middel van vragen die als 'onvoldoende' (O, score 1), 'voldoende' (V, score 2) en 'goed' (G, score 3) beoordeeld en gescoord worden. Op basis van (de som van) de scores op de onderscheiden vragen wordt een criterium beoordeeld als 'goed', 'voldoende' of 'onvoldoende'. Hierna presenteren we de volledige informatie van het criterium 'Doel en gebruik'. De presentatie van de andere criteria beperkt zich tot een korte inleiding en de vragen die bij het betreffende criterium gesteld worden.

De informatie in deze paragraaf is ontleend aan de publicatie 'RCEC Beoordelingssysteem voor de kwaliteit van studietoetsen en examens' (Sanders, Van Dijk, Eggen, Den Otter & Veldkamp, 2016). Het beoordelingssysteem wordt steeds verder ontwikkeld, de meest recente versie is op te vragen via de website van het RCEC, zie www.rcec.nl > Beoordelingssysteem.

6.2.1 Criterium 1: Doel en gebruik

Bij de beoordeling van dit criterium wordt het doel en het gebruik van toetsen en examens beoordeeld. Is duidelijk 'wat' we toetsen of examineren en 'waarom' we dat doen? Het doel van toetsen en examens is te beoordelen of kandidaten over de vereiste kennis, vaardigheden of houdingen beschikken. Het gebruik van toetsen en examens betreft de beslissingen die op basis van de door de kandidaten behaalde resultaten op toetsen en examens genomen worden. Merk op dat in het RCEC-beoordelingssysteem onder examens ook praktijkexamens verstaan worden.

Criterium 1: Doel en gebruik		O	V	G
Basisvraag 1.1	Is aangegeven wat de doelgroep(en) van de toets of het examen is (zijn)?	1	2	3
Bij onvoldoende beoordeling kan men de twee andere vragen van dit criterium overslaan en doorgaan met criterium 2				
Basisvraag 1.2	Is het meetdoel van de toets of het examen beschreven?	1	2	3
Bij onvoldoende beoordeling kan men basisvraag 3 van dit criterium overslaan en doorgaan met criterium 2				
Basisvraag 1.3	Is aangegeven wat het gebruiksdoel van de toets of het examen is?	1		3

Aanwijzingen bij basisvraag 1.1 | Is aangegeven wat de doelgroep(en) van de toets of het examen is (zijn)?

In ieder geval moet de opleiding benoemd zijn waarvoor de toets of het examen wordt ingezet. Maar ook het aangeven van de leeftijd, het beroep, het opleidingsniveau of de relevante voorkennis van kandidaten zijn mogelijkheden om de doelgroep te definiëren. Deze informatie

kan onder andere belangrijk zijn bij het beoordelen van de inhoud van de toets of het examen zoals het taalgebruik en de gehanteerde normen of cesuren.

Aanwijzingen bij basisvraag 1.2 | Is aangegeven wat het meetdoel van de toets of het examen is? Een toets of examen moet vaststellen wat kandidaten na afloop van een onderwijstraject of een EVC-traject wel en niet beheersen. Wat de kandidaten geacht worden te beheersen, kan onder andere aangegeven worden als:

- de beheersing van een bepaald construct (bijvoorbeeld 'leesvaardigheid');
- de beheersing van een exameneenheid van een examenprogramma (bijvoorbeeld de exameneenheid 'havo-examen wiskunde');
- de beheersing van een kerntaak, beroepstaak of werkproces (bijvoorbeeld uit een mbo-kwalificatiedossier);
- de beheersing van een competentie (bijvoorbeeld 'analyseren' van een assistent-drogist).

Een toets of examen die een construct of een competentie meet, zal een gedetailleerde beschrijving met voorbeelden moeten geven van de theorie waarop het bedoelde construct of de competentie gebaseerd is. Dit impliceert dat niet volstaan kan worden met 'deze toets meet het construct leesvaardigheid' of 'dit examen meet de competentie analyseren'. Indien de toets of het examen niet expliciet naar een construct of competentie verwijst, zullen de onderwerpen, kerntaken of exameneenheden die in de toets of het examen aan de orde komen, beschreven moeten worden. In voorkomende gevallen kan volstaan worden met een verwijzing naar relevante brondocumenten.

Bij deze vraag is het van belang dat de relevantie van de inhoud van de toets of het examen voor het beoogde doel aannemelijk gemaakt wordt. Daarbij kan bijvoorbeeld een toetsmatrijs van de toets of het examen goede diensten bewijzen.

Aanwijzingen bij basisvraag 1.3 | Is aangegeven wat het gebruiksdoel van de toets of het examen is?

Een toets of examen kan gebruikt worden voor:

- *Selectie* | Afhankelijk van het toets- of examenresultaat wordt een leerling wel of niet toegelaten tot een opleiding.
- *Classificatie* | Afhankelijk van het toets- of examenresultaat volgen leerlingen verschillende onderwijsprogramma's die tot verschillende diploma's of certificaten leiden (bijvoorbeeld de voormalige Eindtoets Basisonderwijs van Cito).
- *Plaatsing* | Afhankelijk van het toets- of examenresultaat volgen leerlingen verschillende onderwijsprogramma's die tot hetzelfde certificaat of diploma leiden (bijvoorbeeld een zelfbeoordeling ten behoeve van de BOL- of BBL-leerweg in het mbo).
- *s of diplomering* | Afhankelijk van het toets- of examenresultaat wordt wel of niet een diploma of certificaat verstrekt.

Een ander gebruiksdoel van toetsen dat in de onderwijskundige literatuur opgang doet, betreft de drie benaderingen van formatief assessment:

- data-based decision making (DBDM), in de Nederlandse literatuur opbrengstgericht werken (OGW) genoemd;
- assessment for learning (AfL), in de Nederlandse literatuur ook wel toetsen of evaluatie van het leren genoemd;
- diagnostische toetsen (DT).

Voor meer informatie over formatief toetsen verwijzen we naar het proefschrift van Van der Kleij (2013), te downloaden op de website van het RCEC, www.rcec.nl.

Eindoordeel criterium 1: Doel en gebruik		
De som van de beoordelingen op de vragen 1 t/m 3 is 8 of 9	Twee vragen worden met '3' beoordeeld en een andere vraag wordt met minstens '2' beoordeeld	Goed
De som van de beoordelingen op de vragen 1 t/m 3 is 6 of 7	Eén vraag wordt met '3' beoordeeld en de andere vragen worden minstens met '2' beoordeeld of de andere vragen worden met '2' beoordeeld	Voldoende
	Eén van de vragen wordt met '1' beoordeeld	Onvoldoende

6.2.2 Criterium 2: Toets- of examenmateriaal

Bij de beoordeling van dit criterium wordt een onderscheid gemaakt tussen toetsen en examens die met behulp van de computer worden afgenomen, en toetsen en examens die schriftelijk worden afgenomen en uit gesloten en/of open vragen bestaan, en (praktijk)examens die uit één of meer praktijkopdrachten bestaan. Ook kunnen toetsen en examens uit combinaties van gesloten vragen, open vragen en praktijkopdrachten bestaan. Bij gesloten of (meervoudige) meerkeuzevragen moet de kandidaat het goede antwoord (of goede antwoorden) selecteren, bij open vragen het goede antwoord (of goede antwoorden) formuleren en bij een (praktijk) examen de praktijkopdrachten uitvoeren.

Er worden bij dit criterium twee basisvragen gesteld. In het algemeen geldt dat om de score(s) op een toets of examen zinvol te kunnen interpreteren, de toets of het examen zodanig dient te zijn afgenomen en gescoord dat andere, niet-beoogde factoren geen invloed kunnen uitoefenen op de totstandkoming van de score(s). Zo dient bijvoorbeeld de afname en instructie dusdanig gestandaardiseerd te zijn dat de invloed van variatie in instructie of verschil in de toets- of examenleider of van de afnamesituatie op de score is geëlimineerd of in ieder geval binnen de grenzen van het mogelijke is beperkt. Ook moet de scoring zo objectief mogelijk zijn. Voor de beoordeling van een toets of examen die schriftelijk afgenomen wordt, dient men te beginnen met vraag 2.1, voor een computertoets of -examen met vraag 2.8. Indien er van een toets of examen zowel een schriftelijke versie als een computerversie bestaat, dient de kwaliteit van het toets- en examenmateriaal van beide versies te worden beoordeeld.

Criterium 2: Toets- of examenmateriaal		O	V	G
Schriftelijke toets of examen				
Basisvraag 2.1	Zijn de vragen of opdrachten gestandaardiseerd?	1		3
	Bij onvoldoende beoordeling (1) van deze vraag kan men de rest van de vragen van dit criterium overslaan en doorgaan met criterium 3			
Basisvraag 2.2	a Is er sprake van een geautomatiseerd of objectief scoringssysteem? en/of b Als de scoring door beoordelaars gebeurt, is dan het beoordelingsvoorschrift volledig en duidelijk?	1	2	3
	Bij onvoldoende beoordeling (1) van deze vraag kan men de rest van de vragen van dit criterium overslaan en doorgaan met criterium 3			
2.3	Zijn de vragen of praktijkopdrachten, het toets- of examenboekje, de antwoord-schalen en/of het antwoordformulier zodanig ontworpen dat fouten bij de invulling voorkomen worden?	1		3
2.4	Is het scoringssysteem zodanig ontworpen en beschreven dat fouten bij de scoring voorkomen worden?	1		3
2.5	Is de instructie voor de kandidaat volledig en duidelijk?	1	2	3
2.6	Zijn de vragen of opdrachten correct geformuleerd?	1	2	3
2.7	Hoe is de kwaliteit van het toets- of examenmateriaal?	1	2	3

Criterium 2: Toets- of examenmateriaal		O	V	G
Computertoets of computereexamen				
Basisvraag 2.8	Zijn de vragen gestandaardiseerd?	1		3
	Bij onvoldoende beoordeling (1) van deze vraag kan men de rest van de vragen van dit criterium overslaan en doorgaan met criterium 3			
Basisvraag 2.9	Is er sprake van een geautomatiseerd of objectief scoringssysteem?	1	2	3
	Bij onvoldoende beoordeling (1) van deze vraag kan men de rest van de vragen van dit criterium overslaan en doorgaan met criterium 3			
2.10	Is de software zodanig ontworpen dat fouten door onjuist gebruik voorkomen worden?	1	2	3
2.11	Is de instructie voor de kandidaat volledig en duidelijk?	1	2	3
2.12	Zijn de vragen correct geformuleerd?	1	2	3
2.13	Hoe is de kwaliteit van de vormgeving van de gebruikersinterface?	1		3

6.2.3 Criterium 3: Representativiteit

Bij dit criterium wordt de representativiteit van de toets of het examen beoordeeld. Representativiteit heeft betrekking op zowel de inhoud als moeilijkheidsgraad van de toets of het examen. De inhoud van toetsen en examens is gebaseerd op wat een kandidaat wordt onderwezen. Deze leerdoelen worden, afhankelijk van de onderwijssector, geformuleerd als kerndoelen, eindtermen, basiskwalificaties, kerntaken of competenties. Omdat deze doelen nog te algemeen zijn om er toetsen of examens op te kunnen baseren, dienen ze uitgewerkt te worden tot toetsbare leerdoelen. Indelingsschema's of taxonomieën voor menselijk presteren, meestal aangeduid met toetsmatrizen, vormen hierbij een nuttig hulpmiddel om de beoogde leerdoelen uit te werken tot toetsbare leerdoelen.

Omdat de meeste studietoetsen en examens eerder directe metingen van menselijk gedrag zijn (en veelal ook door gebruikers als zodanig opgevat worden) dan metingen van constructen of competenties, wordt bij dit criterium prioriteit gegeven aan de inhoud van de toets of het examen. Mochten bij sommige toetsen of examens wel constructen of competenties worden gemeten, en als daar in de verantwoording van de toets of het examen bewijzen voor worden aangevoerd, dan zal de auditor zijn beoordeling op die bewijzen baseren. Zie hierover ook hoofdstuk 4 van Toetsen op School, te downloaden op www.toetsenschool.nl.

Behalve dat de inhoud van de toets of het examen de leerdoelen dient te representeren, dient de moeilijkheidsgraad van de vragen of praktijkopdrachten, en dus de toets of het examen, ook afgestemd te zijn op de beoogde doelgroep. Dit betekent in de praktijk dat het merendeel van de vragen of praktijkopdrachten niet te moeilijk of te makkelijk mag zijn voor de kandidaten.

Criterium 3: Representativiteit		O	V	G
Basisvraag 3.1	Is de toetsmatrizen, het examenprogramma, examenplan, competentieprofiel of de operationalisatie van het construct een adequate representatie van het meetdoel?	1	2	3
Bij onvoldoende beoordeling (1) van deze vraag kan men de rest van de vragen van dit criterium overslaan en doorgaan met criterium 4				
Basisvraag 3.2	Is de moeilijkheidsgraad van de vragen of de praktijkopdrachten afgestemd op de beoogde doelgroep?	1	2	3

6.2.4 Criterium 4: Betrouwbaarheid

Bij dit criterium wordt de betrouwbaarheid van (de scores van) een toets of examen beoordeeld. Bij betrouwbaarheid gaat het om de vraag of we vertrouwen kunnen hebben in de scores die kandidaten op een toets of examen behalen. De betrouwbaarheid is te kwantificeren met een betrouwbaarheidscoëfficiënt, het percentage misclassificaties en de standaardmeetfout. De betrouwbaarheidscoëfficiënt heeft een ondergrens van 0,0 en een bovengrens van 1,0. Een hoge betrouwbaarheidscoëfficiënt geeft aan dat we vertrouwen hebben in de betrouwbaarheid van het examen. We mogen dan verwachten dat indien de kandidaten twee keer hetzelfde examen zouden maken, zij een vergelijkbare score behalen. Onder misclassificaties verstaan we het aantal kandidaten dat als gevolg van de onbetrouwbaarheid van de toets of het examen ten onrechte gezakt en ten onrechte geslaagd is. Bij een hoge betrouwbaarheid is het aantal misclassificaties gering. Met de standaardmeetfout kunnen we de vraag naar de betrouwbaarheid van de scores van individuele kandidaten beantwoorden, oftewel welke andere score had een kandidaat ook op de toets of het examen kunnen behalen. Bij een hoge betrouwbaarheid ligt de score van het eerste examen erg dicht bij de score op het tweede examen.

Voor meer informatie over het berekenen en interpreteren van de betrouwbaarheid van een examen verwijzen we naar hoofdstuk 3 van Toetsen op School, www.toetsenopschool.nl.

Criterium 4: Betrouwbaarheid		O	V	G
Basisvraag 4.1	Zijn of worden betrouwbaarheidsgegevens verstrekt?	1	2	3
	Bij onvoldoende beoordeling van deze vraag kan men direct doorgaan naar criterium 5			
4.2	Zijn of worden de betrouwbaarheidsgegevens correct berekend?	1	2	3
4.3	Zijn de betrouwbaarheidsgegevens voldoende gezien de beslissingen die met de toets of het examen genomen worden?	1	2	3

6.2.5 Criterium 5: Standaardbepaling en normhandhaving

Bij dit criterium beoordelen we hoe de normen van toetsen of examens bepaald zijn en hoe de normen van vergelijkbare of parallelle toetsen of examens gehandhaafd zijn. Normeren kunnen we onderscheiden in relatief en absoluut normeren. Het doel van relatief normeren is het opstellen van relatieve normen. Kenmerkend voor relatieve normen is dat de score van een kandidaat vergeleken wordt met de scores die andere kandidaten van dezelfde doelgroep behaald hebben. Relatieve normen worden vooral gebruikt bij psychologische tests en studietoetsen die door testuitgevers op de markt worden gebracht. Het doel van absoluut normeren is het opstellen van absolute normen. Kenmerkend voor absolute normen is dat de score van een kandidaat vergeleken wordt met een score die als standaard, norm of cesuur aangeduid wordt.

De bespreking hier beperken we tot absoluut normeren. Het doel van absoluut normeren is het bepalen van absolute normen/standaarden/cesuren. De term standaard wordt gebruikt als afkorting voor prestatiestandaard of beheersingsstandaard. De meest bekende standaard of cesuur van een toets of examen is de score die de grens vormt tussen een voldoende of slagen en een onvoldoende of zakken. Dat betekent dat kandidaten met een score gelijk aan of boven de cesuur een voldoende behalen en dus geslaagd zijn voor de toets of het examen en dat kandidaten met een score onder de cesuur een onvoldoende behalen en dus gezakt zijn. Het is echter ook mogelijk om voor een toets of examen meer dan één standaard of cesuur te bepalen. In dat geval worden de kandidaten niet in twee categorieën (gezakt of geslaagd) maar in meerdere categorieën, bijvoorbeeld de cijfers 1,0 tot en met 10,0 onderverdeeld. Methoden voor standaardbepaling kunnen onderscheiden worden in methoden die gebaseerd zijn op de beoordeling van de vragen/opgaven/opdrachten van een toets of examen, bijvoorbeeld de methode van Angoff, en methoden die gebaseerd zijn op de beoordeling van de kandidaten die een examen maken, bijvoorbeeld de methode van de contrasterende groepen. Voor meer informatie zie hoofdstuk 9 van Toetsen op School, www.toetsenopschool.nl.

Het tweede doel van normeren bij toetsen of examens is het handhaven van eenmaal bepaalde standaarden. De methoden om standaarden (normen) te handhaven worden meestal met normhandhaving aangeduid. Zie voor meer informatie op www.toetswijzer.nl > ToetsSpecials de [ToetsSpecial](#) over normering bij de centrale examens in het voortgezet onderwijs.

Criterium 5: Standaardbepaling en normhandhaving		O	V	G
5.1	Worden standaarden/normen/cesuren verstrekt? Bij onvoldoende beoordeling (1) van deze vraag kan men de rest van de vragen van dit criterium overslaan en doorgaan met criterium 6	1		3
Toets of examen met absolute normen				
5.2	Is de standaardbepaling correct bepaald? a Is de standaardbepalingsmethode op de juiste wijze uitgevoerd? b Zijn de beoordelaars/vakdeskundigen naar behoren geselecteerd en getraind? c Is er voldoende overeenstemming tussen de beoordelaars? Bij onvoldoende beoordeling (1) van deze vraag kan men de rest van de vragen van dit criterium overslaan en doorgaan met criterium 6	1	2	3
Toets of examen met relatieve normen				
5.3	Wat is de kwaliteit van de verstrekte normen? a Zijn de normgroepen groot genoeg? b Zijn de normgroepen representatief? Bij onvoldoende beoordeling (1) van deze vraag kan men de rest van de vragen van dit criterium overslaan en doorgaan met criterium 6	1	2	3
5.4	Worden de betekenis en de beperkingen van de normschaal duidelijk gemaakt voor de gebruiker en is het type normschaal in overeenstemming met het doel van de toets of het examen? Bij onvoldoende beoordeling (1) van deze vraag kan men de rest van de vragen van dit criterium overslaan en doorgaan met criterium 6	1	2	3
5.5	Worden er gegevens verstrekt over: a Gemiddelden, standaardafwijkingen van de scoreverdeling? b Nauwkeurigheid van de meting en daarbij behorende intervallen (standaardmeetfout, standaardschattingsfout, testinformatie/ standaardfout)?	1		3
Voor absolute en relatieve normen				
5.6	Worden normen/standaarden/cesuren gehandhaafd? Bij onvoldoende beoordeling (1) van deze vraag kan men doorgaan met criterium 6	1		3
5.6.a	Is de methode op grond waarvan de norm(en) is (zijn) gehandhaafd correct?	1		3

6.2.6 Criterium 6: Afname en beveiliging

Om de afname van een toets of examen goed te doen verlopen, moet er informatie beschikbaar zijn die met name voor de surveillant bedoeld is (zie ook functieprofiel surveillant op www.nvexamens.nl). Deze informatie dient in overzichtelijke vorm, hetzij op papier hetzij digitaal, beschikbaar te zijn. Voor toetsen of examens die via de computer worden afgenomen, geldt dat tevens specifieke aanwijzingen moeten worden gegeven met betrekking tot de installatie en/of het opstarten en het gebruik van de toets of het examen. Soms is deze informatie gebundeld in een aparte installatiehandleiding.

Een toets of examen dient ook 'goed' beveiligd te zijn, dat wil zeggen dat al het mogelijke gedaan moet worden om de toegang tot de toets of het examen, het toets- of examenmateriaal en de toets- of examenresultaten te beveiligen. Bij de beveiliging van toetsen of examens dienen we een onderscheid te maken tussen enerzijds computertoetsen of -examens en anderzijds schriftelijke toetsen en examens.

Criterium 6: Afname en beveiliging		O	V	G
Basisvraag 6.1	Is er voor de surveillant voldoende informatie over de afname van de toets of het examen beschikbaar?	1		3
	Bij onvoldoende beoordeling (1) van deze vraag kan men de rest van de vragen van dit criterium overslaan			
6.1.a	Is de informatie (voor de surveillant) volledig en duidelijk?	1	2	3
6.1.b	Wordt de mate van deskundigheid die vereist is om de toets of het examen af te nemen vermeld?	1	2	3
Basisvraag 6.2	Is de toets of het examen voldoende beveiligd?	1	2	3
	Bij onvoldoende beoordeling (1) van deze vraag kan men de rest van de vragen van dit criterium overslaan			
Extra vragen voor afname via computer				
6.3	Wordt er informatie gegeven over de installatie van de computersoftware?	1	2	3
6.4	Wordt er informatie gegeven over de bediening en mogelijkheden van de software?	1	2	3
6.5	Zijn er voldoende mogelijkheden voor technische ondersteuning?	1	2	3

6.3 NVAO-beoordelingskader en RCEC-beoordelingssysteem

Hiervoor zagen we dat het NVAO-beoordelingskader gebruikt wordt om de basiskwaliteit van een opleiding te beoordelen. Op basis van die beoordeling wordt een opleiding al of niet geaccrediteerd. Een van de drie standaarden, de derde standaard, heeft betrekking op toetsing: 'De opleiding beschikt over een adequaat systeem van toetsing'. Bij deze standaard wordt de volgende toelichting gegeven: 'Het gerealiseerde niveau blijkt uit de tussentijdse en afsluitende toetsen, de afstudeerwerken en de wijze waarop afgestudeerden in de praktijk of in een

vervolgopleiding functioneren. De toetsen en de beoordeling zijn valide, betrouwbaar en voor studenten inzichtelijk.'

Het RCEC-beoordelingssysteem is ontwikkeld om de kwaliteit van toetsen en examens te beoordelen. Toetsen en examens worden beoordeeld op zes criteria die door bijbehorende vragen bij de criteria geoperationaliseerd worden. De toelichting bij standaard 3 maakt duidelijk dat het NVAO-beoordelingskader deze operationalisaties ontbeert. Vandaar dat hier voorgesteld wordt om het RCEC-beoordelingssysteem in te zetten bij standaard 3 van het NVAO-beoordelingskader. Opleidingen zouden in hun kritische reflectie moeten laten zien dat zij hun toetsen en examens met het RCEC-beoordelingssysteem beoordeeld hebben en aangeven wat de resultaten van die beoordeling waren. Het laatste is met name een taak van de examencommissie, in voorkomende gevallen bijgestaan door een toetscommissie (zie www.nvexamens.nl > Functieprofielen voor een functieprofiel van een lid van een examen- of toetscommissie). Die toetsbeoordelingen zouden niet alleen het werk van de visitatiepanels vereenvoudigen maar ook de objectiviteit en kwaliteit van de beoordelingen ten goede komen. Het voorgaande betekent wel dat niet alleen leden van examencommissies en toetscommissies maar ook ten minste één lid van de visitatiepanels bekend moeten zijn met het RCEC-beoordelingssysteem. Merk op dat de beoordeling van standaard 4 van het NVAO-beoordelingskader afhankelijk is van de beoordeling van standaard 3. Indien standaard 3 als onvoldoende beoordeeld wordt dan kan niet vastgesteld worden of de beoogde eindkwalificaties gerealiseerd zijn.

Het RCEC-beoordelingssysteem kan ook zeer goed ingezet worden bij het gebruik van de kwaliteitspiramide (Sluijsmans et al., 2012), met name bij de beoordeling van de kwaliteit van toetsitems en de toets (Jaspers en Van Zijl, 2011). Het RCEC-beoordelingssysteem kan ook aangepast worden voor de beoordeling van het examenprogramma; deze aanpassing komt met name neer op een meerdimensionale benadering van de criteria betrouwbaarheid en validiteit (Van Rijn et al., 2012).

Het RCEC-beoordelingssysteem is geen statisch systeem (Wools, 2015). Op basis van ervaringen opgedaan bij de beoordelingen zal het beoordelingssysteem indien nodig aangepast worden om de kwaliteit van toetsing in het hoger onderwijs zo goed mogelijk te kunnen beoordelen en op die manier de kwaliteit van opleidingen te verbeteren.

Geraadpleegde literatuur

Geraadpleegde literatuur

AERA (1999). *Standards for educational and psychological testing*. Joint Committee on Standards for Educational and Psychological Testing. Washington DC: American Educational Research Association. Nieuwe editie verschenen in 2014.

Andriessen, D. (2012). *Kwaliteitscriteria voor afstudeeronderzoek voor HBO bacheloropleidingen*. Hogeschool Inholland.

Arter, J., & Chappuis, J. (2007). *Creating & recognizing quality rubrics*. Upper Saddle River, NJ: Pearson / Prentice Hall.

Baartman, L.K., Bastiaens, T.J., Kirschner, P.A., & Vleuten, C.P. van der (2006). The wheel of competency assessment: Presenting quality criteria for competency assessment programs. *Studies in Educational Evaluation*, 32 (2), 153-170.

Baldiga, K. (2012). *Gender differences in willingness to guess and the implications for test scores*. Unpublished doctoral dissertation. Boston: Harvard University.

Berk, R.A. (1996). Standard setting: The next generation (where few psychometricians have gone before!). *Applied Measurement in Education*, 9 (3), 215-235.

Berkel, H.J.M. van, Bax, A., & Joosten-ten Brinke, D. (2014). *Toetsen in het hoger onderwijs*. Houten: Bohn Stafleu van Loghum.

Biggs, J. (1996). Enhancing teaching through constructive alignment. *Higher Education*, 32 (3), 347-364.

Birenbaum, M., & Dochy, F. (Ed.) (1996). *Alternatives in assessment of achievements, learning, processes, and prior knowledge*. Boston: Kluwer.

Boulet, J.R., McKinley, D.W., Whelan, G.P., & Hambleton, R.K. (2003). Quality assurance methods for performance-based assessments. *Advances in Health Sciences Education*, 8 (1), 27-47.

Butler, S.M., & McMunn, N.D. (2006). *A teacher's guide to classroom assessment: Understanding and using assessment to improve student learning*. San Francisco: Jossey-Bass.

Cito (2010). *Toetstechnische begrippenlijst*. Arnhem: Cito.
Online: www.cito.nl/onderzoek%20en%20wetenschap/kenniscentrum/begrippenlijst

Cizek, G.J. (Ed.) (2001). *Setting performance standards: Concepts, methods, and perspectives*. Lawrence Erlbaum Associates.

Cohen, S.A. (1987). Instructional alignment: Searching for a magic bullet. *Educational Researcher*, 16 (8), 16-20.

Collins, A. (1991). Portfolios for biology teacher assessment. *Journal of Personnel Evaluation in Education*, 5 (2), 147-168.

Commissie Bruijn (2012). *Vreemde ogen dwingen: Eindrapport Commissie externe validering examenkwaliteit hoger beroepsonderwijs*. Den Haag: Vereniging Hogescholen.
Online: www.vereniginghogescholen.nl/publicaties-en-verenigingsafspraken/publicaties-gesorteerd-op-verschijningsjaar-1/2012-1/1637-vreemde-ogen-dwingen-1/file

COTAN (2009/2010). *COTAN Beoordelingssysteem voor de kwaliteit van tests*.
Online: www.psynip.nl > Tests / COTAN > COTAN Beoordeling

Curry-Sumner, I. (et al.) (2010). *Onderzoeksvaardigheden: Instructie voor juristen*. Nijmegen: Ars Aequi.

Curry-Sumner, I., & Schaaf, M.F. van der (2011). The theory and practice of teaching and guiding legal research skills. *Recht en Methode in Onderzoek en Onderwijs*, 1 (1), 64-84.

Douglas, K.M., & Mislevy, R.J. (2010). Estimating classification accuracy for complex decision rules based on multiple scores. *Journal of Educational and Behavioral Statistics*, 35 (3), 280-306.

Downing, S.M., & Haladyna, T.M. (2006). *Handbook of test development*. Lawrence Erlbaum Associates.

Eraut, M., & Du Boulay (2000). Developing the attributes of medical professional judgement and competence: A review of the literature. *Cognitive Sciences Research Paper 518*. University of Sussex.
Online: <http://users.sussex.ac.uk/~bend/doh/reporthtml.html>

Expertgroep BKE/SKE (2013). *Verantwoord toetsen en beslissen in het hoger onderwijs: Een voorstel voor een programma van eisen voor een basis- en seniorkwalificatie examinering (BKE/ SKE)*. Den Haag: Vereniging Hogescholen.

Expertgroep Protocol (2014). *Beoordelen is mensenwerk: Bevindingen over de wenselijkheid en mogelijkheid van een gezamenlijk protocol voor het beoordelen van (kern)werkstukken*. Den Haag: Vereniging Hogescholen.

Fitzpatrick, R., & Morrison, E.J. (1971). Performance and product evaluation. In: Thorndike, E.L. (Ed.), *Educational Measurement*. 2nd edition, 237-270. Washington, DC: American Council on Education.

Frederiksen, N. (1984). The real test bias: Influences of testing on teaching and learning. *American Psychologist*, 39 (3), 193-202.

Gielen, S., Peeters, E., Dochy, F., Onghena, P., & Struyven, K. (2010). Improving the effectiveness of peer feedback for learning. *Learning and Instruction*, 20 (4), 304-315.

Goldman, R.L. (1994). The reliability of peer assessments: A meta-analysis. *Evaluation & the Health Professions*, 17 (1), 3-21.

Hafner, J., & Hafner, P. (2003). Quantitative analysis of the rubric as an assessment tool: An empirical study of student peer-group rating. *International Journal of Science Education*, 25 (12), 1509-1528.

Haladyna, T.M. (1997). *Writing test items to evaluate higher order thinking*. Needham Heights, MA: Allyn & Bacon.

- Haladyna, T.M., & Rodriguez, M.C. (2013). *Developing and validating test items*. Routledge.
- Hambleton, R.K., Jeager, R., Plake, B.S., & Mills C.N. (2000). Setting performance standards on complex educational assessments. *Applied Psychological Measurement*, 24 (4), 355-366.
- Hambleton, R.K., & Pitoniak, M.J. (2006). Setting performance standards. In: Brennan, R.L. (Ed.), *Educational measurement*. 4th edition. Westport, CT: American Council on Education / Praeger.
- Hofstee, W.K.B. (1999). *Principes van beoordeling: Methodiek en ethiek van selectie, examineren en evaluatie*. Lisse: Swets & Zeitlinger Publishers.
- Hollingworth, W., Medina, L.S., Lenkinski, R.E., Shibata, D.K., Zurakowski, D., Comstock, B., & Jarvik, J.G. (2006). Interrater reliability in assessing quality of diagnostic accuracy studies using the QUADAS tool: A preliminary assessment. *Academic Radiology*, 13 (7), 803-810.
- Inspectie van het Onderwijs (2009). *Boekhouder of wakend oog?: Verslag van een onderzoek bij examencommissies in het hoger onderwijs over de garantie van het niveau*. Utrecht: Inspectie van het Onderwijs.
Online: www.onderwijsinspectie.nl/publicaties/2009/05/Boekhouder+of+wakend+oog.html
- Inspectie van het Onderwijs (2011). *Alternatieve afstudeertrajecten en de bewaking van het eindniveau in het hoger onderwijs*. Utrecht: Inspectie van het Onderwijs.
Online: www.onderwijsinspectie.nl/publicaties/2011/04/Alternatieve+afstudeertrajecten+hoger+onderwijs.html
- Inspectie van het Onderwijs (2011). *Toezichtkader bve 2012*. Utrecht: Inspectie van het Onderwijs.
Online: www.onderwijsinspectie.nl/publicaties/2011/10/Toezichtkader+bve+2012.html
- Inspectie van het Onderwijs (2015). *Verdere versterking: Onderzoek naar het functioneren van examencommissies in het hoger onderwijs*. Utrecht: Inspectie van het onderwijs.
Online: www.onderwijsinspectie.nl/publicaties/2015/04/verdere-versterking-examencommissies-hoger-onderwijs.html
- Inspectie van het Onderwijs (2016). *De kwaliteit van de toetsing in het hoger onderwijs*. Utrecht: Inspectie van het Onderwijs.
- International Testing Committee (2013). *ITC guidelines on quality control*.
Online: www.intestcom.org/page/19
- Jaspers, M., & Zijl, E. van (2011). *Kwaliteit van toetsing in het hoger onderwijs: Naar een versterkte rol van examencommissies*. Tilburg: Fontys Hogescholen, Dienst Onderwijs en Onderzoek.
Online: <http://fontys.surfsharekit.nl:8080/get/smpid:28091/DS1/>
- Joosten-ten Brinke (2011). *Eigentijds toetsen en beoordelen*. Tilburg: Fontys Lerarenopleiding Tilburg.
- Kamminga, M. (2011). De tien valkuilen van digitaal toetsen. *Euclides*, 86 (5), 207-209.
- Kane, M.T. (1992). *The validity of assessments of professional competence*.
Online: <http://files.eric.ed.gov/fulltext/ED343958.pdf>

Kane, M.T. (2006). Validation. In: Brennan, R.L. (Ed.), *Educational Measurement*. 4th edition, 17-64. Washington, DC: The National Council on Measurement in Education & The American Council on Education.

Karpicke, J.D., & Blunt, J.R. (2011). Retrieval practice produces more learning than elaborative studying with concept mapping. *Science*, 331, 772-775.

Kayzel, R. (2005). *Competentiegericht opleiden*. Cahier 1. Amsterdam: Hogeschool van Amsterdam.

Keijzer-de Ruiter, M. (2011). *Ontwikkelingen in de Delftse Toetspraktijk: De (digitale) toets als diagnose, leermiddel, voortgangscontrole en kwalificatieinstrument binnen de Delftse Toetspraktijk*. Delft: TU Delft.

Online: www.icto.tudelft.nl/icto/publicaties/ontwikkelingen-in-de-delftse-toetspraktijk/

Kleij, F.M. van der (2013). *Computer-based feedback in formative assessment*. Proefschrift. Enschede: Universiteit Twente.

Online: www.rcec.nl/Alumni/alumni/ProefschriftVanderkleij.pdf

Kleijn, R.A.M. de, Prins, F.J., & Tartwijk, J. van (2013). *Students' use of a rubric for research papers*. Paper gepresenteerd tijdens de conferentie van de European Association for Research on Learning and Instruction (EARLI), München, Duitsland.

Linn, R., & Burton, E. (1994). Performance-based assessment: Implications of task specificity. *Educational Measurement: Issues and Practices*, 13 (1), 5-8.

McMillan, J.H. (2011). *Classroom Assessment: Principles and practice for effective standards-based instruction*. Fifth Edition. Boston, MA: Allyn & Bacon / Merrill Education.

Meijers, A.W.M., Borghuis, V.A.J., Mutsaers, E.J.P.J., Overveld, V.C., & Perrenet, J.C. (2005). *Criteria voor academische bachelor en master curricula*. Eindhoven: Technische Universiteit.

Messick, S. (1994). The interplay of evidence and consequences in the validation of performance assessments. *Educational Researcher*, 23 (2), 13-23.

Ministerie van OCW. *Regeling standaarden examenkwaliteit mbo 2012*.

Online: <http://wetten.overheid.nl/BWBR0030427>

Ministerie van OCW. *Monitor Trends in Beeld*. Kwaliteit en prestaties van de Nederlandse onderwijs-, cultuur- en wetenschapsstelsels in woord, beeld en cijfers. Website.

Online: www.trendsinbeeldocw.nl

Ministerie van OCW. *Wet op het hoger onderwijs en wetenschappelijk onderzoek*.

Online: <http://wetten.overheid.nl/BWBR0005682>

Ministerie van OCW. *Wet versterking kwaliteitswaarborgen hoger onderwijs*.

Online: <http://wetten.overheid.nl/BWBR0034472>

Ministerie van OCW (2010). *Wet van 4 februari 2010 tot wijziging van de Wet op het hoger onderwijs en wetenschappelijk onderzoek en enige andere wetten onder meer in verband met de verbetering van het bestuur bij de instellingen voor hoger onderwijs, de collegegeldsystematiek en de rechtspositie van studenten (versterking besturing)*. Staatsblad nr. 119, 18 maart 2010.
Online: <https://zoek.officielebekendmakingen.nl/stb-2010-119.html>

Moss, P.A. (1994). Can there be validity without reliability? *Educational Researcher*, 23 (2), 5-12.

Nitko, A.J., & Brookhart, S.M. (2011). *Educational Assessment of Students*. 6th edition. Boston, MA: Pearson.

NVAO (2004). *Dublin descriptoren*. Den Haag: NVAO.

NVAO (2014). *Beoordelingskaders accreditatiestelsel hoger onderwijs Nederland*. Den Haag: NVAO.

NVAO (2015). *Richtlijn beoordeling eindwerken tijdens visitaties*. Den Haag: NVAO.
Online: www.nvaio.net/system/files/procedures/Richtlijn%20NVAO%20beoordeling%20eindwerken%202015.pdf

NVE (2013/2015). *Functieprofielen medewerkers examenorganisatie*.
Online: www.nvexamens.nl/functieprofielen.html

Onderwijsraad (2002). *Competenties: Van complicatie tot compromis*. Studie. Den Haag: Onderwijsraad.
Online: www.onderwijsraad.nl/upload/publicaties/414/documenten/studie_competenties.pdf

Onderwijsraad (2004). *De blik naar buiten: Transparantie en kwaliteitsborging van examinering in het hoger onderwijs*. Studie. Den Haag: Onderwijsraad.
Online: www.onderwijsraad.nl/publicaties/2004/de-blik-naar-buiten/item636

Orsmond, P., & Merry, S. (2013). The importance of self-assessment in students' use of tutors' feedback: A qualitative study of high and non-high achieving biology graduates. *Assessment and Evaluation in Higher Education*, 38 (6), 737-753.

Panadero, E., & Jonsson, A. (2013). The use of scoring rubrics for formative assessment purposes revisited: A review. *Educational Research Review*, 9, 129-144.

Popham, W.J. (2007). *Classroom assessment: What teachers need to know*. 5th edition. Boston: Pearson.

Porter, A., Polikoff, M.S., Barghaus, K.M., & Yang, R. (2013). Constructing aligned assessments using automated test construction. *Educational Researcher*, 42 (8), 415-423.

Radulova, E., & Berkel, H. van (2013). Examencommissies in het hoger onderwijs. *Examens*, 4, 28-30.

Reckase, M.D. (1995). Portfolio assessment: A theoretical estimate of score reliability. *Educational Measurement: Issues and Practice*, 14 (1), 12-14.

ResearchNed (2012). *Externen in examencommissie hoger onderwijs: Een verkenning van praktische varianten*. Nijmegen: ResearchNed.

Online: www.researchned.nl/wp-content/uploads/2013/05/Eindrapport-extern-lid-examencommissies-juni-2012.pdf

Rijn, P.W. van, Béguin, A.A., & Verstralen, H.H.F.M. (2012). Educational measurement issues and implications of high stakes decision making in final examinations in secondary education in the Netherlands. *Assessment in Education: Principles, Policy & Practice*, 19 (1), 117-136.

Sadler, D.R. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, 18 (2), 119-144.

Sadler, D.R. (2009). Indeterminacy in the use of preset criteria for assessment and grading. *Assessment & Evaluation in Higher Education*, 34 (2), 159-179.

Sadler, D.R. (2010). Beyond feedback: Developing student capability in complex appraisal. *Assessment & Evaluation in Higher Education*, 35 (5), 535-550.

Sanders, P.F. (Red.) (2013). *Toetsen op School*. Cito: Arnhem.

Online: www.toetsenopschool.nl

Sanders, P.F., Dijk, P. van, Eggen, T., Otter, D. den, & Veldkamp, B. (2016). *RCEC Beoordelingssysteem voor de kwaliteit van studietoetsen en examens*. Enschede: RCEC.

Online: www.rcec.nl/Beoordelingssysteem

Schaaf, M.F. van der, Goris, M., & Stokking, K.M. (2000). *Begeleiden en beoordelen van praktische opdrachten bij Grieks en Latijn*. Utrecht: Universiteit Utrecht, ICO-ISOR.

Schaaf, M.F. van der, & Stokking, K.M. (2008). Developing and validating a design teacher portfolio assessment. *Assessment & Evaluation in Higher Education*, 33 (3), 245-262.

Schaaf, M.F. van der, Stokking, K.M., Verloop, N. (2008). Teacher beliefs and teacher behaviour in portfolio assessment. *Teaching and Teacher Education*, 24 (7), 1691-1704.

Shapley, K.S., & Bush, M.J. (1999). Developing a valid and reliable portfolio assessment in the primary grades: Building on practical experience. *Applied Measurement in Education*, 12, 111-132.

Shavelson, R.J., Baxter, G.P., & Gao, X. (1993). Sampling variability of performance assessments. *Journal of Educational Measurement*, 30, 215-232.

Sluijsmans, D.M.A. (2002). *Student involvement in assessment: The training of peer assessment skills*. Unpublished doctoral dissertation. Heerlen: Open Universiteit.

Sluijsmans, D.M.A., Joosten-ten Brinke, D., & Schilt-Mol, T. van (Red.) (2014/2015). *Kwaliteit van toetsing onder de loep: handvatten om de kwaliteit van toetsing in het hoger onderwijs te analyseren, verbeteren en borgen*. Antwerpen/Apeldoorn: Garant Uitgevers.

Sluijsmans, D.M.A., Peeters, A., Jakobs, L., & Weijzen, S. (2012). De kwaliteit van toetsing onder de loep. *Onderwijsinnovatie*, 17-25.

Sluijsmans, D.M.A., & Prins, F.J. (2006). A conceptual framework for integrating peer assessment in teacher education. *Studies in Educational Evaluation*, 32 (1), 6-22.

Stokking, K.M., & Schaaf, M.F. van der (1999). *Beoordeling van onderzoeksvaardigheden van leerlingen: Richtlijnen, alternatieven en achtergronden*. Studiehuisreeks, 27. Tilburg: Mesoconsult.

Online: www.mesoconsult.nl/wp-content/uploads/2011/02/27-Beoordelen-van-onderzoeksvaardigheden-van-leerlingen.pdf

Stokking, K.M., & Schaaf, M.F. van der (2000). *Ontwikkeling en beoordeling van onderzoeksvaardigheden: Verslag van een onderzoek in de Tweede Fase VO*. Utrecht: Universiteit Utrecht, ICO-ISOR.

Stokking, K., Schaaf, M.F. van der, Jaspers, J., & Erkens, G. (2004). Teachers' assessment of students' research skills. *British Journal of Educational Research*, 30 (1), 93-115.

Online: http://dspace.library.uu.nl/bitstream/handle/1874/25742/schaaf_04_teachersassessmentofstudentsresearchskills.pdf

Straetmans, G.J.J.M. (1985). *Evaluatie in het tandheelkundig onderwijs: Beoordelen van werkstukken en meten van probleemoplosvaardigheid*. Proefschrift. Nijmegen: Katholieke Universiteit.

Straetmans, G.J.J.M. (1998). *Toetsing van competenties*. Arnhem: Cito.

Straetmans, G.J.J.M. (2006). Assessment van competentie: Definitie, knelpunten, ontwerpbeslissingen. In: Roelofs, E., & Straetmans, G. (Red.). *Assessment in actie: Competentiebeoordeling in opleiding en beroep*. Perspectief op assessment, nr. 6. Arnhem: Cito.

Straetmans, G.J.J.M. (2006). Protocol portfolio scoring: Conclusies trekken over de verwerving van competenties op grond van een beoordelingsportfolio. In: Schramade, P. (Red.), *Handboek Effectief Opleiden*, 40/67, sectie 9.1-5.

Straetmans, G.J.J.M. (2014). Voorwaarden voor valide beoordeling van beroepsbekwaamheid: Een handreiking voor het hbo. *Examens: Tijdschrift voor de toetspraktijk*, 11 (2), 1-8.

Straetmans, G.J.J.M., Roelofs, E., & Peters, M. (2011). Vaststellen van didactische bekwaamheid tijdens de LIO-stage. *Tijdschrift voor Lerarenopleiders (VELON/VELOV)*, 32 (1), 4-11.

Straetmans, G.J.J.M., & Sanders, P.F. (2001). *Beoordelen van competenties van docenten*. EPS-reeks 05. Educatief Partnerschap.

Online: www.rcec.nl/publicaties/overige%20publicaties/EPSbrochure5_Beoordelen.pdf

SURF (2013). *VGTogether: Samenwerken in voortgangstoetsing*. Eindrapportage Projecten programma Toetsing en Toetsgestuurd Leren. Utrecht: SURF.

Online: www.surf.nl/binaries/content/assets/surf/nl/kennisbank/2013/eindrapportage-project-vgtogether.pdf

Trijffel, E. van, Pol, R.J. van de, Oostendorp, R.A.B., & Lucas, C. (2010). Interrater reliability for measurement of passive physiological movements in lower extremity joints is generally low: A systematic review. *Journal of Physiotherapy*, 56, 223-235.

Universiteit van Amsterdam (2012). *Handreiking voor examencommissies*.

Online: www.uva.nl/onderwijs/onderwijskwaliteit/organisatie/opleiding/examencommissies/examencommissies-aan-de-uva.html

Vereniging Hogescholen (2007). *HBO-Monitor 2006* (voorheen Kunstenmonitor). Den Haag: Vereniging Hogescholen/HBO-raad.

Vereniging Hogescholen (2011). *Geslaagd!: Handreiking examencommissies*. Den Haag: Vereniging Hogescholen.

Vereniging Hogescholen (2015). *Handreiking examencommissies 2015*. Den Haag: Vereniging Hogescholen.

Wiggins, G.P. (1993). *Assessing student performance: Exploring the purpose and limits of testing*. San Francisco, CA: Jossey-Bass.

Wiggins, G.P. (1998). *Educative assessment: Designing assessments to inform and improve student performance*. San Francisco, CA: Jossey-Bass.

Wools, S. (2015). *All about validity: An evaluation system for the quality of educational assessment*. Proefschrift. Enschede: Universiteit Twente.

Online: www.cito.nl/~media/cito_nl/files/over%20cito/all%20about%20validity%20-%20saskia%20wools%202015.ashx

De websites zijn geraadpleegd op 8 juni 2016.

Afkortingen

Afkortingen

AfL	assessment for learning
BKE	basiskwalificatie examinering
BKO	basiskwalificatie onderwijs
BSA	bindend studieadvies
bve	beroepsonderwijs en volwasseneneducatie
CAT	computergestuurd adaptief toetsen / computergestuurde adaptieve toets
CEFR	Common European Framework of Reference
CGO	competentiegericht opleiden
COTAN	Commissie Testaangelegenheden Nederland
DT	diagnostisch toetsen
DUO	Dienst Uitvoering Onderwijs van het ministerie van OCW
ECTS	European credit transfer and accumulation system
ERK	Europees Referentiekader
EVC	erkenning van (elders) verworven competenties
havo	hoger algemeen voortgezet onderwijs
hbo	hoger beroepsonderwijs
ho	hoger onderwijs
LIO	leraar in opleiding
mbo	middelbaar beroepsonderwijs
NVAO	Nederlands-Vlaamse Accreditatieorganisatie
NVE	Nederlandse Vereniging voor Examens
OCW	Ministerie van Onderwijs, Cultuur en Wetenschap
OGW	opbrengstgericht werken
OER	onderwijs- en examenregeling
pabo	pedagogische academie basisonderwijs
RCEC	Research Center voor Examinering en Certificering
SKE	seniorkwalificatie examinering
vwo	voortgezet wetenschappelijk onderwijs
WHW	Wet op het hoger onderwijs en wetenschappelijk onderzoek
wo	wetenschappelijk onderwijs

Index

Index

21+-toets 11

A

aanvaardbaarheid 55
absolute normering 98
absolute standaardbepaling 28, 57, 98
actualiteit 47
afstudeeronderzoek 56, 59, 60
afstudeeropdracht 52–61
afstudeerproduct 52–58
afstudeerstage 60
assessmentdomein 35, 36, 41
assessment for learning (AfL) 16, 17, 23, 94
assessmentopdracht 36, 37, 45
assessmentprogramma 40–49
associate degree 13
attitude 55, 56, 61
auditcommissie 87, 89
authenticiteit 39, 40, 41, 43, 44, 47, 53

B

bachelor-master (bama) 12–14, 25, 86, 91
bacheloropleiding 11–14, 25, 56, 86, 91
basiskwalificatie examinering (BKE) 66, 72, 78
basiskwalificatie onderwijs (BKO) 38, 72
begeleidingsgesprek 53, 54, 58, 60, 61
benchmark 54
beoordelingsaspect 44–49
beoordelingsdoel 37, 44, 54, 61, 93–94
beoordelingsmodel 52, 54, 57–58, 92
beoordelingsportfolio 39, 47, 61
beoordelingsschaal 26, 27, 37, 44, 45–46
beoordelingsstandaard 44, 48, 49, 52, 54, 56–59, 98–99
beperkte opleidingsbeoordeling 87–90
beroepsproduct (afstudeerproduct) 44, 52, 54, 57, 89
beroepstaak 32–49, 94
betrouwbaarheid 16, 26, 28, 41, 55, 61, 65, 93, 97–98, 101
betrouwbaarheidscoefficiënt 97
bewijsdossier 40, 47, 48–49

bindend studieadvies (BSA) 11, 12
buitenpromovendus 13

C

CAT (computergestuurde adaptieve toetsing) 16, 26
CEFR (Common European Framework of Reference for Languages) 12, 19
Centraal Register Hoger Onderwijs (CROHO) 86
centrale selectie 10
certificeren van examinatoren 38, 66
certificeren van studenten 10, 12–16, 23, 94
cesuur 39, 54, 57, 64, 68, 70, 73, 74, 79, 94, 98–99
checklist 37, 76, 78
civiel effect 13
classificatiesysteem (taxonomie) 24, 75, 97
classificeren van studenten 94
clustergewijze visitatie 86, 88, 89, 91
colloquium doctum (21+-toets) 11
commissie Bruijn 16, 44, 65, 92
Commissie Testaangelegenheden Nederland (COTAN) 86
compensatoire beslisregel 29, 30
compenserende model 57, 58
competentie 13, 14, 30, 32–49, 55–61, 75, 91, 94, 97
competentiebeoordeling 32–49, 56, 61, 91, 94, 97
competentieverricht opleiden (CGO) 32, 44, 49, 89
complementaire beslisregel 29, 30
computertoets 12, 16, 26, 74, 95, 100
conjunctieve beslisregel 29
conjunctieve model 57
construct 23, 56, 94, 97
constructgerichte beoordeling 56, 94, 97
contrasterende groep 98
COTAN-beoordelingssysteem 86
credit 13
criteriumgerichte interview 54
criteriumgerichte methode 57
criteriumsituatie 39–41

D

data-based decision making (DBDM) 94
decentrale selectie 11

deelcijfer 29–30
 Delftse toetspraktijk 16–19
 diagnostisch toetsen (DT) 16, 18–19, 23, 94
 DIALANG 19
 digitale toetsing 16, 17, 18, 19, 27, 74, 81, 100
 diploma 10–16, 47, 49, 64–84, 76, 86, 94
 diplomasupplement 72
 directe toelating 10
 disjunctieve model 57
 doctor 13
 doctoraalexamen 14
 doctoraat 13, 14
 doelgroep van de toets 93, 97
 Dublin-descriptoren 24, 33, 77

E

ECTS (European Credit Transfer and Accumulation System) 13, 15
 eindcijfer 18, 23, 29–30, 53, 90
 eindwerk 55, 64, 66, 70, 72, 81, 88–92
 eredoctoraat 14
 ERK (Europees referentiekader) 12, 19
 European Qualifications Framework (EQF) 77
 evaluatiebureau 88, 90, 92
 EVC (Erkenning van verworven competenties) 14, 94
 examen 12–16, 22–30, 55, 57, 64–84, 86–101
 examencommissie 14, 16, 23, 24, 28, 30, 64–84, 87, 90, 91, 92, 101
 examencommissielid 67, 82
 exameneenheid 94
 examenorganisatie 64, 66, 68, 70, 74, 75–78, 80
 examenprogramma 24, 25, 64, 66, 75–80, 94, 97, 101
 examiner 23, 26, 28, 34–49, 64–84, 90
 Expertgroep Protocol 32, 67, 75, 76, 77, 92
 expliciete opdracht 42
 extrapoleerbaarheid 47

F

formatieve toetsing 10, 16, 22, 23, 55, 59, 61, 94
 functieprofiel 38, 66, 67–68, 100, 101

G

gebruiksdoel van de toets 10–19, 22–24, 93–94, 99
 generaliseerbaarheid 41, 47
 gesloten vraag 95
 getuigschrift 64, 71–73, 81, 83
 gewogen loting 10
 groepsbeoordeling 52, 54, 58–59
 groepsopdracht 18, 52, 58, 72

H

halo-effect 38
 harde knip 12
 horn-effect 38
 huishoudelijk reglement 82–84

I

impliciete opdracht 42
 individuele beoordeling 52, 54, 58–59
 ingenieur 14
 inhoudsvaliditeit 24
 Inspectie van het Onderwijs 65–67, 86, 92
 instaptoets Wiskunde 19
 instellingsoverstijgende toetsing 16, 19, 66, 69, 70
 instellingstoets kwaliteitszorg 87
 interbeoordelaarsbetrouwbaarheid 61
 interbeoordelaarsovereenstemming 37, 38
 interuniversitaire Voortgangstoets Geneeskunde (iVTG) 15, 19, 26

J

job performance appraisal 42

K

kennis 10, 12, 13, 14, 24, 35, 36, 40, 44, 53, 55, 56, 66, 71, 91, 93
 kennistoets 29, 34, 49
 kernkwalificatie 33
 kerntaak 33, 39, 40, 44, 94, 97
 klachtencommissie 68, 70, 73–74, 81, 82, 83, 84
 kritische reflectie 32, 89, 90, 101
 kwaliteitsaspect 35, 37, 38, 45, 48, 49, 59, 75–81
 kwaliteitsnorm 33, 76, 78
 kwaliteitspiramide 75, 92, 101
 kwaliteitsstandaard 27, 86–92, 100–101

L

leerdoel 23, 24, 30, 52, 77–80, 91, 97
locatiebezoek 89–90
loting 10, 11

M

mandaat 83
master of arts 14
master of science 14
masteropleiding 12–14, 19, 23, 91
maximum performance 42
mbo-kwalificatiedossier 94
meeliftgedrag 47, 58, 79
meester in de rechten 14
meetdoel van de toets 18, 23, 55, 93, 94, 97
Meijers-criteria 24
methodemix 30, 41–43, 48, 77
methode van Angoff 57, 98
misclassificatie 29, 30, 97
mystery guest 42, 43

N

Nationaal Kenniscentrum EVC 14
natuurgetrouwheid (fidelity) 34, 35, 39, 41, 42, 43
Nederlandse Vereniging voor Examens (NVE) 38, 66, 67–68, 84
Nederlands kwalificatiekader (NLKK) 77
Nederlands-Vlaamse
Accreditatieorganisatie (NVAO) 24, 30, 32, 65, 78, 86–92, 100–101
normeringseffect 38
normgerichte methode 57
normhandhaving 65, 93, 98–99
numerus fixus 10, 11
NVAO-databank 90–92
NVE-functieprofiel 38, 66, 67–68, 100, 101
NVE-organogram examenorganisatie 66, 68

O

onderwijsdoel 32, 34, 36, 47
onderwijs- en examenregeling (OER) 14, 15, 22, 64–84, 75
Onderwijsraad 12, 13, 32, 33
onderzoeksverslag (afstudeerproduct) 52–56
opbrengstgericht werken (OGW) 16–17, 23, 94

open vraag 18, 25, 27, 95
opleidingsdoel 32, 33, 44, 49, 77
Overall Toets Accountability (OAT) 15, 19
Oxford Online Placement Test (OOPT) 12

P

parallelle toets 25, 26, 27, 98
peer beoordeling 18, 52, 54, 58, 59–60, 61
performance assessment 34, 42, 43
personal-bias error 38
persoonlijkheidseffect 38
plaatsen van studenten 10, 12, 23, 94
plaatsingstoets 12, 19
Placement Test Engels 19
praktijkopdracht 53, 55, 95, 96, 97
praktijktoets 34, 46
praktische bruikbaarheid 55
presentatie (afstudeerproduct) 52, 53, 56
prestatie criterium 33, 36–49
prestatie-indicator 44–49
prestatie standaard 39, 98
proeve van bekwaamheid 42, 43
promotie 13, 14
promovenda 13, 14
promovendus 13, 14

Q

Questionmark Perception 26

R

RCEC-beoordelingssysteem 27, 86, 93–101
reflectiegesprek (afstudeerproduct) 52
relatieve normering 15, 98
relatieve standaardbepaling 28, 98
representativiteit 24, 34, 40, 93, 97
Research Center voor Examinering en Certificering (RCEC) 86, 93
respons 42
rijexamen 43
rubriek (rubric) 26, 27, 28, 52, 59

S

schaalpunt 46
scorerubriek 36, 37, 39, 44, 46
scoretabel 28, 29

selecteren van studenten 10–12, 23, 94
seniorkwalificatie examinering (SKE) 66, 78, 82
stage 25, 40, 42, 43, 52, 53, 60, 71, 72, 76
stagebegeleider 60
standaard 28, 39, 48–49, 52, 54, 56–59, 98
standaardbepaling 28, 57, 93, 98–99
standaardmeetfout 30, 46, 97, 99
stimulus 42, 43
studentgecentreerde benadering 57
studiepunt 11, 12, 13, 15, 71, 76
summatieve toetsing 10, 12, 16, 22, 23, 27, 55, 59, 61
surveillant 27, 69, 100

T

taakgerichte beoordeling 56
taaksituatie 32–49
taakspecifieke prestatie 36, 37, 44, 45, 46, 47
taxonomie van Bloom 24
TestVision 26
titel in hoger onderwijs 13, 14
toelating 10, 11, 13
toetsbaar leerdoel 97
toetsbeleid 69, 75, 76, 78, 80, 91
toetscommissie 64, 69, 76, 77, 78, 80, 91, 101
toetsconstructie 22–30, 39, 75, 79
toetscyclus 75, 76, 78, 79
toetsdoel 10–19, 22–24, 93–94
toetsen van het leren 10, 94
toetsen voor het leren 10
toetsgecentreerde methode 57
toetsmatrijs 22, 24–25, 35, 75, 76, 79, 80, 94
typical performance 42

V

vaardigheid 10, 12, 13, 14, 23, 24, 36, 44, 53, 55, 56, 61, 66, 71, 77, 91, 93
validiteit 26, 28, 33, 34, 41, 55, 61, 65, 101
Vereniging Hogescholen 32, 52, 65, 66, 67, 81
vergelijkingseffect 38
visitatie 24, 28, 88–92
visitatiegroep 89
visitatiepanel 88–90, 101

volledigheid (comprehensiveness) 34, 35
voortgangstoets 15, 16, 19, 26
vrijstelling 14, 68, 71, 73, 83

W

werkproces 94
wetenschappelijk onderzoek 14, 52
Wet op het hoger onderwijs en wetenschappelijk onderzoek (WHW) 14, 64–84
WISCAT-pabo 26

Z

zak/slaaggrens 22, 27–28
zelfbeoordeling 52, 59–60

Cito

Corporate

Toetsen op School

Hoger onderwijs

Piet Sanders (redactie)

**Bedankt voor het lezen van Toetsen op School
hoger onderwijs.**

Wilt u weten wat Cito voor u kan betekenen?
Bel met een van onze adviseurs: (026) 352 12 81
of mail naar citoconsult@cito.nl

Zij helpen u graag verder!

Cito: dé expert in examinering

- ✓ Al bijna 50 jaar nationaal marktleider op het gebied van beoordeling en toetsconstructie
- ✓ Internationaal toonaangevend psychometrisch instituut
- ✓ Wereldwijd erkend en behorend tot de top 5 van testing and assessment companies
- ✓ Gespecialiseerd aanbod toetsen en examens voor het onderwijs en het bedrijfsleven
- ✓ Deskundige en ervaren trainers

Cito helpt mensen richting te geven aan hun toekomst

Wij geloven in de ontwikkeling van mensen. Daar hoort inzicht in mogelijkheden en in grenzen bij. Door kennis, vaardigheden en competenties objectief meetbaar te maken en de ontwikkeling ervan te volgen, kun je het beste uit jezelf halen, verantwoorde keuzes maken en beter richting geven aan je toekomst. Wij dragen daaraan bij door wereldwijd, voor het onderwijs én het bedrijfsleven, werk te maken van goed en eerlijk toetsen. | www.cito.nl

Fotografie: Ron Steemers

