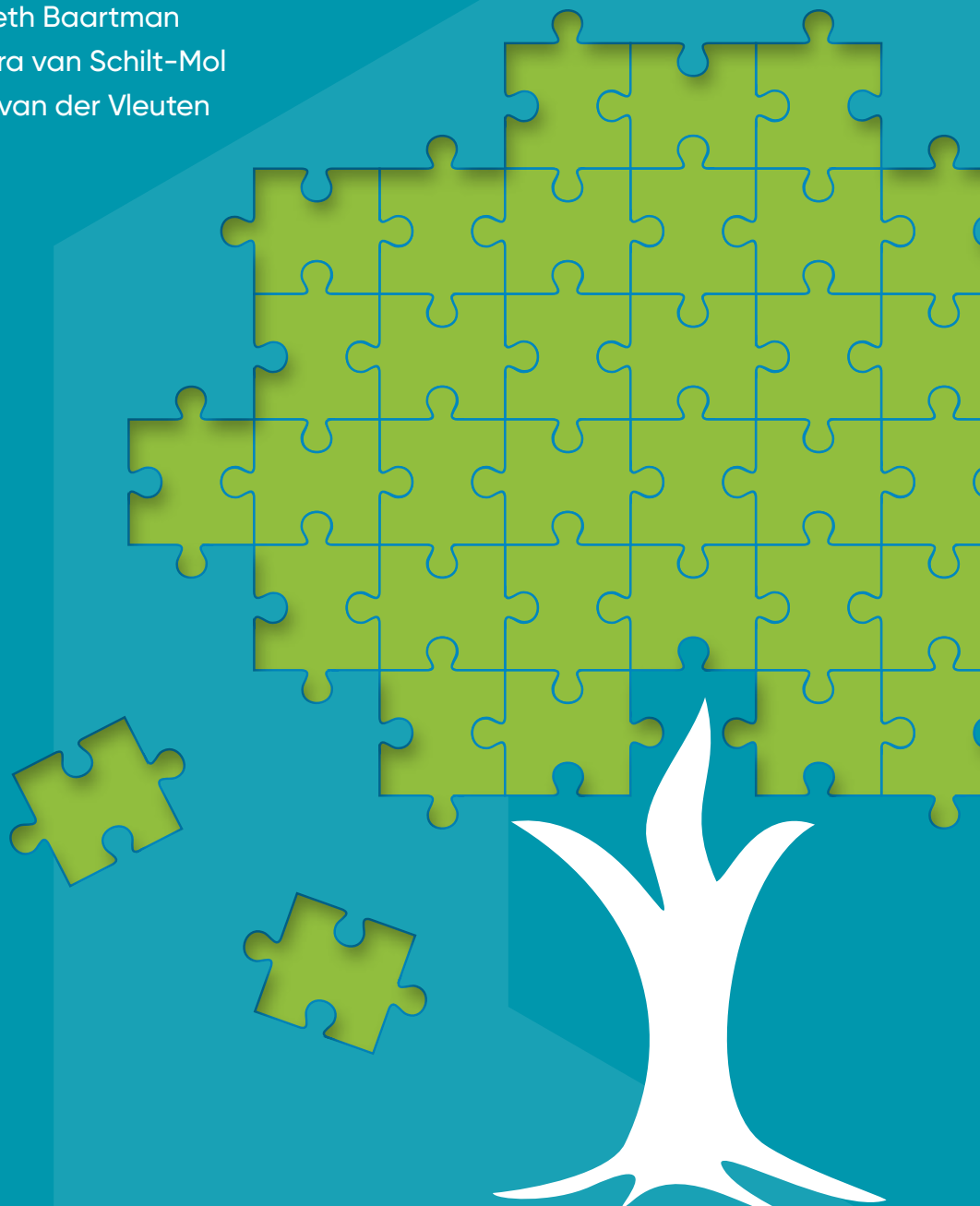


Liesbeth Baartman
Tamara van Schilt-Mol
Cees van der Vleuten

Boom



PROGRAMMATISCH TOETSEN

Voorbeelden en ervaringen
uit de praktijk

Programmatisch toetsen

Programmatisch toetsen

Voorbeelden en ervaringen uit de praktijk

Liesbeth Baartman
Tamara van Schilt-Mol
Cees van der Vleuten

Boom

Opmaak binnenwerk: Textcetera, Den Haag
Basisontwerp omslag: Dog & Pony, Amsterdam
Omslagontwerp: Haagsblauw, Den Haag

© Liesbeth Baartman, Tamara van Schilt-Mol en Cees van der Vleuten & Boom uitgevers Amsterdam,
2020

Behoudens de in of krachtens de Auteurswet gestelde uitzonderingen mag niets uit deze uitgave worden verveelvoudigd, opgeslagen in een geautomatiseerd gegevensbestand, of openbaar gemaakt, in enige vorm of op enige wijze, hetzij elektronisch, mechanisch, door fotokopieën, opnamen of enige andere manier, zonder voorafgaande schriftelijke toestemming van de uitgever.

Voor zover het maken van reprografische verveelvoudigingen uit deze uitgave is toegestaan op grond van artikel 16h Auteurswet dient men de daarvoor wettelijk verschuldigde vergoedingen te voldoen aan de Stichting Reprorecht (Postbus 3051, 2130 KB Hoofddorp, www.reprorecht.nl). Voor het overnemen van (een) gedeelte(n) uit deze uitgave in bloemlezingen, readers en andere compilatiewerken (art. 16 Auteurswet) kan men zich wenden tot de Stichting PRO (Stichting Publicatie- en Reproductierechten Organisatie, Postbus 3060, 2130 KB Hoofddorp, www.stichting-pro.nl).

No part of this book may be reproduced in any form, by print, photoprint, microfilm or any other means without written permission from the publisher.

ISBN 978 90 2443 409 1
ISBN 978 90 2443 423 7 (e-book)
NUR 841

www.boomhogeronderwijs.nl

Inhoudsopgave

Inleiding		9
1	Programmatisch toetsen als concept <i>Cees van der Vleuten, Liesbeth Baartman en Tamara van Schilt-Mol</i>	15
1.1	Inzicht in de ontwikkeling van de student ontstaat door een mix van verschillende datapunten	17
1.2	Elk datapunt is feedbackgericht en kent geen zak-/slaagbeslissing	18
1.3	De leeropbrengsten vormen de ruggengraat van het toetsprogramma	19
1.4	Er is een constante dialoog over het gebruik van feedback voor zelfsturing	21
1.5	Het aantal datapunten en de zwaarte van de beslissing zijn proportioneel aan elkaar gerelateerd	22
1.6	De zwaarte van een beslissing is leidend voor de hoeveelheid benodigde beoordelaarsexpertise	23
2	Programmatisch toetsen van klinische vaardigheden <i>Tamara van Schilt-Mol, Femke Jongen-Hermus, Marjanneke Krauss, Liesbeth Baartman en Cees van der Vleuten</i>	27
3	Programmatisch toetsen tijdens de stage bij de Master Voeding en Diëtetiek <i>Liesbeth Baartman, Janica Jamieson, Gemma Jenkins, Tamara van Schilt-Mol en Cees van der Vleuten</i>	43
4	Programmatisch toetsen tijdens de coschappen bij de Master Diergeneeskunde <i>Liesbeth Baartman, Harold Bok, Tamara van Schilt-Mol en Cees van der Vleuten</i>	59
5	Programmatisch toetsen om te komen tot een lerende cultuur bij de Bachelor Ergotherapie <i>Tamara van Schilt-Mol, Marleen Kaijen, Liesbeth Baartman en Cees van der Vleuten</i>	75
6	Niveau-assessments bij de Bachelor Communicatie <i>Liesbeth Baartman, Wendy Peeters, Tamara van Schilt-Mol en Cees van der Vleuten</i>	87

7	Eigentijds beoordelen en beslissen bij de Master Ontwerpen van Eigentijds Leren	107
	<i>Tamara van Schilt-Mol, Marijke van Vijfeijken, Liesbeth Baartman en Cees van der Vleuten</i>	
8	Programmatisch toetsen bij de Basisopleiding Didactiek voor hbo-docenten	121
	<i>Tamara van Schilt-Mol, Marieke Kloek, Ton de Kraay, Liesbeth Baartman en Cees van der Vleuten</i>	
9	Eerste stappen richting programmatisch toetsen bij de iPabo	133
	<i>Liesbeth Baartman, Monique Leijgraaf, Tamara van Schilt-Mol, Cees van der Vleuten</i>	
10	Didactische vaardigheden: wat vraagt programmatisch toetsen van docenten?	147
	<i>Tamara van Schilt-Mol, Wessel Peeters, Sasja de Vries, Liesbeth Baartman en Cees van der Vleuten</i>	
11	Programmatisch toetsen als concept in de praktijk. Een terugblik op de praktijkvoorbeelden in dit boek	159
	<i>Liesbeth Baartman, Tamara van Schilt-Mol en Cees van der Vleuten</i>	
	Literatuur	179
	Over de auteurs	183



Inleiding

Waarom dit boek?

Toetsen zijn een essentieel onderdeel van het onderwijs. Hoewel er geen discussie bestaat over het belang van toetsen, is er wel in groeiende mate onvrede over de wijze waarop toetsen worden gebruikt. Die onvrede wordt niet zozeer beïnvloed door de toetsen zelf, maar vooral door de consequenties en effecten van toetsen op studenten. Studenten hebben vaak het gevoel een marathon te lopen waarbij ze constant over allerlei hordes moeten springen, te leren voor een toets om het geleerde vervolgens te vergeten, en ze hebben vaak moeite om de regie te nemen over hun leerproces. Docenten besteden veel tijd aan het geven van feedback, maar vinden dat studenten deze feedback te weinig oppakken ten behoeve van hun leerproces. Zij hebben het gevoel veel tijd kwijt te zijn aan toetsen, en die tijd niet te kunnen besteden aan onderwijs. Terwijl toetsing juist een integraal onderdeel zou moeten zijn van onderwijs! Juist toetsen – in de brede zin van het woord: alle leeractiviteiten die informatie opleveren over het leren van de student – kunnen studenten houvast geven over hun ontwikkeling ('waar sta ik en wat wordt mijn volgende stap?') en kunnen docenten informatie geven over wat wel en niet werkt in hun onderwijs, en hoe zij kunnen inspelen op de ontwikkeling van studenten en hierin maatwerk kunnen leveren.

De conventionele manier van toetsen lijkt dan ook niet (meer) aan te sluiten bij heersende opvattingen over onderwijs en de beoogde leerprocessen in het beroepsonderwijs, gericht op het opleiden van professionals die kunnen omgaan met de complexe dynamiek van de beroepspraktijk. Oftewel, er is geen sprake van *constructive alignment*: aansluiting tussen de beoogde leerresultaten, de leer- en onderwijsactiviteiten, en de toetsactiviteiten. Toetsing en onderwijs of leren worden binnen de conventionele manier van toetsen vaak los van elkaar gezien – waarbij toetsing primair wordt ingezet als afsluiting van het leerproces – terwijl juist de verbinding van het leren aan onderwijs- en toetsactiviteiten van waarde kan zijn voor de ontwikkeling van de student. Steeds vaker beschouwen docenten toetsen – wederom in de brede zin van het woord – als integraal onderdeel van het onderwijs, waardoor het mogelijk wordt om de student te begeleiden in zijn¹ ontwikkeling op basis van de informatie uit de toetsen, en de student zelf de regie over zijn ontwikkeling te laten nemen. Daarbij wordt voortdurend de balans opgemaakt om te bepalen waar de student staat ten opzichte van de te bereiken leerresultaten en welke stappen hij nog te

1 We gebruiken in dit boek omwille van de eenvoud de mannelijke aanspreekvorm, maar vanzelfsprekend worden studenten en docenten van alle geslachten bedoeld.

zetten heeft. Docenten willen toetsresultaten met andere woorden steeds vaker inzetten om het ontwikkelingsproces van de student te sturen en te stimuleren, en tegelijkertijd valide en betrouwbare beslissingen nemen over voortgang en het toekennen van studiepunten.

Vanuit deze ontwikkeling hebben wij als auteurs van dit boek een groeiende aandacht ervaren voor het concept programmatisch toetsen. De afgelopen jaren kregen wij steeds meer vragen over programmatisch toetsen en verzoeken om lezingen, workshops, professionaliseringstrajecten, advisering, praktijkgericht onderzoek et cetera. Veel vragen die we kregen zijn heel praktisch, bijvoorbeeld: Hoe kan ik programmatisch toetsen inrichten? Hoe kies ik geschikte datapunten? Is programmatisch toetsen ook geschikt voor een bacheloropleiding? Kan ik met iemand praten die al ervaring heeft in de praktijk? Wat vinden studenten ervan? Uit al deze praktijkvragen ontstond het idee om praktijkvoorbeelden van programmatisch toetsen te verzamelen en te bundelen.

In dit boek willen we handvatten geven voor opleidingen die overwegen om te gaan werken volgens de uitgangspunten van programmatisch toetsen. Maar we willen ook de diversiteit aan praktijkvoorbeelden laten zien, waaruit blijkt dat de principes van programmatisch toetsen een diversiteit aan verschijningsvormen kennen in de praktijk. Ook gaan we in op het proces dat de opleidingen in dit boek hebben doorlopen tijdens het herontwerp en de implementatie van programmatisch toetsen, en hoe zij, ook na implementatie, stappen blijven zetten in de doorontwikkeling van programmatisch toetsen, door te evalueren en te leren van hun ervaringen in de praktijk.

Wat vind je in dit boek?

In het eerste hoofdstuk van dit boek vind je een introductie op het concept programmatisch toetsen. In dit hoofdstuk leggen we uit wat de belangrijkste uitgangspunten van programmatisch toetsen zijn, maar ook waarom programmatisch toetsen als toetsconcept is ontwikkeld. In dit hoofdstuk vind je ook een uitleg van de belangrijkste begrippen bij programmatisch toetsen, zoals datapunten, low stakes en high stakes beslissingen en de ruggengraat van beoogde leeropbrengsten die de basis vormt voor het ontwerp.

Na het eerste hoofdstuk volgen de praktijkvoorbeelden: opleidingen die programmatisch toetsen hebben geïmplementeerd of bezig zijn met een herontwerp waarin programmatisch toetsen een belangrijke plek heeft. Verschillende praktijkvoorbeelden passeren de revue (zie Tabel 1.1). Een aantal opleidingen heeft programmatisch toetsen geïmplementeerd in een gedeelte van de opleiding; partiële implementatie. Dit zijn het Skillslab van de Bachelor Geneeskunde van de Universiteit Maastricht, de Master Voeding en Diëtetiek (Edith Cowan University, Australië) en de Master Diergeneeskunde (Universiteit Utrecht), waar programmatisch toetsen is geïmplementeerd tijdens het werkplekleren (de stages en de coschappen). In de hoofdstukken 4, 5 en 6 komen de

ontwerpers en docenten van drie opleidingen aan het woord die de uitgangspunten van programmatisch toetsen hebben geïmplementeerd in hun gehele curriculum. De docenten van de Bachelor Ergotherapie van de Hogeschool van Arnhem en Nijmegen, en de Bachelor Communicatie van Hogeschool Utrecht hebben al enkele jaren praktijkervaring met programmatisch toetsen en vertellen ook over de ervaringen van hun studenten en docenten. De opleiders van de Master Ontwerpen van Eigentijds Leren (Hogeschool van Arnhem en Nijmegen) vertellen over hun ervaringen bij het ontwerpproces, waarin programmatisch toetsen een belangrijk uitgangspunt was. De Basisopleiding Didactiek van Hogeschool Rotterdam (hoofdstuk 8) is gericht op een andere doelgroep dan bachelor- of masterstudenten, namelijk hbo-docenten. Ook hier hebben de docenten al enkele jaren ervaring met programmatisch toetsen. Dat geldt ook voor de docenten van de Master Ontwerpen van Eigentijds Leren, waarbij de masterstudenten allen werkzaam zijn in het onderwijs. Hier is de dubbele lading interessant: de deelnemende (hbo-)docenten maken kennis met programmatisch toetsen in een professionaliseringstraject en kunnen deze ervaringen weer mee terug nemen naar hun eigen opleiding. In hoofdstuk 9 komen opleiders van de iPabo aan het woord. Zij werken aan meer geïntegreerde toetsing van verschillende schoolvakken, als een tussenstap richting programmatisch toetsen. Ook wordt gewerkt aan het professionaliseren van docenten in het geven van feedback, als belangrijke voorwaarde om programmatisch toetsen later goed te kunnen implementeren. Hoofdstuk 10, ten slotte, sluit hierop aan. In dit hoofdstuk vertellen twee onderwijskundigen over de professionalisering van docenten die programmatisch toetsen vraagt, en de ervaringen die zij hebben opgedaan bij twee opleidingen van de Hogeschool van Arnhem en Nijmegen.

We sluiten dit boek af met een reflectiehoofdstuk, waarin we terugblikken op de negen praktijkvoorbeelden in dit boek. Aan de hand van de uitgangspunten van programmatisch toetsen (zie hoofdstuk 1) vergelijken we de verschillende praktijkvoorbeelden. Welke keuzes hebben de opleidingen gemaakt bij elk uitgangspunt? Welke verschillen en overeenkomsten zien we? En welke lessen kunnen we daaruit leren?

Tabel 1.1 Overzicht van de praktijkvoorbeelden in dit boek

Hfs.	Studie	Partiële / gehele implementatie	Bachelor- / masteropleiding	Stadium van implementatie
2	Skillslab Bachelor Geneeskunde, Universiteit Maastricht	Partiële implementatie	Bachelor	Geïmplementeerd, aantal jaren praktijkervaring
3	Stages Master Voeding en Diëtetiek, Edith Cowan University	Partiële implementatie	Master	Geïmplementeerd, aantal jaren praktijkervaring
4	Coschappen Master Diergeneeskunde, Universiteit Utrecht	Partiële implementatie	Master	Geïmplementeerd, aantal jaren praktijkervaring
5	Bachelor Ergotherapie, Hogeschool Arnhem en Nijmegen	Gehele implementatie	Bachelor	Geïmplementeerd, aantal jaren praktijkervaring
6	Bachelor Communicatie, Hogeschool Utrecht	Gehele implementatie	Bachelor	Geïmplementeerd, aantal jaren praktijkervaring
7	Master Ontwerpen van Eigentijds Leren, Hogeschool Arnhem en Nijmegen	Gehele implementatie	Master	Ontwerp afgerond, eerste praktijkervaringen
8	Basisopleiding Didactiek, Hogeschool Rotterdam	Gehele implementatie	Professionalisering hbo-docenten	Geïmplementeerd, enkele jaren praktijkervaring
9	Bachelor leraar basisonderwijs, Hogeschool iPabo	Gehele implementatie	Bachelor	Ontwerpfase
10	Het belang van professionalisering van docenten bij programmatisch toetsen		Professionalisering hbo-docenten	

Alle praktijkvoorbeelden in dit boek zijn gebaseerd op interviews met een of meerdere mensen van de betreffende opleiding. In de hoofdstukken zijn dan ook veel citaten uit de interviews te vinden. De mensen van de opleidingen vertellen in het hoofdstuk het eigen verhaal van hun opleiding. Dat zijn heel eerlijke verhalen, met daarin ook de hobbels en valkuilen die de betrokkenen tegenkwamen tijdens het ontwerpproces, bijvoorbeeld bij het meekrijgen van collega's of de examencommissie, en tijdens de daadwerkelijke uitvoering in de praktijk. Op basis van de interviews hebben wij – Liesbeth Baartman, Tamara van Schilt-Mol en Cees van der Vleuten – een eerste versie van het hoofdstuk van elk praktijkvoorbeeld geschreven. Deze eerste versie is vervolgens aangevuld en gecorrigeerd door de geïnterviewden, en soms is een tweede aanvullend

interview gehouden. Alle geïnterviewden hebben actief meegeschreven aan hun eigen hoofdstuk en zijn ook coauteur van hun hoofdstuk. Zij vertellen hun eigen verhaal over hoe zij programmatisch toetsen passend hebben gemaakt voor hun eigen opleiding.

In de interviews met de betrokkenen van de opleidingen kwam steeds een aantal hoofdonderwerpen of hoofdvragen terug. Die hoofdvragen zijn ook terug te vinden in de opzet van elk praktijkhoofdstuk. Elk hoofdstuk begint met een korte schets van de betreffende opleiding en de belangrijkste ontwikkelingen daar. Vervolgens zijn we in de interviews ingegaan op de volgende vragen:

- Waarom heeft de opleiding gekozen voor programmatisch toetsen?
- Hoe is het toetsprogramma ingericht?
- Wat is de ruggengraat van het toetsprogramma?
- Uit welke datapunten bestaat het toetsprogramma?
- Hoe en wanneer worden beslissingen genomen over studenten?
- Hoe is het portfolio opgebouwd?
- Wat zijn de praktijkervaringen van docenten en studenten met programmatisch toetsen?
- Wat doet programmatisch toetsen met het leerproces van de studenten?
- Hoe verliep het ontwerp- en implementatietraject?

We hopen met dit boek te illustreren dat programmatisch toetsen een toets-CONCEPT is, een opvatting of visie op toetsing. Een toetsconcept is geen toetsRECEPT. Dit boek biedt dus geen kant-en-klare oplossingen of handvatten om programmatisch toetsen vorm te geven. Of zoals een van de geïnterviewden voor dit boek zei: 'Programmatisch toetsen is geen Billykast van Ikea.' Er bestaat geen handleiding over hoe programmatisch toetsen stap voor stap te implementeren. Wel geven de uitgangspunten van programmatisch toetsen richting aan de ontwerpkeuzes die je als opleiding kunt maken, passend bij je eigen studenten, je eigen visie op leren, en het beroep waartoe je opleidt. We hopen dat de verhalen en ervaringen van de opleidingen die deze ontwerpkeuzes al hebben gemaakt inspiratie geven om voor de eigen opleiding het eigen verhaal te schrijven, om een eigen passende invulling te geven aan de uitgangspunten van programmatisch toetsen.



Programmatisch toetsen als concept

1

De gebruikelijke manier van toetsing in de onderwijspraktijk is eenvoudig. Aan het einde van een onderwijsperiode worden een of meerdere toetsen afgenomen en daarmee wordt die onderwijsperiode afgesloten. Scoren studenten hoger dan de minimumgrens, dan slagen zij en ontvangen ze de studiepunten die voor het vak of de periode staan. Wanneer ze onder de minimumgrens scoren, moeten zij de toets(en) herkansen. Meestal wordt de waardering in een schoolcijfer uitgedrukt. Dit type toetsing past goed bij een onderwijsconcept dat gericht is op beheersingsleren: we gaan er dan van uit dat studenten die slagen de stof tot in de eeuwigheid beheersen.

Modernere onderwijsconcepten zijn echter meer constructivistisch van aard. Lerenden creëren zelf hun kennis en kunde, samen met anderen. Meestal gebeurt dit met leertaken die uitdagender zijn dan bij het beheersingsleren. Die taken zijn bovendien nuttiger voor hun opleiding tot professional. De leertaken zijn namelijk vaak afgeleid van de taken binnen het beroepenveld, doordat er een vertaling plaatsvindt van concrete beroepstaken naar een onderwijsomgeving. Beroepsbeoefenaren uit de praktijk hebben vaak een directe rol, bijvoorbeeld als opdrachtgever, expert of medebeoordelaar. Daardoor gaat binnen het onderwijs veel aandacht uit naar de transfer van het geleerde naar de beroepspraktijk. Deze zogeheten hele taakbenadering oftewel holistische benadering, waarbij gekeken wordt naar de beheersing op de hele taak en niet naar de beheersing van afzonderlijke onderdelen binnen de taak, staat dan ook centraal in veel instructietheorieën (Van Merriënboer & Kirschner, 2007).

De loopbaan van toekomstige professionals staat in het teken van een leven lang leren. Immers, het is goed mogelijk dat studenten in de toekomst functies zullen uitoefenen waarvan we het bestaan nu nog niet kennen. Om dit leven lang leren te bevorderen, is er in moderne onderwijsconcepten veel aandacht voor zelfsturing van studenten. Zelfreflectie en planning van de studie zijn belangrijke onderdelen in dit type onderwijs, aangezien deze vaardigheden ook in de latere beroepspraktijk belangrijk zijn. Tevens is er in het beroepenveld – en daarmee ook in beroepsopleidingen – veel aandacht voor complexe vaardigheden zoals bijvoorbeeld samenwerken, communiceren, professionaliteit en leiderschap. Ook deze complexe vaardigheden moeten studenten beheersen, om goed te kunnen functioneren in hun toekomstige beroep. Die vaardigheden zijn echter alleen te beoordelen middels het evalueren van het gedrag in de praktijk. Bovendien is het niet goed mogelijk om deze complexe

vaardigheden binnen het tijdsbestek van één onderwijsperiode te ontwikkelen en te beoordelen. Het past dan ook beter om de ontwikkeling van complexe vaardigheden te volgen in een meer longitudinale manier van beoordelen.

Moderne opleidingen maken dus veel gebruik van leertaken die direct ontleend zijn aan het toekomstige beroep van studenten, waarbij er veel aandacht is voor complexe vaardigheden en zelfsturing door de student. Het conventionele model van toetsing sluit echter slecht aan bij deze moderne visie op leren. Men spreekt daardoor in dit kader ook wel over *constructive misalignment*: het onderwijs, het beoogde leerproces van de student, de begeleiding en de toetsing sluiten niet goed op elkaar aan. Zowel studenten als docenten beschrijven de toetsing vaak als een hordenloop. Dat leidt meer dan eens tot het terugvallen op oppervlakkige leerstrategieën, gericht op reproductie van informatie (Cilliers et al., 2012). Evenmin is er voldoende aandacht voor groei of ontwikkeling, al helemaal niet over verschillende onderwijsperioden heen. Na het voltooien van alle onderwijseenheden en de daarbij behorende toetsen, sluiten studenten hun opleiding af. De conventionele wijze van toetsing is daarmee eigenlijk een atomistische benadering: vanuit allemaal korte momentopnamen over de student, trekken docenten de conclusie dat de student beroepsbekwaam is. Dit leidt regelmatig tot vragen en twijfels bij docenten, zowel over het kennisniveau van studenten als over hun geschiktheid voor het beroep. Voor veel docenten is een herkenbare twijfel: ‘Deze student blijft maar struikelen over deze toets, terwijl ik hem in totaliteit toch een prima beroepsbeoefenaar vind.’

Programmatisch toetsen is een manier van beoordelen die aansluit bij een constructivistische benadering van onderwijs. Het is een concept met een aantal uitgangspunten om de toetsing vorm te geven en dat betekent dat het nadrukkelijk géén recept of voorschrift is. Een van de doelen van dit boek is dan ook om te laten zien dat er vele manifestaties van programmatisch toetsen kunnen zijn. Programmatisch toetsen komt voort uit de ervaring van velen dat opleidingen met uitsluitend summatieve toetsen de plank vaak misslaan. Nu is elke vorm van toetsing een compromis tussen het wenselijke en het haalbare. In de ideale situatie zouden docenten elke student gedurende langere tijd in de beroepspraktijk willen observeren. Dat is natuurlijk onhaalbaar. Om toch de beroepspraktijk te benaderen, kiezen ze vaak voor het afnemen van een individuele, mondelinge en/of schriftelijke toets, in de vorm van een reflectieverslag over een stageperiode. Dat roept echter de volgende vraag op over de validiteit van het onderzoek en de opgedane ervaring: meet je zo echt gedrag in de beroepspraktijk? Ook het punt van betrouwbaarheid leidt tot vragen, zoals: is de student in staat om eerlijk over zichzelf te rapporteren? En wat is dat ‘eerlijk’ dan?

Afhankelijk van het doel van de toets of beoordeling en de context waarin deze wordt afgenomen, maak je andere afwegingen. De vraag is dus wanneer je wat wilt optimaliseren met welk doel voor ogen. Programmatisch toetsen maakt

hierin scherpe keuzes. Een belangrijk uitgangspunt van programmatisch toetsen is dat het geheel aan toetsen – dus anders dan individuele toetsen – wel van hoge kwaliteit (betrouwbaar, valide) kan zijn. In programmatisch toetsen wordt afgestapt van het “afvinken” van onderdelen in individuele toetsen, en wordt gekeken naar de ontwikkeling van de student naar het beoogde eindniveau als geheel. Programmatisch toetsen is dus een holistische benadering: er wordt niet afgevinkt, maar er wordt gekeken naar de gehele ontwikkeling van een lerende. De benadering van programmatisch toetsen is daarmee een holistische en kent de volgende uitgangspunten:²

1. Inzicht in de ontwikkeling van de student ontstaat door een mix van verschillende datapunten.
2. Elk datapunt is feedbackgericht en kent geen zak-/slaagbeslissing.
3. De leeropbrengsten vormen de ruggengraat van het toetsprogramma.
4. Er is een constante dialoog over het gebruik van feedback voor zelfsturing.
5. Het aantal datapunten en de zwaarte van de beslissing zijn proportioneel aan elkaar gerelateerd.
6. De zwaarte van een beslissing is leidend voor de hoeveelheid benodigde beoordelaarsexpertise.

Hieronder gaan we wat dieper op de afzonderlijke uitgangspunten in.

1.1 Inzicht in de ontwikkeling van de student ontstaat door een mix van verschillende datapunten

Bij programmatisch toetsen wordt de ontwikkeling van studenten gevolgd door middel van zogenoemde datapunten. Gemaakte opdrachten, beroepsproducten, presentaties, feedback van studenten, medestudenten, cliënten of opdrachtgevers, maar bijvoorbeeld ook uitkomsten van kennistoetsen fungeren als informatiebrokjes over het leerproces. Studenten krijgen feedback bij elk datapunt, zodat zij stil kunnen staan bij hun leerproces (‘Waar sta ik nu?’) en de vervolgstappen die nodig zijn (‘Waar moet ik heen en hoe kom ik daar?’).

Elke vorm van beoordeling kan als datapunt worden benut. Mogelijke datapunten zijn:

- een score op een meerkeuzetoets;
- een reflectieverslag van een ervaringsgebeurtenis;
- een video met opgenomen gedrag;
- een peer-evaluatie;

2 Meer over de uitgevoerde wetenschappelijke onderzoeken die ten grondslag liggen aan deze uitgangspunten zijn te vinden in onder meer Van der Vleuten et al. (2005), Driessen et al. (2007), Van der Vleuten et al. (2015). Een aantal van deze publicaties zijn te vinden op www.ceesvandervleuten.com.

- een 360-gradenbeoordeling;
- een onderzoeksverslag;
- een beroepstaak.

Een datapunt kan ook de leertaak zelf betreffen, zoals het maken van een schrijfpodricht of het uitvoeren van een consult. Elk datapunt staat volledig in het teken van het leerproces van de student. De keuze voor de vorm van een bepaald datapunt moet dus ook ten dienste staan van het leren en effectief zijn voor het leerproces. Een kennistoets lokt bijvoorbeeld ander leergedrag uit dan een praktijkopdracht.

Bij programmatisch toetsen wordt het beoogde effect bekeken per datapunt, maar ook in afstemming met het toetsprogramma als geheel. Er moet een bewuste planning van datapunten worden gemaakt, in het licht van het programma als geheel. Want ook in de beroepspraktijk moeten professionals op het ene moment schrijven, op een ander moment zich verbaal uitdrukken of debatteren, schematiseren, et cetera. Geen enkele toets kan dekkend zijn en alle vaardigheden beoordelen, maar het programma als geheel kan dat wel. Er bestaan dus ook geen voorkeurstypen van datapunten of toetsmethoden. Elke methode – gestandaardiseerd, ongestandaardiseerd, objectief, subjectief, authentiek of niet – is geoorloofd, zolang het datapunt maar aansluit op het beoogde leerproces en in staat is rijke informatie op te leveren waar de student baat bij heeft. Wel is het zo dat observaties, professionele oordelen en feedback van betrokkenen zoals cliënten of opdrachtgevers onontbeerlijk zijn voor het beoordelen van gedrag. Professionele oordelen kunnen komen van bijvoorbeeld docenten, medestudenten, collega's en cliënten. Die oordelen zijn per definitie subjectief, maar een groot aantal subjectieve oordelen samen levert wel een robuuste beoordeling van de student op.

1.2 Elk datapunt is feedbackgericht en kent geen zak-/slaagbeslissing

Belangrijk bij programmatisch toetsen is dat op basis van één enkel datapunt geen zak-/slaagbeslissing genomen kan worden. Aan één datapunt zijn dus geen studiepunten gekoppeld. Een datapunt is immers slechts een momentopname en is onvoldoende betrouwbaar om zo'n beslissing te rechtvaardigen. Het is te vergelijken met een enkele pixel van een foto, dat geeft geen beeld van de hele foto. Ook een enkel datapunt geeft nog geen beeld van het daadwerkelijke kennen en kunnen van de student.

Een enkel datapunt is wél heel geschikt voor feedback. Die feedback kan kwantitatief zijn, en bijvoorbeeld bestaan uit een overzicht van deelscores op een meerkeuzetoets, waarbij per score wordt aangegeven hoe de student scoort ten opzichte van de gehele groep. Feedback kan ook kwalitatief van aard zijn.

Kwalitatieve feedback is veelal narratief: de feedbackgever drukt de eigen bevindingen uit in woorden. Vooral voor het leren van complexe vaardigheden is kwalitatieve informatie bijzonder nuttig, en veelal betekenisvoller dan kwantitatieve informatie. Schoolcijfers zijn heel beperkte informatiedragers, aangezien ze geen richting geven aan het verbeteren van prestaties of leertaken. De kunst is om bij individuele datapunten zo veel mogelijk informatie te geven waarmee het leren kan worden bevorderd. Bij programmatisch toetsen wordt dus maximaal ingezet op de *leerwaarde* van elk datapunt.

Omdat aan een enkel datapunt geen beslissing hangt, kan de wijze van beoordeling heel subjectief zijn. Dat is prettig, zeker als het om het beoordelen van complexe vaardigheden gaat. Doordat er geen zwaarwegende consequenties aan de beoordeling van de datapunten hangen, is het eenvoudiger voor begeleiders om eerlijk te zijn in hun terugkoppeling. Die terugkoppeling is vooral een vorm van coaching, waarin de begeleider aangeeft hoe hij de prestatie van de student waardeert ('Waar sta je nu?') en met de student vervolgstappen bedenkt ('Waar moet je heen en hoe kom je daar?'). Het is bovendien onnodig om beoordelaars te kalibreren naar één standaard, waarbij alle studenten exact op dezelfde wijze worden beoordeeld – als dat al mogelijk zou zijn. Vanuit een perspectief op leren is het immers juist prettig om verschillende visies te oogsten. Door de angst van de zak-/slaagbeslissing uit een datapunt te halen, ontstaat er veel meer vrijheid voor het geven van feedback, voor verschillende perspectieven op kwaliteit, en voor de dialoog met de studenten over het verbeteren van de kwaliteit van hun werk.

1.3 De leeropbrengsten vormen de ruggengraat van het toetsprogramma

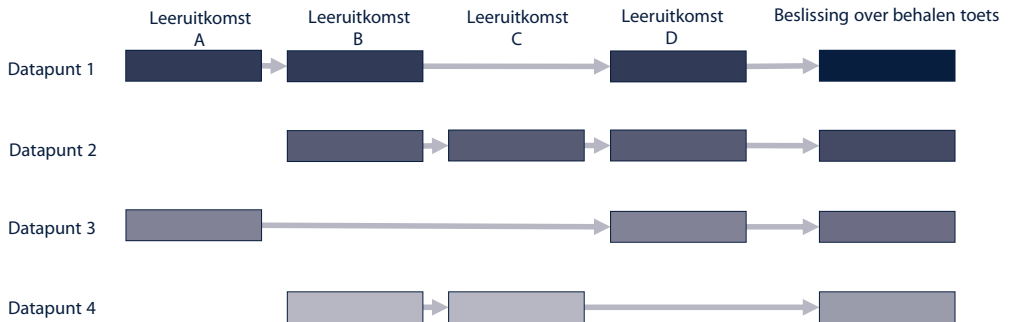
Bij programmatisch toetsen wordt het toetsprogramma van een opleiding integraal en doelbewust uitgedacht, met de beoogde leeropbrengsten als uitgangspunt (Van der Vleuten, Schut, & Heeneman, 2018). Een raamwerk van competenties, vaardigheden, beroepstaken of leeruitkomsten vormt de ruggengraat van zowel het onderwijsprogramma als het toetsprogramma. Dat raamwerk omvat de leerlijnen binnen het gehele programma. Alle datapunten in het toetsprogramma worden gekoppeld aan dit raamwerk. Dat maakt voor student en docent inzichtelijk dat alle opdrachten, feedback, oefeningen et cetera, een bijdrage leveren aan het langlopende leerproces van de student, en de beoogde leerresultaten.

Dit raamwerk of deze ruggengraat biedt voor begeleiders tevens de mogelijkheid om de ontwikkeling van studenten te monitoren. Een (elektronisch) portfolio is daarbij behulpzaam. Het raamwerk structureert ook dit portfolio: vanuit de competenties, vaardigheden, beroepstaken of leeruitkomsten wordt bepaald welke informatie in het portfolio moet worden opgenomen. Het monitoren en

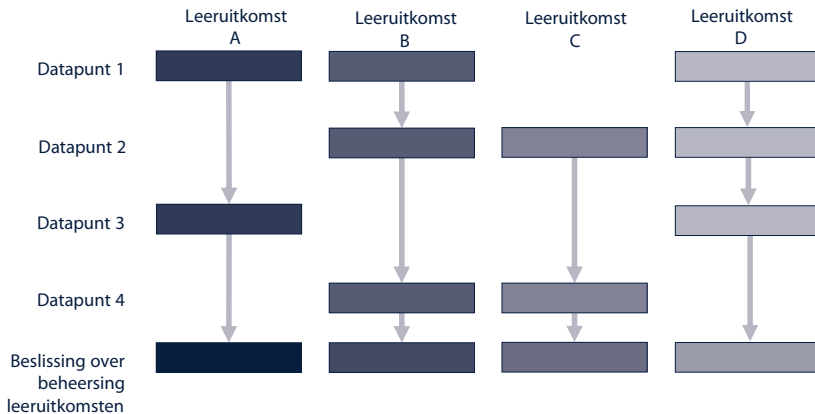
begeleiden van studentontwikkeling gebeurt door koppeling van de datapunten aan de ruggengraat. Bijvoorbeeld: een student maakt een opdracht en krijgt feedback op de eigen prestaties aan de hand van een beoordelings-/feedback-formulier waarop de competenties zijn benoemd. De student bespreekt met zijn begeleider dus niet alleen hoe hij deze specifieke opdracht heeft uitgevoerd, maar ook hoe hij daarbij zijn competenties heeft ingezet. De feedback die de student ontvangt, gaat dus tevens over zijn competentieontwikkeling. En dat is feedback die de student bij de volgende opdracht opnieuw kan benutten, omdat ook deze opdracht is gekoppeld aan de ruggengraat, en de competenties er opnieuw in terugkomen en zichtbaar zijn. In het huidige onderwijsconcept is dit vaak niet het geval, omdat opdrachten inhoudelijk niet aan elkaar zijn gekoppeld door zo'n raamwerk. De student krijgt dan wel feedback, maar kan die vaak niet benutten bij een volgende opdracht.

Ook bij het nemen van een (summatieve) beslissing over studenten is bij programmatisch toetsen de ruggengraat leidend. De beslissing wordt altijd genomen op basis van die ruggengraat, het gaat dus om een beslissing over bijvoorbeeld competenties of leeruitkomsten. Alle verzamelde datapunten – de informatiebrokjes – vertellen het verhaal van de student over die leeruitkomsten. Stel dat een opleiding vier leeruitkomsten definieert, dan wordt de informatie uit de verschillende datapunten gebruikt om een beslissing te nemen over de vraag of de student die vier leeruitkomsten beheerst, over alle datapunten heen. Dit is een andere manier van informatie samenvoegen dan vaak in het hoger onderwijs gebeurt. Bij de conventionele manier van toetsen wordt een beslissing genomen over het wel/niet behalen van een toets, module of vak, terwijl mogelijk hierbinnen verschillende leeruitkomsten aan bod komen. Opleidingen kiezen ervoor om binnen de conventionele manier van toetsen op niveau van een toets, module of vak een beslissing te nemen doordat hieraan studiepunten (ECTS) zijn gekoppeld. Bij programmatisch toetsen wordt de informatie uit datapunten niet bekeken op het niveau van een afzonderlijke toets, maar op het niveau van de leeruitkomst waarover een beslissing genomen moet worden. In Figuur 1.1 wordt dit inzichtelijk gemaakt: bij de conventionele manier van toetsen wordt, ondanks dat in de toets meerdere leeruitkomsten aan de orde komen, een beslissing genomen op het niveau van de toets. Bij programmatisch toetsen wordt informatie op het niveau van de leeruitkomsten gedestilleerd uit alle beschikbare datapunten en wordt over deze datapunten heen een beslissing genomen over de beheersing van de leeruitkomst.

Beslissing binnen conventioneel toetsconcept: toets (datapunt) als uitgangspunt



Beslissing binnen programmatisch toetsen: leeruitkomsten als uitgangspunt



Figuur 1.1 Aggregatie van informatie

1.4 Er is een constante dialoog over het gebruik van feedback voor zelfsturing

Een voorwaarde om studenten echt te laten leren van feedback uit de datapunten, is een continue dialoog met de student. Daarin wordt de student gestimuleerd om te reflecteren op de gegeven feedback, en om feedback van anderen (uit de datapunten) te vergelijken met zijn eigen beeld van zijn ontwikkeling. De student kan zich dan bijvoorbeeld realiseren dat hij bepaalde feedback verschillende keren heeft gekregen, en dat er op dit punt dus werk aan de winkel is voor een volgende opdracht.

Het is bekend dat feedback krijgen geen garantie is voor het gebruiken ervan. Ook toont onderzoek aan dat studenten begeleid moeten worden bij het nemen van zelfregie: de ontwikkeling van zelfsturing vraagt om sturing. Bij pro-

grammatisch toetsen biedt een coach of mentor die sturing. De coach volgt de ontwikkeling van de student en voert regelmatig evaluatiegesprekken met de student. Meestal ligt de coachrol bij een docent uit het onderwijsprogramma. Deze coach begeleidt een (beperkte) groep studenten. Om dit goed te kunnen doen heeft de coach toegang tot het portfolio, en spreekt hij zijn studenten regelmatig. In die gesprekken komt de ontwikkeling van de student aan bod, docent en student bespreken die aan de hand van de datapunten die tot dan toe zijn verzameld. Om te stimuleren dat er een constructieve dialoog plaatsvindt tussen de student en de coach, is het de bedoeling dat de student deze gesprekken voorbereidt. Ook de student gebruikt daarvoor de informatie uit de datapunten.

Het geven en ontvangen van feedback is niet voor iedereen eenvoudig, maar het is een vaardigheid die geleerd kan worden. Dat vergt echter zowel bij studenten als bij docenten kennis, tijd en aandacht. Vaak bestaat feedback uit een dialoog tussen de feedbackontvanger en de feedbackgevende. Bij voorkeur bestaat zo'n dialoog uit het stellen en beantwoorden van (open) vragen. Een voorbeeld van vragen die een docent kan stellen die zojuist een student heeft geobserveerd met een cliënt:

Waar dacht je aan in het begin van het gesprek?

Had het ook iets anders kunnen zijn?

Was het gesprek dan anders gelopen?

Zag je de reactie van de cliënt toen je zei ?

Hoe zou je dat een volgende keer anders kunnen doen?

Wil je dit gesprek even kort samenvatten en uploaden naar het portfolio?

Aan deze vragen zie je dat het “beoordelen” op de achtergrond staat. Het gaat vooral om het sturen van het leren. Zelf ontdekken en reflectie zijn belangrijke elementen van zo'n dialoog, meer nog dan informatieoverdracht. Relaties in zo'n dialoog zijn belangrijk. Het creëren van vertrouwde en veilige leersituaties is daarom van het grootste belang. Als dit goed gebeurt ontstaat een feedback-cultuur, waarin het geven en ontvangen van feedback als normaal wordt gezien.

1.5 Het aantal datapunten en de zwaarte van de beslissing zijn proportioneel aan elkaar gerelateerd

Bij programmatisch toetsen spreken we niet meer over formatieve of summatieve toetsen. We spreken over een continuüm van wat er op het spel staat: heeft de beslissing voor de student kleine of grote consequenties?

In Figuur 1.2 wordt dit inzichtelijk gemaakt. Feedback op één enkel datapunt leidt niet tot een beslissing en voor de student staat er dus weinig op het spel. De student mag fouten maken. Het belangrijkste bij een datapunt is informa-

tierijke feedback voor de student. In het midden van het continuüm bevindt zich de tussentijdse beslissing. Deze beslissing vindt bijvoorbeeld halverwege de onderwijsperiode plaats, er staat dan iets meer op het spel. Op dat moment worden het gehele portfolio en de tot dan toe verzamelde datapunten beoordeeld. Zo'n tussentijds oordeel is diagnostisch van aard ('Wat gaat goed, wat kan beter?'), therapeutisch ('Waar moet aan gewerkt worden?') en voorspellend ('Waar gaat dit naartoe?'). Belangrijk is dat een tussentijdse beslissing op tijd wordt ingepland, zodat de student nog tijd heeft om te werken aan de geconstateerde leervragen, en datapunten kan verzamelen in het portfolio om groei te laten zien.



Figuur 1.2 Continuüm van zwaarte van beslissingen

Aan het andere uiterste van het continuüm staat de eindbeslissing. Bij een eindbeslissing wordt een beslissing genomen over het afsluiten van een onderwijsperiode, en de toegang tot de volgende onderwijsperiode. Er wordt, met andere woorden, een zak-/slaagbeslissing genomen. Wanneer de student slaagt, worden alle studiepunten die behoren bij de onderwijsperiode ineens uitgekeerd (bijvoorbeeld 30 ECTS-punten als de onderwijsperiode een half studiejaar omvat of 60 ECTS-punten voor een heel jaar). Soms worden aan de zak-/slaagbeslissingen nog andere kwalificaties toegevoegd, bijvoorbeeld om te kunnen bepalen of een student cum laude slaagt. Wanneer de student zakt, volgt meestal een advies voor de verdere voortgang. Bij een eindbeslissing staat veel op het spel en er zijn dan ook veel datapunten nodig om een robuuste beslissing te kunnen nemen. De pixels van de afzonderlijke datapunten verenigen zich in een totaalbeeld over de student. Als het beeld duidelijk is, kan een legitiem besluit plaatsvinden.

1.6 De zwaarte van een beslissing is leidend voor de hoeveelheid benodigde beoordelaarsexpertise

Gezien de heterogeniteit van de datapunten is het onmogelijk om een rekenkundig oordeel te vellen over de totaliteit ervan. Immers, de informatie in het portfolio kan kwantitatief of kwalitatief van aard zijn en er kan dus geen reken-

kundig gemiddelde worden berekend. Bovendien kan het portfolio allerlei soorten datapunten bevatten, bijvoorbeeld feedback van experts en peers, een lijst van verrichtingen, zelfreflecties of andere informatie waaruit de expertise van de student blijkt. De eindbeoordeling vereist dus een holistisch professioneel oordeel.

Om tot een zo grondig en objectief mogelijke beoordeling van de student te komen, wordt gezorgd voor onafhankelijkheid van beoordelaars en wordt de hoeveelheid beoordelaarsexpertise opgevoerd indien nodig. Belangrijk is dat deze beoordelaars niet bij het leerproces van de student betrokken zijn geweest. Dit creëert een interessant spanningsveld, dat ook wel bekend staat als het *fire-wall dilemma*. Immers, de persoon die het best de student zou kunnen beoordelen is de coach, want de coach kent de student het best en heeft dus veel informatie tot zijn beschikking. Maar als de coach betrokken is bij een zwaarwegende beslissing (de beoordeling van de student), is de relatie met de student anders: het is dan moeilijker deze veilig te laten zijn en feedback te geven. Als de veiligheid op het spel staat, kunnen er “spelletjes” worden gespeeld tussen student en coach. Een coach kan zich bijvoorbeeld gedwongen voelen om een bepaald oordeel te geven om de goede relatie te bewaren, of de student kan zich belemmerd voelen in het benoemen van ontwikkelingspunten omdat dit een lagere beoordeling tot gevolg zou kunnen hebben. Er zijn verschillende oplossingen voor dit dilemma. Zo kan een andere coach tussentijds beoordelen. Of een andere coach kan advies geven aan de onafhankelijke eindbeoordelaars – al dan niet aangevuld met de visie van de student. Dat advies weegt dan mee, maar is niet doorslaggevend voor de eindbeoordeling.

Afhankelijk van de beoordeling die plaats moet vinden, wordt gekozen voor het inzetten van meer of minder beoordelaars, en daarmee meer of minder beoordelaarsexpertise. In het meest ideale geval neemt een groep of een commissie van onafhankelijke beoordelaars de zwaarwegende eindbeslissingen. Bij programmatisch toetsen maakt men meestal gebruik van holistische beoordelingscriteria en beschrijvingen van niveaus om het oordelen voor beoordelaars te vergemakkelijken. Deze uitwerkingen, vaak opgenomen in rubrics, kunnen afgeleid zijn van het overkoepelende raamwerk van de opleiding. Zo kan er bijvoorbeeld een korte beschrijving zijn van wat in welke fase van de opleiding verwacht mag worden van de competentie communicatie. Die omschrijvingen zijn dan mijlpalen of uitwerkingen van niveaus, die richting geven aan de beslissing van de beoordelaars. Het is ook mogelijk om de beoordelaarsexpertise alleen te vergroten bij studenten voor wie dit nodig is. Voor (verreweg) het grootste deel van de studenten zijn de te nemen beslissingen eenvoudig en heel duidelijk: bij deze studenten wijzen de datapunten allemaal in dezelfde goede ontwikkelingsrichting. Het is dan onnodig om veel beoordelaarsexpertise in te zetten. Bij sommige studenten zal dat wel nodig zijn en moet een student worden besproken in een voltallige commissie. De beoordelaarsexpertise verzekert dat een beslissing ook uitgelegd of gerechtvaardigd kan worden. Dit

draagt bij aan de legitimiteit van de beslissing. Als dit proces van holistisch beoordelen zorgvuldig wordt omkleed met onafhankelijkheid en voldoende beoordelaarsexpertise, doorstaat een eindbeslissing zelfs een juridische toets. De tussentijdse beslissingen en het proces van coaching dragen hieraan bij. Belangrijk bij programmatisch toetsen is dan ook dat de eindbeslissing geen verrassing mag zijn voor de student.

Samenvattend

Programmatisch toetsen is een generiek toetsconcept: het is geschikt voor ieder type onderwijs, tenminste, wanneer de visie op leren van de opleiding aansluit op een constructivistisch onderwijsconcept. Als zo'n onderwijsconcept niet wordt onderschreven, heeft programmatisch toetsen geen enkele zin. Dat is bijvoorbeeld het geval als het programma bestaat uit afzonderlijke vakken, die allemaal behaald moeten worden. Er is dan weinig ruimte voor programmatisch toetsen. Is er sprake van lange leerlijnen in een onderwijsprogramma, zijn generieke vaardigheden belangrijk en wordt zelfsturing van de student verwacht? Dan kan programmatisch toetsen heel geschikt zijn. Bij programmatisch toetsen staat het leerproces van de student centraal. Toetsen, of eigenlijk de datapunten – die veel breder worden opgevat dan alleen toetsen – hebben geen gevolgen in termen van 'halen' of 'niet halen'. Daardoor staan deze datapunten veel meer in het teken van leren. Door het werken met een ruggegraat van datapunten kunnen studenten de feedback die ze krijgen, benutten in vervolgopdrachten of activiteiten, die ze soms ook zelf kunnen kiezen. Bij programmatisch toetsen kijkt men ook anders naar betrouwbaarheid en validiteit. Binnen de conventionele manier van toetsen worden betrouwbaarheid en validiteit veelal beschouwd vanuit de losse toetsen. Bij programmatisch toetsen worden verschillende soorten datapunten en feedback van verschillende beoordelaars samengenomen, en worden beslissingen genomen op basis van dit totaalplaatje van de student. Dit zorgt voor een robuuste, betrouwbare beslissing over de student en het vervolgtraject in zijn opleiding.

Programmatisch toetsen biedt een geheel andere kijk op het curriculum en de toetsing dan de traditionele summatieve zienswijze van toetsing. Het is voor veel docenten niet gemakkelijk om te denken buiten het conventionele (summatieve) paradigma. Zoals voor elke onderwijsinnovatie moet er draagvlak zijn onder de stakeholders van het onderwijsprogramma. Zonder dat draagvlak zal programmatisch toetsen niet succesvol kunnen worden geïmplementeerd. Programmatisch toetsen is een toetsconcept. Zoals met elk concept, kan dat op verschillende manieren worden vertaald naar een concrete implementatie in het onderwijs. In dit boek komen verschillende implementaties aan bod, bij uiteenlopende opleidingen. Het laatste hoofdstuk geeft weer welke lessen we daarvan kunnen leren.

Binnen het beroepsonderwijs is steeds meer belangstelling voor programmatisch toetsen. Dat is een toetsconcept waarin toetsen en leren geïntegreerd zijn en de competentieontwikkeling van de student centraal staat. Gedurende de opleiding wordt continu informatie over het leerproces van de student verzameld, de zogeheten datapunten. Hiermee kunnen opleiders het leerproces monitoren en bijsturen. Feedback en de ontwikkeling van zelfregulatie staan centraal. Zo krijgen studenten de benodigde tijd en ruimte om zich te ontwikkelen en écht te leren van toetsen.

In dit boek komen opleiders aan het woord, die hun onderwijs- en toetsprogramma hebben ingericht volgens de uitgangspunten van programmatisch toetsen. Zij vertellen hun eigen verhaal: Waarom hebben zij gekozen voor programmatisch toetsen? Hoe is het toetsprogramma ingericht? Wat zijn ervaringen van studenten en docenten? En hoe is het implementatieproces verlopen?

De voorbeelden laten zien dat programmatisch toetsen geen recept is. Het moet aansluiten bij de onderwijsvisie van een opleiding, bij het beroep waartoe je opleidt en bij het type studenten. Elke opleiding maakt daarom eigen keuzes bij de uitwerking van programmatisch toetsen.

Dit boek is bedoeld voor onderwijsprofessionals in het middelbaar en hoger beroepsonderwijs en het wetenschappelijk onderwijs die overwegen om de uitgangspunten van programmatisch toetsen toe te passen binnen het onderwijs en op zoek zijn naar theorie, voorbeelden en inspiratie.

