

# PROGRAMMATISCH TOETSEN, IETS VOOR MIJN OPLEIDING? \_

whitepaper

Marion van de Wijdeven

Eline den Tuinder

Kyle van den Langenberg

Tamara van Schilt-Mol

*Academie Educatie, Onderzoeksteam Kwaliteiten van Leraren*

1 juli 2024

## INHOUDSOPGAVE

<b>INLEIDING .....</b>	<b>3</b>
<b>1 PROGRAMMATISCH TOETSEN, WAT HOUDT DAT IN? .....</b>	<b>4</b>
Ontstaan van programmatisch toetsen.....	6
Ontwerpprincipes .....	8
Ontwerpkeuzes .....	10
<b>2 PROGRAMMATISCH TOETSEN, IETS VOOR MIJN OPLEIDING?.....</b>	<b>12</b>
Visie op onderwijs is leidend .....	12
Vakoverstijgend denken (ook wel heldere ruggengraat) .....	13
Lerende cultuur .....	14
De rol van feedback en feedbackgeletterdheid van alle actoren .....	15
Moed .....	15
Samenvatting van de belangrijkste aandachtspunten .....	15
<b>3 BENIEUWD NAAR PRAKTIJKERVARINGEN? .....</b>	<b>16</b>
<b>LITERATUUR .....</b>	<b>17</b>

## **INLEIDING**

Programmatisch toetsen is een concept dat momenteel vaak ter sprake komt bij opleidingen in het beroepsonderwijs die overwegen om de toetsing anders in te richten.

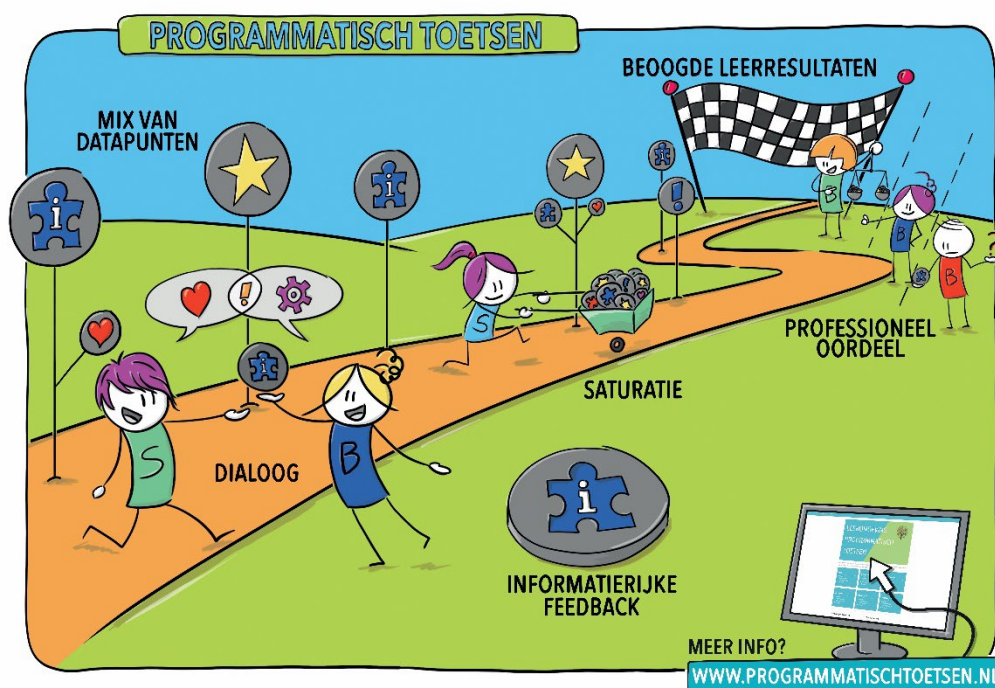
Vanuit het lectoraat Eigentijds Beoordelen en Beslissen merken we dat er bij HAN-opleidingen, die overwegen om te gaan werken volgens de principes van programmatisch toetsen, behoefte is aan informatie over en handvatten voor invoering van programmatisch toetsen. Met dit whitepaper willen we HAN-opleidingen ondersteunen meer wegwijs te worden in het concept programmatisch toetsen en het ontstaan daarvan. Vervolgens reiken we enkele aspecten aan aan de hand waarvan je kunt bepalen of programmatisch toetsen mogelijk iets is voor jouw opleiding. Of misschien juist wel helemaal niet.

## 1 PROGRAMMATISCH TOETSEN, WAT HOUDT DAT IN?

Van Schilt-Mol en Baartman (2023) beschrijven dat programmatisch toetsen inhoudt dat studenten tijdens een (langere) leerperiode informatie verzamelen over hun ontwikkeling, door middel van datapunten. Een datapunt is iets dat een student doet of maakt, bijvoorbeeld een essay, handeling in de praktijk, tentamen, et cetera. Studenten ontvangen feedback op de datapunten, bijvoorbeeld van een docent, medestudent of iemand uit de praktijk. Deze feedback is bedoeld om het leren te sturen; informatierijke feedback over waar de student staat ten opzichte van het doel dat centraal staat en welke stappen de student nog kan zetten om dit doel te behalen. Met andere woorden: het gaat om het geven van een (waarde)oordeel, zonder dat er bijvoorbeeld studiepunten aan gekoppeld zijn. Wanneer er voldoende en diverse informatie is verzameld op datapunten, kan er pas een betrouwbare uitspraak worden gedaan over het wel of niet behalen van de leeruitkomsten. Er wordt bij programmatisch toetsen dus gekeken naar de gehele ontwikkeling van studenten op het beoogde eindniveau, in plaats van afvinken van onderdelen in individuele toetsen (Baartman et al., 2020). Deze holistische manier van beoordelen sluit aan bij een constructivistische benadering van onderwijs (Baartman et al., 2020); leren wordt gezien als een actief en sociaal proces, waarbij de verantwoordelijkheid van het leerproces bij de student ligt (Last & Jongen, 2021). In onderstaande afbeelding (Figuur 1) is het proces van programmatisch toetsen visueel weergegeven in de vorm van een praatplaat:

**Figuur 1**

*Praatplaat programmatisch toetsen*



*Opmerking. Deze praatplaat is in opdracht van Tamara van Schilt-Mol en Liesbeth Baartman ontwikkeld door Studio Duikboot.*

Programmatisch toetsen verschilt met de ‘klassieke modulaire aanpak’ van toetsing, zoals Van der Vleuten et al. (2018) dit noemen. In deze aanpak is het gebruikelijk dat elk onderdeel wordt afgesloten met een toets; deze manier van toetsen behoort tot een model van leren dat gericht is op deelbeheersing van de leerinhoud. Echter, deze manier van toetsing past vaak minder goed bij de opleidingen die zich focussen op (complexe) vaardigheden en competenties waar de arbeidsmarkt behoefte aan heeft. Bij holistisch beoordelen gaat het veel meer om het kijken naar complexe vaardigheden die tot uiting komen in het gedrag van de student. Van een student wordt in eigentijdse opleidingen zelfsturing verwacht, zeker met het oog op het idee van een leven lang ontwikkelen dat veel van deze opleidingen hun studenten mee willen geven. In de ‘klassieke aanpak’ van toetsing valt er echter weinig te sturen, aangezien het ritme van de student wordt bepaald door vooraf vastgelegde toetsen. Kortom: in de ‘klassieke aanpak’ wordt het leren vooral gestuurd door (summatieve) toetsen, waardoor het leren niet goed aansluit bij, zoals Van der Vleuten et al., (2018) het noemen, het moderne opleiden (Van der Vleuten et al., 2018). In onderstaande tabel staan een aantal belangrijke verschillen tussen de klassieke (modulaire) en de moderne aanpak van toetsing, ook wel hele taak-benadering genoemd (gebaseerd op Van der Vleuten et al., 2018).

**Tabel 1**

*Vergelijking klassieke modulaire aanpak en hele taakbenadering*

<b>Klassieke modulaire aanpak / losse vakken</b> <b>Beheersingsleren/"test-theorie"</b>	<b>Hele taakbenadering</b> <b>Constructivisme/"holistisch beoordelen"</b>
(elk onderdeel van) een onderwijsprogramma wordt afgesloten met een toets en/of beroepsproduct	geheel van (informatie uit) toetsen geeft inzicht in de beheersing van het geheel van leerdoelen van de opleiding
minimumgrens behaald = voldoende/geslaagd minimumgrens niet behaald = herkansing	informatie uit toetsen kan zonder negatieve gevolgen worden gebruikt om het leerproces vorm te geven (fouten maken mag en is veilig)
toetsen zijn gericht op leerdoelen/inhouden van het thema, iedere toets leidt tot een ‘zwaarwegend besluit’	toetsen zijn gerelateerd aan de beoogde leerresultaten van de opleiding, er is samenhang binnen en tussen vakken
leren is veelal gericht op halen van de toets (performance goal)	toetsen maken wordt gezien als onderdeel van ontwikkelingsproces (mastery goal)

*Opmerking.* Gebaseerd op Van der Vleuten, C., Schut, S., & Heeneman, S. (2018). Programmatisch toetsen als motor voor professioneel leren in het hoger onderwijs. In: Sluijsmans, D. en M. Segers (red.), *Toetsrevolutie. Naar een feedbackcultuur in het hoger onderwijs*. Phronese.

## Ontstaan van programmatisch toetsen

Het idee voor programmatisch toetsen is afkomstig uit het medisch onderwijs. Om complexe competenties te kunnen beoordelen, is het van belang om informatie uit verschillende bronnen (instrumenten, beoordelaars en contexten) te halen om zo recht te doen aan validiteit en betrouwbaarheid (Van der Vleuten & Schuwirth, 2005). Om hieraan tegemoet te komen schreven Van der Vleuten en Schuwirth al in 2005 een pleidooi voor het benaderen van toetsing als systeem, waarbij werd gedacht vanuit toetsprogramma's, in plaats van vanuit 'losse toetsen' ofwel *individual assessment methods* (Van der Vleuten & Schuwirth, 2005). Aanvullend hierop beschreven Van der Vleuten et al. (2012) een model voor programmatisch toetsen, bedoeld om handvatten te bieden bij het vormgeven van een toetsprogramma in het onderwijs. Het model is gebaseerd op een aantal principes (gebaseerd op empirisch onderzoek, voor verdere toelichting zie Van der Vleuten, et al. (2010)), bijvoorbeeld dat toetsing het leren stimuleert en oordelen van experts (bijvoorbeeld docent, mentor, medestudent, et cetera) noodzakelijk zijn. Het doel van het model is om zowel de leerfunctie als de beslisfunctie van toetsing te optimaliseren, en de kwaliteit van het onderwijs te evalueren (Van der Vleuten et al., 2012).

### *Leerfunctie*

Bij programmatisch toetsen gaat het om de "*longitudinale ontwikkeling van de lerende door middel van leertaken, passende feedback en (ondersteunde) zelfsturing*" (Van der Vleuten et al., 2012, p. 210). Deze ontwikkeling wordt gevolgd door middel van datapunten, waarbij elk datapunt volledig in het teken staat van het leerproces van de student (Baartman et al., 2020). Het leren wordt geoptimaliseerd omdat er op basis van 'losse' datapunten geen zwaarwegende beslissingen worden genomen (Heeneman et al., 2021). Er is een continuüm van beslissingen die genomen kunnen worden op basis van datapunten. Al deze beslissingen zijn van belang. De beslissingen verschillen echter in wat er op het spel staat voor de student, wat de 'stake' is. Studenten ontvangen feedback op hun datapunten waarmee het leren (verder) kan worden bevorderd (Baartman et al., 2020). Voor meer informatie over de leerfunctie van programmatisch toetsen, zie <https://www.programmatischtoetsen.nl/4-leerfunctie.html>.

### *Beslisfunctie*

Aan het eind van een onderwijsperiode (bijvoorbeeld een onderwijseenheid, semester, et cetera) wordt er een zak-/slaagbeslissing genomen. Om een robuuste en betrouwbare beslissing te nemen, is het van belang dat er feedback op verschillende soorten datapunten en van verschillende beoordelaars wordt samengenomen, zodat de beslissing op basis van dit totaalbeeld van de student gemaakt kan worden (Baartman et al., 2020) (zie ter illustratie figuur 2). Deze zwaarwegende beslissing zou geen verrassing moeten zijn voor zowel de student, als de beoordelaars (Van der Vleuten et al., 2012). Deze beslissing wordt veelal genomen door een besliscommissie. Voor meer informatie over de beslisfunctie van programmatisch toetsen, zie <https://www.programmatischtoetsen.nl/5-beslisfunctie.html>.

**Figuur 2**

*Het verzamelen van meerdere 'pixels' (datapunten) geeft een duidelijker beeld van de student*



*Opmerking.* Deze afbeeldingen zijn in opdracht van Tamara van Schilt-Mol en Liesbeth Baartman ontwikkeld door Studio Duikboot.

### *Continuüm van 'stakes'*

Er wordt bij programmatisch toetsen, met andere woorden, dus een continuüm van stakes gehanteerd. De dichotomie tussen de formatieve en summatieve functies van toetsen die bij meer traditionele toetssystemen wordt gehanteerd, wordt bij programmatisch toetsen vervangen door dit continuüm van stakes, waarbij low, medium en high stakes beslissingen onderscheiden worden (Agricola et al., 2024; Baartman et al., 2020; Van der Vleuten et al., 2012).

De 'low stakes-beslissingen' zijn gebaseerd op de informatie uit een enkel datapunt. Deze informatie wordt benut om het leren van de student te sturen en stimuleren. Deze beslissing wordt 'low stakes' genoemd omdat er voor de student weinig op het spel staat: de feedback op één enkel datapunt leidt niet tot een zak/slaagbeslissing, de student mag fouten maken. Het is geen 'no stakes-beslissing' omdat er niet niets op het spel staat: de informatie wordt aan het eind van een leerperiode namelijk wel gebruikt bij de interpretatie van het geheel van alle informatie uit meerdere datapunten. Tijdens 'medium stakes-beslissingen' – tussentijdse beslissingen – wordt vastgesteld hoe de student vordert en wat nog kan worden verbeterd. Het doel van de beslissingen is diagnostisch ('wat gaat goed, wat kan beter?'), therapeutisch ('waar moet aan gewerkt worden?') en voorspellend ('waar gaat dit naar toe?'). Deze beslissingen worden genomen op basis van informatie uit meerdere datapunten. 'High stakes-beslissingen' – bijvoorbeeld het toekennen van studiepunten - worden genomen op basis van een veelheid aan informatie uit datapunten.

## Ontwerpprincipes

In de literatuur zijn verschillende publicaties te vinden met ontwerpprincipes voor programmatisch toetsen, die in de basis duidelijk overeenkomen met elkaar (zie bijvoorbeeld Baartman et al., 2020; Heeneman et al., 2021; Van der Vleuten et al., 2018; Van der Vleuten et al., 2012). Van Schilt-Mol et al. (nog te verschijnen) hebben drie recente uitwerkingen van principes van programmatisch toetsen naast elkaar gelegd en weergegeven in een tabel:

1. **Rijen:** Baartman et al. (2020). Zes principes, gebaseerd op onder andere Van der Vleuten et al., 2012; De Jong et al., 2019; Jamieson et al., 2017; Schut et al., 2020.
2. **Cellen:** Heeneman et al. (2021). Twaalf principes, gebaseerd op Van der Vleuten et al. (2010; 2012; 2017; 2019); Van der Vleuten & Schuwirth, (2005) en een consensusbijeenkomst.
3. **Kolommen:** Torre et al. (2021). Driedeling op basis van de twaalf principes van Heeneman et al. (2021).



**Tabel 2**

*Vergelijking van de principes van programmatisch toetsen.*

	<u>Continue zinvolle feedback om de dialoog te bevorderen</u>	<u>Mix van toetsvormen binnen een continuüm van 'stakes'</u>	<u>Het opzetten van rechtvaardige en geloofwaardige besluitvormingsprocessen</u>
<b>Beoogde leerresultaten vormen de ruggengraat van het programma en vormen de basis voor datapunten en aggregatie</b>		<i>De keuze voor datapunten is gebaseerd op de passendheid bij de beoogde leerresultaten.</i>	<i>Informatie uit datapunten wordt geaggregeerd naar de leerresultaten waarover een beslissing wordt genomen</i>
<b>Een mix van verschillende datapunten wordt gebruikt om inzicht te krijgen in de ontwikkeling van studenten.</b>	<i>Alle toetsvormen zijn slechts datapunten.</i>	<i>Er is een mix aan verschillende soorten datapunten.</i>	
<b>Alle datapunten zijn gericht op het optimaliseren van feedback en hebben geen betrekking op zak / slaag-beslissingen.</b>	<i>Elk datapunt wordt geoptimaliseerd voor leren door betekenisvolle feedback te verstrekken.</i>	<i>Zak/slaagbeslissingen worden niet genomen op basis van een enkel datapunt.</i>	
<b>Een continue dialoog over datapunten, feedback en zelfregulering</b>	<ul style="list-style-type: none"> <li>• <i>Tussentijdse medium stakes beslissingen hebben als doel voortgang te bespreken en vervolgstappen in leren te bepalen.</i></li> <li>• <i>Studenten hebben regelmatig gesprekken met docenten waarin verzamelde datapunten worden besproken en geanalyseerd.</i></li> <li>• <i>Het eigenaarschap van studenten wordt bevorderd doordat studenten eigen keuzes maken voor vervolgstappen in leren, gebaseerd op datapunten.</i></li> </ul>		
<b>Er is een evenredige verhouding tussen het aantal datapunten en de zwaarte van de beslissing.</b>		<i>Een low-high stakes continuüm vervangt het onderscheid tussen formatief en summatief.</i>	<i>Hoe zwaarder de beslissing, hoe meer datapunten nodig zijn als basis voor die beslissing.</i>
<b>Hoe hoger de inzet van de beslissing, hoe meer expertise van de beoordelaar nodig is.</b>			<i>High stakes beslissingen zijn holistisch van aard en worden genomen op een robuuste wijze.</i>

*Opmerking.* Overgenomen uit *Programmatisch toetsen: Een verdieping vanuit theorie en praktijk* (p. 8-9) door T. Van Schilt-Mol, L. Baartman en C. Van der Vleuten, nog te verschijnen. Boom Hoger onderwijs.

## Ontwerpkeuzes

Heeneman et al. (2021) geven aan dat de principes belangrijk zijn, maar dat ze op veel verschillende manieren vormgegeven kunnen worden. Zoals ook Baartman et al. (2020) aangeven: programmatisch toetsen is een toetsconcept waarbinnen je keuzes kan maken, geen toetsrecept dat je moet volgen. Het is van belang dat opleidingen hun onderwijs inrichten zodat deze passend is voor de context van de opleiding, waarbij onderwijs en toetsing op elkaar zijn afgestemd (Heeneman et al, 2021). Hierbij maken opleidingen, op basis van de principes, verschillende ontwerpkeuzes die passend zijn bij hun eigen context. Binnen het NRO-project 'Naar een nieuwe balans in toetsfuncties: van formatief en summatief naar een continuüm van beslissingen' is onderzoek gedaan naar de ontwerpkeuzes die opleidingen maken bij het implementeren van programmatisch toetsen. Op basis hiervan zijn 11 ontwerpkeuzes geïdentificeerd (Van Schilt-Mol et al., nog te verschijnen):

1. wijze waarop kennis al dan niet als datapunt wordt geprogrammeerd in het curriculum;
2. wijze waarop een prestatie zoals een criterium gericht interview (CGI) wel of niet ingezet wordt tijdens de high stakes beslissing;
3. wijze waarop beoogde leerresultaten separaat of holistisch worden beoordeeld tijdens high stakes beslissing;
4. wijze waarop vaste en/of vrije datapunten worden geprogrammeerd in het curriculum;
5. wijze waarop het feedbackformulier voor de afzonderlijke datapunten is gebaseerd op de beoogde leerresultaten;
6. wijze waarop medium stakes beslissingen worden ingezet;
7. wijze waarop studenten vrijheid hebben bij het kiezen van feedbackperspectieven op de datapunten;
8. wijze waarop selfassessment in het toetsprogramma is opgenomen als datapunt;
9. wijze waarop de leerprocesbegeleider (mentor/coach/...) een rol heeft in het beslisproces bij de high stakes beslissing;
10. wijze waarop de momenten voor het geven/krijgen van feedback op de datapunten zijn vastgelegd;
11. de omvang van onderwijseenheden waarover de high stakes beslissingen worden genomen.

Naast deze specifieke ontwerpkeuzes zijn uit het onderzoek (Agricola et al., 2024) ook algemene ontwerpkeuzes naar voren gekomen, die elke opleiding (met verschil in exacte invulling en uitvoering) maakt. Voorbeelden hiervan zijn dat opleidingen bij het ontwerpen van het onderwijs en de toetsing een set van leeruitkomsten hanteren die centraal staan (ruggengraat van beoogde resultaten) en dat er een mix van verschillende datapunten wordt gehanteerd. Hiernaast worden er ook altijd leerprocesbegeleiders ingezet en wordt er gebruik gemaakt van een e-portfolio-systematiek. Verdere toelichting op deze algemene ontwerpkeuzes is te lezen in Agricola, et al. (2024).

De ruimte in het maken van de ontwerpkeuzes kan prettig zijn, maar dit kan er ook toe leiden dat er geen sprake (meer) is van programmatisch toetsen. Bijvoorbeeld wanneer ...

- ... de datapunten als voldoende beoordeeld moeten zijn voordat de student 'op mag' voor een high stake beslissing (wél programmatisch toetsen wanneer er verplichte toetsen buiten het programmatisch toetsen curriculum geplaatst worden).
- ... er een afsluitend criterium gericht interview (CGI) is waarin het portfolio wordt beoordeeld en waaraan de beslissing is gekoppeld (wél programmatisch toetsen wanneer het CGI wordt gebruikt als extra datapunt om aanvullende informatie te verzamelen).
- ... de high stake beslissing wordt genomen op basis van een gemiddelde beoordeling van de datapunten (wél programmatisch toetsen wanneer er door de besliscommissie wordt gekeken naar het geheel van alle beoordelingen (holistisch beoordelen), waarbij de narratieve feedback leidend is voor de beslissing).

## 2 PROGRAMMATISCH TOETSEN, IETS VOOR MIJN OPLEIDING?

Het concept programmatisch toetsen past niet bij iedere opleiding. Wat zijn overwegingen om wel of juist niet voor dit concept te kiezen? Onderzoek (Baartman et al., 2020) laat zien dat het essentieel is dat opleidingen weloverwogen kiezen voor programmatisch toetsen, zodat programmatisch toetsen doelgericht en bewust kan worden ingezet als middel om het gewenste doel van de opleiding te realiseren. Programmatisch toetsen moet dus een middel zijn, en geen doel op zich.

Wanneer je programmatisch toetsen niet weloverwogen invoert, en wanneer je het doel ervan nog niet helder voor ogen hebt, kan dit consequenties hebben voor de kwaliteit van het onderwijs en de toetsing binnen je opleiding. Er ontstaat daardoor namelijk een risico dat er geen zorgvuldige en betrouwbare beslissingen genomen kunnen worden over studenten, ondanks dat je dat met de invoering van programmatisch toetsen juist wel hebt beoogd.

Programmatisch toetsen is inmiddels ingevoerd in verschillende onderwijscontexten, in verschillende onderwijsdomeinen en met verschillende typen lerenden (Baartman et al., 2022). Er zijn vanuit onderzoek en vanuit ervaringen in de praktijk diverse aspecten te benoemen die je als opleiding mee kunt wegen in je keuze om al dan niet te gaan werken met het concept programmatisch toetsen. Deze aspecten lichten we hieronder toe.

### Visie op onderwijs is leidend

De visie op onderwijs, waar de visie op leren en de visie op toetsing onderdeel van uitmaken, zou altijd leidend moeten zijn voor de inrichting van toetsing (HAN, z.d). Er moet namelijk altijd sprake zijn van 'constructive alignment', wat inhoudt dat onderwijs en toetsing inherent aan elkaar zijn verbonden (Van Schilt-Mol & Baartman, 2023). Uit praktijkervaringen blijkt dat opleidingen, waarbij programmatisch toetsen succesvol is geïmplementeerd, voor dit concept gekozen hebben omdat zij dit zien als een middel om hun visie op onderwijs te realiseren (Baartman et al., 2020). Het concept past, volgens Van Schilt-Mol en Baartman (2023), bij de kijk op leren van deze opleidingen. En het past bij de signalen en ontwikkelingen die deze opleidingen oppikken uit de beroepspraktijk waartoe zij studenten opleiden.

Wil je als opleiding de studenten stimuleren om meer regie te nemen op het eigen leerproces? Wil je de zelfregulatie meer bevorderen en ruimte creëren voor reflectie en urgente leervragen? Dan kan programmatisch toetsen, volgens Biemond et al. (2021), passen bij jullie visie op onderwijs. Opleidingen die kiezen voor programmatisch toetsen proberen hiermee het zogenaamde 'hordelopen' te verminderen en het leerproces voor de langere termijn te stimuleren (Baartman et al., 2020).

Het concept moet ook passen bij het beroep waar je als opleiding toe opleidt. Programmatisch toetsen past goed bij een opleiding waarin wordt uitgegaan van de beheersing van de taak als geheel, waarin

gewerkt wordt met langere leerlijnen en waarin complexe vaardigheden centraal staan. Het past minder goed bij een onderwijsconcept dat gebaseerd is op beheersingsleren. Dus een concept waarin de beheersing van afzonderlijke onderdelen binnen een taak centraal staat (Platform Leren van Toetsen, z.d.). Eerder in dit whitepaper kwam dit al aan bod.

Alle onderwijs- en toetsconcepten hebben voor- en nadelen. Wat goed is, is het maken van keuzes passend bij jullie visie op onderwijs. Als het concept programmatisch toetsen niet past bij de onderwijsvisie van de opleiding is het dus geen goed idee om voor dit concept te kiezen.

Heb je als opleiding nog geen gedeelde visie op onderwijs? Biemond et al. (2021) raden aan om dan eerst eens bij deze visie stil te staan en pas daarna te bepalen of het concept programmatisch toetsen passend is bij die visie.

### **Vakoverstijgend denken (ook wel heldere ruggengraat)**

In het concept programmatisch toetsen wordt gewerkt met holistische leeruitkomsten. Een student heeft hierin een bepaalde mate van vrijheid om zelf, binnen bepaalde kaders keuzes te maken, zich te specialiseren of aan te sluiten bij vraagstukken die horen bij de eigen praktijk (Van Schilt-Mol & Baartman, 2023). De hoeveelheid vrijheid die de student heeft, is uiteraard afhankelijk van de ontwerpkeuzes die de opleiding maakt. Programmatisch toetsen vereist daardoor de omslag om niet te denken vanuit vakken, maar vanuit de leeruitkomsten (Van Schilt-Mol & Baartman, 2023).

Om als opleiding te kunnen denken vanuit (holistische) leeruitkomsten, is het van belang om deze leeruitkomsten ook doelbewust en integraal uit te denken (Baartman et al., 2020). Biemond et al. (2021) spreken in dit kader over de metafoor van een 'heldere ruggengraat' als raamwerk. Op het Platform Leren van Toetsen (z.d.) wordt dit op een heldere manier omschreven: "Een herkenbare ruggengraat van competenties, vaardigheden, beroepstaken of leeruitkomsten die de basis kunnen vormen voor de datapunten en de beslissingen". Wanneer er van een dergelijke ruggengraat nog geen sprake is, is het aan te raden deze als opleiding eerst uit te denken, alvorens over te gaan op programmatisch toetsen.

Programmatisch toetsen past minder goed bij opleidingen die het curriculum niet via leeruitkomsten hebben opgebouwd, maar dit hebben gedaan aan de hand van losstaande vakken. Ook voor opleidingen waar docenten zich sterk identificeren met hun vak, maar wellicht minder met het beroep dat centraal staat in de opleiding, is programmatisch toetsen minder passend. Bij programmatisch toetsen is het namelijk belangrijk dat docenten niet denken vanuit hun vak, dus alleen het vak centraal stellen, maar dat zij denken vanuit de leeruitkomsten, passend bij het beroep waar men toe opleidt. Docenten zouden zichzelf dus eigenlijk de vraag moeten stellen: Hoe verhouden de kennis en de vaardigheden uit mijn vak zich tot de leeruitkomsten, hoe kan ik daar een bijdrage aan leveren en hoe

kan ik mijn vak aanbieden zodat studenten zich op de leeruitkomsten ontwikkelen? (Van Schilt-Mol & Baartman, 2023).

Uit de praktijk blijkt dat het concept van programmatisch toetsen goed past bij curricula die doorgaande leerlijnen kennen. Is je curriculum georganiseerd in kleine en vooral losse modules die weinig met elkaar samenhangen, dan is programmatisch toetsen niet passend. Een curriculumherziening zou dan eerst nodig zijn (Platform Leren van toetsen, z.d.)

### **Lerende cultuur**

Ook de houding of cultuur van het team, maar eigenlijk van alle betrokkenen bij de opleiding, zijn sterk van belang wanneer een opleiding overweegt of programmatisch toetsen zou kunnen passen. De implementatie van programmatisch toetsen is een ingewikkeld proces waarin studenten, docenten en werkveldbegeleiders hun bestaande opvattingen over toetsing moeten aanpassen. Dit vraagt een lerende cultuur in de opleiding en een omslag wanneer van zo'n lerende cultuur nog geen sprake is. Het is van belang dat het geven van feedback, open staan voor feedback en het durven maken van fouten onderdeel wordt van de cultuur. Een student gaat namelijk, binnen een systeem van programmatisch toetsen, niet alleen feedback verzamelen om te laten zien dat deze iets beheerst, maar daarnaast ook vooral om van die feedback te kunnen leren. Zowel docenten als studenten zouden hierin geprofessionaliseerd moeten worden (Platform Leren van toetsen, z.d.).

Programmatisch toetsen vereist dus een cultuur waarin alle actoren omarmen dat studenten fouten mogen maken zonder dat dit (direct) consequenties heeft. De student verzamelt gedurende het leerproces feedback op verschillende datapunten. Deze feedback wordt vervolgens meegenomen in een portfolio op basis waarvan de besliscommissie uiteindelijk haar beslissingen neemt. Doordat er sprake is van een leerproces, en in het portfolio ook feedback op datapunten is opgenomen waaruit blijkt dat de student nog verdere ontwikkelstappen moe(s)t zetten, heeft het portfolio niet alleen een beeld van de beheersing, maar juist ook van de ontwikkeling van de student (Van Schilt-Mol & Baartman, 2023).

Voor het nemen van de zak-/slaagbeslissingen gaat een besliscommissie niet af op de feedback op individuele datapunten, maar gaat de commissie op zoek naar patronen in de feedback. Er wordt gekeken of het geheel de besliscommissie voldoende vertrouwen geeft dat de student zich voldoende heeft ontwikkeld. Met andere woorden: worden de leeruitkomsten door de student op het juiste niveau beheerst?

## **De rol van feedback en feedbackgeletterdheid van alle actoren**

Het verkrijgen van kwalitatief goede feedback is voor een student heel belangrijk, zeker in een systeem van programmatisch toetsen. Feedback dient in dit concept dan ook een dubbele functie: enerzijds is het een handvat voor studenten om sturing te geven aan hun leren en anderzijds baseert een besliscommissie zich op de feedback om uiteindelijk een beslissing te kunnen onderbouwen (Van Schilt-Mol & Baartman, 2023).

Zowel bij datapunten die vastgelegd zijn, als op andere momenten zoals bij leertaken, opdrachten en workshops is het van belang dat studenten gebruik maken van feedback. En niet alleen in de opleiding, maar ook in het werkveld (Biemond et al., 2021). Programmatisch toetsen vereist daarom een hoge mate van feedbackgeletterdheid. Als dit ontbreekt, wordt een succesvolle implementatie van programmatisch toetsen sterk bemoeilijkt. Het gaat dan niet alleen om de feedbackgeletterdheid van studenten, maar ook van docenten en partners in het werkveld. Daar past een leer- en feedbackcultuur bij, zoals eerder geschreven. Docenten zouden onder andere bereid moeten zijn om met elkaar mee te kijken, om elkaar feedback te geven en om studenten om feedback te vragen (Biemond et al., 2021).

## **Moed**

Biemond et al. (2021) spreken daarnaast ook over 'moed' die nodig is om programmatisch te toetsen; moed om 'gewoon te proberen', om fouten te maken en daarvan te leren, om zichzelf tijd te gunnen. Het invoeren van programmatisch toetsen gaat immers gepaard met grote veranderingen. Ideeën over en ervaringen met onderwijs en toetsing worden vervangen door nieuwe denkbeelden. Het toetsbeleid wijzigt en er is een andere beslisprocedure nodig. De rol van de docent krijgt een meer coachende vorm. Dit alles blijkt, zoals eerder gezegd, vaak een spannend proces.

## **Samenvatting van de belangrijkste aandachtspunten**

Allereerst is het belangrijk om je onderwijs op zo'n manier in te richten dat deze passend is voor de context, waarbij onderwijs en toetsing op elkaar afgestemd zijn. Mogelijk is programmatisch toetsen hierbij een passend concept. Hiervoor is het belangrijk dat:

- Er weloverwogen voor programmatisch toetsen wordt gekozen.
- Het concept passend is bij het onderwijsconcept en de (gedeelde) visie op onderwijs.
- Er gewerkt wordt met doorgaande leerlijnen (ook wel heldere ruggengraat).
- Er sprake is van een lerende cultuur.
- Er voldoende ruimte en aandacht is voor (de rol van) feedback(geletterdheid).
- Er moed wordt getoond!

### **3 BENIEUWD NAAR PRAKTIJKERVARINGEN?**

Op de website <https://www.programmatischtoetsen.nl/8-portretten-praktijk.html> zijn een aantal praktijkvoorbeelden opgetekend van opleidingen die programmatisch toetsen inzetten. Deze portretten geven je meer beelden uit de praktijk van de aspecten die we hierboven beschreven hebben als belangrijk in de keuze voor programmatisch toetsen. Ook kan de informatie op deze website je helpen als je de keuze hebt gemaakt voor programmatisch toetsen, want hoe ga je dan verder?



## LITERATUUR

- Agricola, B., De Vos, M., Baartman, L., & Van Schilt-Mol, T. (2023). *Naar een nieuwe balans in toetsfuncties: van formatief en summatief naar een continuüm van beslissingen. Elf ontwerpkeuzes in programmatisch toetsen* [Factsheet]. Hogeschool Utrecht, Utrecht, Nederland. <http://dx.doi.org/10.13140/RG.2.2.30798.64324>.
- Agricola, B., De Vos, M., Baartman, L., & Van Schilt-Mol, T. (2024). *Naar een nieuwe balans in toets functies: van formatief en summatief naar een continuüm van beslissingen* (Projectnummer 40.5.22945.016). NRO Overzichtsstudies voor het Hoger Onderwijs.
- Baartman, L. K. J., van Schilt-Mol, T., & Van der Vleuten, C. P. M. (2020). *Programmatisch toetsen. Voorbeelden en ervaringen uit de praktijk*. Boom Hoger onderwijs.
- Baartman, L., Van Schilt-Mol, T., & Van der Vleuten, C. (2022). Programmatic assessment design choices in nine programs in higher education. In *Frontiers in Education* (Vol. 7, p. 931980). Frontiers. <https://doi.org/10.3389/educ.2022.931980>.
- Biemond, I., Kaijen, M., & Peeters, W. (2021). *Programmatisch toetsen in het hbo, dat vraagt moed!*. Geraadpleegd op 12 maart 2024, van <https://lerenvantoetsen.nl/programmatisch-toetsen-in-het-hbo-dat-vraagt-moed/>
- De Jong, L. H., Bok, H. G., Kremer, W. D. J., & van der Vleuten, C. P. M. (2019). Programmatic assessment: Can we provide evidence for saturation of information? *Medical Teacher*, 41(6), 678–682. <https://doi.org/10.1080/0142159X.2018.1548909>.
- HAN (z.d.). *Het Toetsweb*. Geraadpleegd op 22 maart 2023, van <https://www.han.nl/artikelen/2021/03/het-toetsweb-duurzame-kwaliteit-van-toetsing/>.
- Heeneman, S., De Jong, L. H., Dawson, L. J., Wilkinson, T. J., Ryan, A., Tait, G. R., ... & van der Vleuten, C. P. (2021). Ottawa 2020 consensus statement for programmatic assessment–1. Agreement on the principles. *Medical teacher*, 43(10), 1139-1148. <https://doi.org/10.1080/0142159X.2021.1957088>.
- Jamieson, J., Jenkins, G., Beatty, S., & Palermo, C. (2017). Designing programmes of assessment: A participatory approach. *Medical Teacher*, 39(11), 1182-1188. <https://doi.org/10.1080/0142159X.2017.1355447>.
- Last, B., & Jongen, S., (2021). *Blended Learning en onderwijsontwerp: Van theorie naar praktijk*. Boom Uitgevers.
- Platform leren van toetsen. (z.d.). *Programmatisch toetsen*. Geraadpleegd op 12 maart 2024, van <https://lerenvantoetsen.nl/programmatisch-toetsen/>
- Platform leren van toetsen. (z.d.). *FAQ Programmatisch toetsen*. Geraadpleegd op 12 maart 2024, van <https://lerenvantoetsen.nl/programmatisch-toetsen-2/>
- Van der Vleuten, C., Heeneman, S., & Schut, S. (2019). Programmatic assessment: an avenue to a different assessment culture. In *Assessment in health professions education* (pp. 245-256). Routledge.

- Van der Vleuten, C., Heeneman, S., & Schuwirth, L. W. (2017). Programmatic assessment. In Dent, J., Harden, R., Hunt, D. (eds). *A Practical Guide for Medical Teachers*. Elsevier.
- Van der Vleuten, C., Schut, S., en Heeneman, S. (2018). Programmatisch toetsen als motor voor professioneel leren in het hoger onderwijs. In: Sluijsmans, D. en M. Segers (red.), *Toetsrevolutie. Naar een feedbackcultuur in het hoger onderwijs*. Phronese.
- Van der Vleuten, C. P., & Schuwirth, L. W. (2005). Assessing professional competence: From methods to programmes. *Medical Education*, 39(3), 309–317. <https://doi.org/10.1111/j.1365-2929.2005.02094.x>.
- Van der Vleuten, C. P., Schuwirth, L. W. T., Driessen, E. W., Dijkstra, J., Tigelaar, D., Baartman, L. K. J., & Tartwijk, J. van. (2012). A model for programmatic assessment fit for purpose. *Medical Teacher*, 34(3), 205–214. <https://doi.org/10.3109/0142159X.2012.652239>.
- Van der Vleuten, C. P. M., Schuwirth, L. W. T., Scheele, F., Driessen, E. W., & Hodges, B. (2010). The assessment of professional competence: building blocks for theory development. *Best practice & research Clinical obstetrics & gynaecology*, 24(6), 703-719. doi:10.1016/j.bpobgyn.2010.04.001.
- Van der Vleuten, C., Lindemann, I., & Schmidt, L. (2018). Programmatic assessment: the process, rationale and evidence for modern evaluation approaches in medical education. *Medical journal of Australia*, 209(9), 386-388. <https://doi.org/10.5694/mja17.00926>.
- Van Schilt-Mol, T., & Baartman, L. (2023). Programmatisch toetsen. In Van Berkel, H., Bax, A., Joosten-ten Brinke, D., Beekman, K., & Van Schilt-Mol, T. (reds). *Toetsen in het hoger onderwijs*. Boom Hoger onderwijs.
- Van Schilt-Mol, T., Baartman, L., & Van der Vleuten, C. (nog te verschijnen). *Programmatisch toetsen: Een verdieping vanuit theorie en praktijk*. Boom Hoger onderwijs.
- Schut, S., Van Tartwijk, J., Driessen, E., Van der Vleuten, C., & Heeneman, S. (2020). Understanding the influence of teacher-learner relationships on learners' assessment perception. *Advances in Health Sciences Education*, 25(2), 441-456. <https://doi.org/10.1007/s10459-019-09935-z>.

**OPEN UP**  
**NEW** **HAN\_** UNIVERSITY  
OF APPLIED SCIENCES  
**HORIZONS.**