

# Programmatisch toetsen: waarom, wat en hoe?

written by Wessel Peeters

17 juni 2019



*Binnen het hogerberoepsonderwijs is een nieuwe vorm van toetsen in opkomst: programmatisch toetsen. Een vorm van toetsen en leren met als doel om breder te kijken naar de student als geheel, om zo een valide oordeel te vellen over de competenties van de student. Waarom is dat wenselijk? Wat is programmatisch toetsen dan precies? Hoe kun je er mee aan de slag? Dat en meer lees je in dit artikel!*

## Waarom programmatisch toetsen

De laatste jaren wordt er steeds kritischer gesproken en geschreven over toetsen als de cito, examens en tentamens. Daarbij gaat het vooral over de vraag wat een toets nu precies zegt over het niveau van een leerling of student of in hoeverre

het bijdraagt aan het leren of dat het dat juist, mede door de bijkomende stress, in de weg zit. Toetsen zouden een duidelijke meerwaarde moeten hebben (betekenisvol moeten zijn) en een onderdeel van het leerproces moeten zijn.

Het gesprek is onderdeel van een breder maatschappelijke gesprek over onderwijs, waarbij de focus steeds meer komt te liggen op het leren als continue proces, in plaats van het beoordelen en becijferen van dat leren. Zo gaat het bijvoorbeeld steeds meer over het belang van heldere doelen, succescriteria, feedback en hulp bij de vervolgstappen van de leerlingen of student ([feedup](#), [feedback en feedforward](#)): een programma vol met toetsen geeft hier weinig ruimte voor en kan zelfs een negatieve werking hebben op het leerrendement. Ook zou onderwijs meer ruimte moeten bieden aan individuele verschillen, oftewel een [meer gepersonaliseerd](#) (gedifferentieerd) aangeboden moeten kunnen worden. De gesprekken hebben geleid voor een toegenomen aandacht voor het [sociaal-constructivisme](#) (zie ook [John Dewey](#)): een stroming die het leerproces ziet als een actief proces van kennisverwerving, waarbij de kennis ontstaat en gedeeld wordt met anderen. Belangrijke principes daarbij zijn het voortbouwen op de aanwezige voorkennis (zie ook de [Zone van Naaste Ontwikkeling](#) van [Lev Vygotsky](#)), het stimuleren van het zelfregulerende vermogen van leerlingen (zie ook de [Cognitive Apprenticeship](#) theorie), fouten maken mag en het leren binnen de context.

Er zijn daarbij verschillende onderzoeken die aansluiten bij het gesprek over onderwijs en leren. Voorbeelden daarvan zijn onderzoeken waaruit naar voren komt hoe belangrijk [het inspelen op voorkennis](#) is (ook wel [scaffolding](#) genoemd), onderzoeken waaruit naar voren komt dat leerlingen en studenten [stoppen met leren zodra het cijfer bekend is](#) en onderzoeken die ingaan op het belang van het geven van [effectieve feedback](#). Daarnaast komt uit onderzoek ook naar voren dat de vaak kort gegeven toetsen van 30-60 minuten [niet erg veel zeggen](#) over de capaciteiten van leerlingen en studenten: ze gaan slechts kort in op één specifieke onderdeel van een groot geheel:

Toets-Tijd in Hours	MCQ <sup>1</sup>	Korte casus gerichte Essay <sup>2</sup>	Simulaties <sup>1</sup>	Mon-de-ling <sup>3</sup>	Patient examen <sup>4</sup>	OSCE <sup>5</sup>	KKB <sup>6</sup>	Praktijk Video Observatie <sup>7</sup>	In-cognito SPS <sup>8</sup>
1	0.62	0.68	0.36	0.50	0.60	0.54	0.73	0.62	0.61
2	0.76	0.73	0.53	0.69	0.75	0.69	0.84	0.76	0.76
4	0.93	0.84	0.69	0.82	0.86	0.82	0.92	0.93	0.82
8	0.93	0.82	0.82	0.90	0.90	0.90	0.96	0.93	0.86

<sup>1</sup>Norcini et al., 1985

<sup>2</sup>Stalenhoef-Halling et al., 1990

<sup>3</sup>Swanson, 1987

<sup>4</sup>Wass et al., 2001

<sup>5</sup>Van der Vleuten, 1988

<sup>6</sup>Norcini et al., 1999

<sup>7</sup>Ram et al., 1999

<sup>8</sup>Gorter, 2002

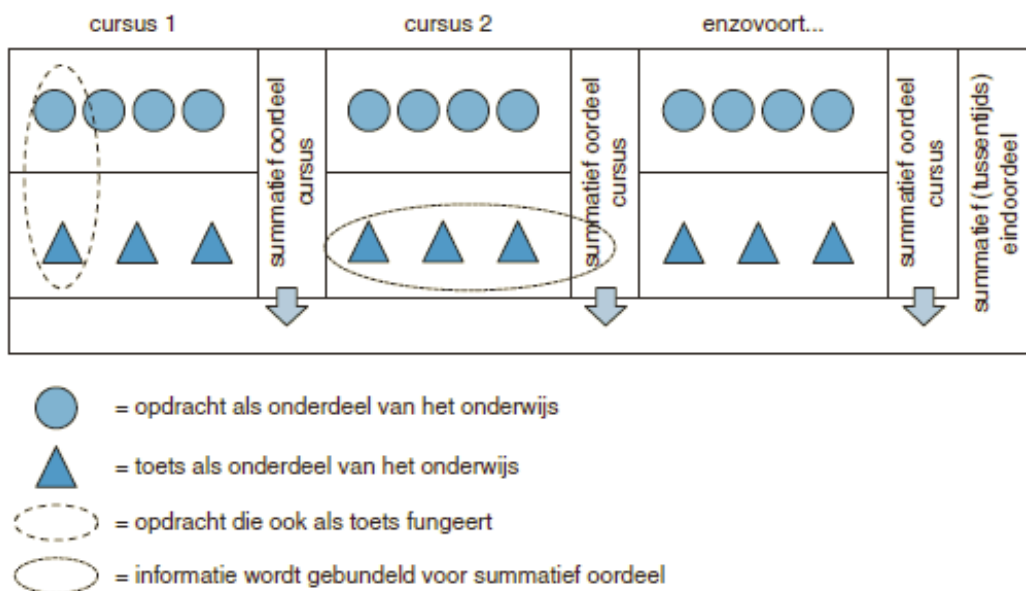
De percentiel score geeft aan hoe betrouwbaar een soort toets is naarmate je er meer tijd aan besteed. Hoe hoger de score, hoe betrouwbaarder de toets.

De maatschappelijke beweging en onderzoeken geven richting aan een nieuwe vorm van leren en toetsen. Een inmiddels ingeburgerde ontwikkeling daarbij is het [formatief evalueren](#) of handelen, waarbij veel van de hierboven genoemde aspecten rondom leren naar voren komen. Daarbij blijft in de praktijk echter vaak wel staan dat dat je uiteindelijk naar één toets(moment) toe werkt, waarbij het gaat over één soort product, wat onderdeel is van een groot geheel. In hoeverre zegt dat ene moment nu iets over de student in z'n geheel? En in hoeverre is het noodzakelijk dat aparte toetsmoment te hebben, als er tijdens de lessen of daarbuiten al voldoende wordt aangetoond? En hoeverre kun je binnen die toetsmomenten belangrijke aspecten meenemen die niet in de toets zitten, zoals de zelfstandigheid van een leerling of student? In hoeverre kan het professionele oordeel van de docent onderdeel zijn van het beoordelingsproces, aangezien niet alles met papier te meten is? Dergelijke vragen gaan over de validiteit van toetsen - iets anders dan de objectiviteit. Het vraagt om een meer holistische benadering van leren en toetsen, en dat is precies waar het programmatisch toetsen over gaat.

## Wat is programmatisch toetsen

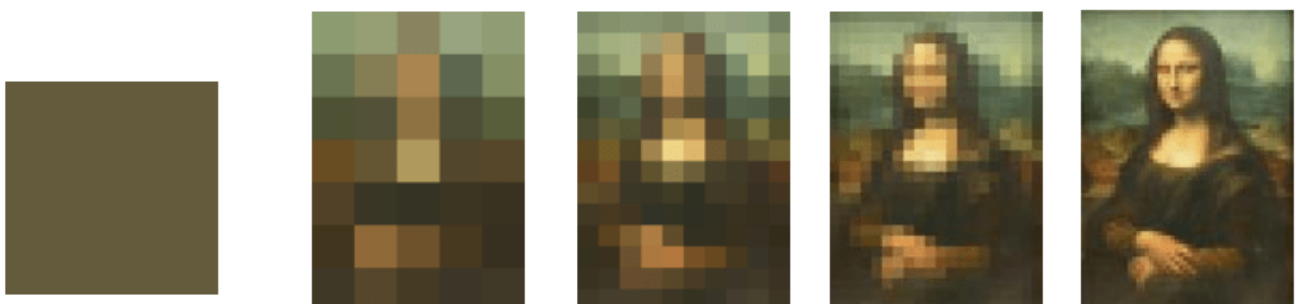
Programmatisch toetsen is een toetsvorm waarmee enkele hbo opleidingen in Nederland momenteel werken en kan worden gezien als [een sleutel tot de feedbackcultuur](#). Programmatisch toetsen gaat uit van de gedachte dat de student

tijdens zijn opleiding complexe vaardigheden en competenties ontwikkeld, die hij nodig heeft op de arbeidsmarkt. Deze complexe vaardigheden zijn niet in één toets te vatten en vragen om een bredere kijk. Daarbij is het zo dat bij de traditionele manier van toetsen een zak- of slaagregeling hangt aan ieder toets. Bij programmatisch toetsen zijn toetsen onderdeel van het leerproces van de student en gebruik je deze om te sturen en stimuleren, en om beslissingen over de voortgang te onderbouwen. Er wordt daarbij uitgegaan van langere leerlijnen en veel verschillende toetsmomenten. Deze momenten zijn bedoeld om feedback te geven en het zo tot een leermoment te maken. Het kan worden gezien als een ver doorgevoerde vorm van formatief evalueren.



Van Vleuten et al. (2012)

Het bovenstaande figuur kun je vergelijken met een foto: als je één enkele pixel ziet, heb je nog geen idee hoe het totaalplaatje eruit ziet. Bij programmatisch toetsen verzamelt de student voldoende informatie om een haarscherp beeld te creëren. Op basis van het totale plaatje worden beslissingen genomen.



Cees van der Vleuten

Meer concreet gaat het er bij programmatisch toetsen om dat je als docent(en) uiteindelijk een beslissing neemt (beoordeling geeft) op basis van verschillende informatie (data). Deze data is als het goed is veelzijdig en tot stand gekomen door een leerproces waar feedback een cruciaal onderdeel van is geweest. Het beoordelingsmoment zegt dus als het goed is veel over de competenties van de student.

## Hoe ga je aan de slag met programmatisch toetsen

Om met programmatisch toetsen aan de slag te gaan, is het belangrijk om aandacht te hebben voor het curriculum, het systeem en het team.

### Het curriculum

Programmatisch toetsen bestaat uit een aantal onderdelen: low-stakes (datapunten), medium-stake momenten en high-stake momenten. De datapunten zijn toetsen, lesmomenten, stagemomenten, etc. op basis waarvan studenten feedback verzamelen over de [leeruitkomsten](#) of indicatoren waaraan zij werken, van hun peers (zie ook [dit artikel over peerfeedback](#)), verschillende docenten, stagebegeleider, etc. Daarbij is het belangrijk dat studenten op verschillende momenten en op verschillende manieren aan de leeruitkomsten kunnen werken en dat de datapunten verschillende type producten en handelingen zijn, om zo een allround beeld te geven. Concreet zou een student bijvoorbeeld gedurende een periode een stageverslag, casuspresentatie, groepsopdracht in de les en een stagebezoek kunnen gebruiken om feedback te vragen op de leeruitkomsten die daar voor staan, wellicht aangevuld met eigen datapunten en bijbehorende feedback. Deze feedback is naast kwalitatief, ook kwantitatief: het moet wat zeggen over in hoeverre een student een bepaalde competentie beheerst. Zo kan er bijvoorbeeld gekozen worden voor een [single-point-rubric](#) met daaronder een schaal van 1 tm 5: wat ging goed, wat kan beter en hoe heeft de student het gedaan bij die leeruitkomst. Deze kwalitatieve data kan dan bijvoorbeeld worden gebruikt in een spinnenweb van de competenties of eindkwalificaties. Bij datapunten staat er niet veel op het spel, oftewel 'low-stakes'.

Medium-stake momenten zijn formatieve evaluaties, waarbij de student met zijn begeleider (bijv. slb'er) kijkt hoe hij zich de afgelopen weken heeft ontwikkeld op

de verschillende competenties of eindkwalificaties, waar de leeruitkomsten uit voortkomen. Dit zijn momenten waarop studenten feedback ontvangen, die meer nog dan inhoudelijk, vaak gaat over de algemene houding, progressie, etc. Op basis van deze formatieve momenten kan de begeleider uiteindelijk ook adviseren of een student goed past bij de opleiding en wat hij of zij nodig heeft om verder te komen.

High-stake momenten zijn summatieve evaluaties waarbij een panel, vaak bestaande uit verschillende docenten en waar mogelijk ook externe partners, op basis van de datapunten en medium-stake momenten beoordelen of een student studiepunten krijgt toegewezen: alles of niets ('high stakes'). Daarbij wordt er allereerst gekeken naar het grote geheel: de medium-stake momenten. Waar nodig wordt er verder ingezoomd, zodat de losse datapunten en bijbehorende feedback zichtbaar zijn. Waar nodig wordt er weer verder ingezoomd zodat de losse producten zichtbaar zijn. Kortom: een panel graaft zo diep als nodig is om tot een valide oordeel te komen over een student. Het belangrijkste is dat een high-stake moment een moment is dat er écht toe doet (high-stakes) en dan ook écht wat zegt over de student, maar géén verrassing mag zijn: er is immers continue feedback gegeven aan de student.

Concreet zou een programmatisch programma er als volgt uit kunnen zien (1e semester, propedeuse):

Semester 1	Zelfvervalsleren	Doelgroepeleren	Praktijkleren	
Week 1				
Week 2			<b>Datapunt 3</b> - leeruitkomst 5 - leeruitkomst 7	
Week 3	<b>Datapunt 1</b> - leeruitkomst 1 - leeruitkomst 2	<b>Datapunt 2</b> - leeruitkomst 4 - leeruitkomst 6 - leeruitkomst 3		<b>Low Stake 1 (gesprek)</b> - Datapunt 1, 2 en 3 + Algemene indruk
Week 4				
Week 5				
Week 6				
Week 7		<b>Datapunt 5</b> - leeruitkomst 5 - leeruitkomst 6		
Week 8	<b>Datapunt 4</b> - leeruitkomst 1 - leeruitkomst 7 - leeruitkomst 4			<b>Low Stake 2 (gesprek)</b> - Datapunt 1, 2 en 3 - Datapunt 4 en 5 + Algemene indruk
Week 9				
Week 1				
Week 2			<b>Datapunt 6</b> - leeruitkomst 2 - leeruitkomst 6 - leeruitkomst 7	
Week 3				
Week 4		<b>Datapunt 7</b> - leeruitkomst 3 - leeruitkomst 4		
Week 5			<b>Datapunt 8</b> - leeruitkomst 5 - leeruitkomst 7	<b>Low Stake 3 (gesprek)</b> - Datapunt 4 en 5 - Datapunt 6, 7, 8 en 9 + Algemene indruk
Week 6	<b>Datapunt 9</b> - leeruitkomst 1 - leeruitkomst 2 - leeruitkomst 3			
Week 7				
Week 8				
Week 9: Portfolio	<ol style="list-style-type: none"> <li>1. De student kan op basisniveau afstemmen op de ander, een werkrelatie opbouwen en deze onderhouden.</li> <li>2. De student legt doel en werkwijze van zijn interventie helder uit aan de ander.</li> <li>3. De student observeert en interpreteert aan de hand van het LECS-model en beschrijft daarbij binnen de lichamelijke aspecten de uitdrukingsmotoriek.</li> <li>4. De student voert in overeenstemming met de cliënt een interventie uit op basis van een rationale.</li> <li>5. De student toont aan een kleinschalig onderzoek te kunnen uitvoeren met behulp van een meetinstrument en literatuur.</li> <li>6. De student heeft een realistisch beginnend beroepsbeeld en kan dit verwoorden naar een medestudent.</li> <li>7. De student houdt in de samenwerking met medestudenten en docenten rekening met formele regels en randvoorwaarden en communiceert helder met alle betrokken partijen.</li> </ol>			<b>High Stake 1</b>  Beoordeling portfolio. Hierin wordt gekeken naar advies + rubrics. Waar nodig naar low-stake momenten en waar verder nodig naar producten.  <ol style="list-style-type: none"> <li>1. Advies SLB'er</li> <li>2. Beoordelingsrubrics</li> <li>3. Low Stake 1, 2 en 3</li> <li>4. Datapunt 1 tm 9</li> </ol>

Het hierboven getoonde voorbeeld betreft er een waarbij is gekozen voor 3 leerlijnen (vakken). De datapunten zijn in dit geval 9 producten en handelingen die naar de studenten toe verder zijn gespecificeerd, zoals een verslag of casuspresentatie. In dit voorbeeld krijgen studenten na één semester 30 of 0 studiepunten: er is dan dus veel ruimte om te leren en feedback om te ontvangen.

Het volgende high-stake moment is na periode 3 (15 studiepunten) en periode 4 (15 studiepunten): deze knip is gemaakt in verband met het bindend studieadvies (BSA), waarbij het wenselijk is om na periode 3 studiepunten te kunnen geven. In hoofdfase 1, 2 en 3 kan er worden gewerkt in twee semesters of zelfs een heel studiejaar. In OS OER zijn het in dit geval drie onderwijseenheden, met ieder een portfolio toets als afsluiting.

Dagdelenonderwijs met meerdere docenten is bij programmatisch toetsen aan te raden, zo niet voorwaardelijk voor een succes: het stelt studenten in staat om gedurende langere tijd ergens aan te werken en feedback te verzamelen van verschillende docenten. Het geeft docenten daarnaast de mogelijkheid om meer zicht te hebben op de progressie van studenten en het onderwijs aan te passen aan de studenten.

### **Het systeem**

Al hoewel het digitale systeem niet bepalend is, speelt het bij programmatisch toetsen een relatief grote rol: het waterdicht verzamelen van feedback, uploaden van bewijslast, registreren van medium-take momenten, etc, vragen om een digitale infrastructuur die daarbij kan ondersteunen: dit is iets om niet te onderschatten. Wil je bezig met programmatisch toetsen, kijk hier dan op tijd naar: het succes van programmatisch toetsen staat of valt er in dit geval mee. Er zijn gelukkig verschillende bedrijven bezig met software op dit gebied. Een voorbeeld daarvan is [het programma Scorion](#). Mocht je bezig gaan met programmatisch toetsen, dan zul je er vanzelf meer tegenkomen. Mocht je vragen hebben? Laat dit dan gerust weten middels een reactie of mail.

### **Het team**

Het allerbelangrijkste bij programmatisch toetsen is het (docenten)team: hoe goed je een curriculum en systeem ook inricht, uiteindelijk draait het - zoals altijd- om het docentengedrag. Zo is het belangrijk korte lijntjes met elkaar te houden, te willen vallen en opstaan en om een gezamenlijke visie op leren te hebben en te weten hiernaar te handelen: begrippen als feedback en [scaffolding](#) moeten onderdeel zijn van de (professionele) cultuur om programmatisch toetsen tot een succes te maken, en dat kost tijd. Het is daarbij aan te raden om actief te werken aan de didactische en pedagogische bekwaamheid van een team, bijvoorbeeld door het te integreren in het BKE of SKE traject.

### **Update: het eerste jaar**



Hoe heeft de ontwerp in de praktijk uitgepakt? [Lees hier deel 1 van het tweeluik](#) wat door een docent is geschreven naar aanleiding van het eerste jaar.

## Verder lezen

Programmatisch toetsen is interessant, maar ook relatief nieuw en complex. Wil je je verder verdiepen? Hieronder een aantal interessante links:

- 12 tips voor programmatisch toetsen: <https://www.e-xamens.nl/nieuws/programmatisch-toetsen>
- Lezing over programmatisch toetsen (Cees van Vleuten): <https://lerenvantoetsen.nl/keynote-cees-van-der-vleuten-symposium-2016/>
- Powerpoint over programmatisch toetsen (Cees van Vleuten): <https://lerenvantoetsen.nl/wp-content/uploads/2016/06/keynote-cees-van-der-vleuten.pdf>
- Boek over programmatisch toetsen: <https://lerenvantoetsen.nl/boek-over-programmatisch-toetsen/>

Wil je als opleiding of school aan de slag met programmatisch toetsen? Met onze ervaring en kennis hierover helpen we graag. [Neem vrijblijvend contact met ons](#) op voor een kennismaking.

---

Dit artikel is afkomstig van [Vernieuwenderwijs.nl](http://Vernieuwenderwijs.nl).

Bekijk de meeste recente versie van ons artikel op onze website.



**Vernieuwenderwijs**