

EIGENTIJDS BEOORDELEN EN BESLISSEN



Tamara van Schilt-Mol

HAN_UNIVERSITY
OF APPLIED SCIENCES

EIGENTIJDS BEOORDELEN EN BESLISSEN

Een pleidooi om beoordelen weloverwogen en bewust in te zetten als integraal onderdeel van onderwijs zodat zorgvuldige beslissingen over het leren van leerlingen genomen kunnen worden.

Intreerede, in verkorte vorm uitgesproken bij de aanvaarding van het ambt van lector Eigentijds Beoordelen en Beslissen aan de Hogeschool van Arnhem en Nijmegen op 15 april 2021.

Dr. Tamara van Schilt-Mol

COLOFON

HAN University of Applied Sciences, Academie Educatie
Onderzoeksteam Kwaliteiten van Leraren
Lectoraat Eigentijds beoordelen en beslissen
Tamara van Schilt-Mol

REALISATIE

HAN MCV Marketing- en communicatie, Academie Educatie

VORMGEVING

HAN Studio MC, Roswitha Teerink

Uitgave HAN University of Applied Sciences Press, Arnhem, maart 2022

© Alles uit deze uitgave mag worden verveelvoudigd en/of openbaar gemaakt door middel van druk, fotografie, microfilm, geluidsband, of op welke andere wijze dan ook, zonder voorafgaande toestemming van de auteur en uitgever, mits er zorgvuldig wordt verwezen naar de auteur en de uitgever.

NHOUDSOPGAVE

NHOUDSOPGAVE	5
SAMENVATTING	7
1. INLEIDEND	13
2. WAAROM AANDACHT VOOR EIGENTIJDS TOETSEN EN BEOORDELEN?	17
3. KWALITEIT VAN TOETSING	23
Kwaliteit van toetsing geoperationaliseerd	23
Ontwikkeling van het toetsweb	25
4. EIGENTIJDS BEOORDELEN EN BESLISSEN	35
Eigentijds beoordelen en beslissen: een definitie	35
Onderscheid tussen beoordelen en beslissen	36
Informatierijk beoordelen	37
Zorgvuldige beslissingen	38
5. FUNCTIES VAN TOETSEN	43
Leren, beslissen en evalueren	43
Leerfunctie: toetsen om het leren van leerlingen te stimuleren en te sturen	45
Beslisfunctie: toetsen om zorgvuldige beslissingen te kunnen nemen	50
Decision-driven data-collection	56
Evaluatiefunctie: toetsen om de kwaliteit van het onderwijs te optimaliseren	57
Balans tussen functies: toetssystemen	62
6. EIGENTIJDS BEOORDELEN EN BESLISSEN IN DE PRAKTIJK: PRAKTISCHE HANDVATTEN	69
7. ONDERZOEKSAGENDA VAN HET LECTORAAT	75
8. TOT SLOT	81
REFERENTIES	84

SAMENVATTING

Het lectoraat “Eigentijds beoordelen en beslissen” focust op de vraag hoe eigentijds beoordelen en beslissen ontworpen, ingericht en uitgevoerd kan worden zodat het bijdraagt aan de drie functies die onderscheiden worden binnen toetsing:

1. de leerfunctie: het stimuleren en sturen van het leren van leerlingen;
2. de beslisfunctie: het nemen van zorgvuldige beslissingen, en
3. de evaluatiefunctie: het optimaliseren van de kwaliteit van het onderwijs.

Onder “eigentijds beoordelen en beslissen” wordt het volgende verstaan: *“het bewust en weloverwogen inzetten van toetsen als integraal onderdeel van het onderwijs, waarbij op basis van informatierijke beoordelingen zorgvuldige beslissingen genomen worden over (het leren van) leerlingen en de kwaliteit van het onderwijs wordt geoptimaliseerd”*.

Met ‘bewust en weloverwogen inzetten van toetsen’ wordt bedoeld op het maken van evidence-informed (weloverwogen) ontwerpkeuzes die aansluiten bij de visie op onderwijs (bewust). Door het bewust en weloverwogen inzetten van toetsen kan recht worden gedaan aan zowel de leerfunctie (sturen en stimuleren van het leren) als de beslisfunctie (het nemen van zorgvuldige beslissingen). Door met informatie uit toetsen te reflecteren op hun eigen handelen, kunnen docenten recht doen aan de evaluatiefunctie van toetsen (het optimaliseren van de kwaliteit van het onderwijs).

In de definitie wordt een onderscheid gemaakt tussen beoordelen en beslissen. Met beoordelen wordt bedoeld het vaststellen van waar de leerling zich bevindt ten opzichte van zijn leerdoel, met andere woorden het relateren van de geleverde prestatie(s) aan de beoogde leerresultaten. Op basis van de beoordeling wordt gerichte feedback en feedforward gegeven. Met beslissen wordt bedoeld op het verbinden van een consequentie aan een of meerdere beoordelingen. Deze beslissingen kunnen gericht zijn op de (verdere) ontwikkeling van de leerling, op de voortgang ten aanzien van de leerdoelen en op het vaststellen van wat geleerd is (certificeren, kwalificeren, selecteren, etc.).

BEOORDELEN:**VASTSTELLEN WAAR DE LEERLING ZICH BEVINDT TEN OPZICHTE VAN ZIJN LEERDOEL**

De beoordeling geeft altijd handvatten voor de ontwikkeling van de leerling doordat deze altijd wordt voorzien van feedback (feedup, feedback én feedforward). Beoordelingen kunnen worden gegeven op basis van leer/toetsactiviteiten als kennistoetsen, stage, beroepsproducten, peerfeedback, selfassessment, mentorgesprek, etc.

BESLISSEN:**VERBINDEN VAN EEN CONSEQUENTIE AAN DE BEOORDELING**

Het belang van de beslissing is afhankelijk van mate waarin er voor de leerling iets op het spel staat.

Weinig belang: beslissing is gericht op (verdere) ontwikkeling, informatie wordt gebruikt om het onderwijs(leerproces) bij te sturen.

Redelijk belang: beslissing is gericht op voortgang, informatie wordt gebruikt om het onderwijs(leerproces) bij te sturen.

veel belang: beslissing is gericht op het vaststellen van wat geleerd is, informatie wordt gebruikt om kwalificaties als zakken/slagen, (on)voldoende, (niet) overgaan te kunnen onderbouwen.

RELATIE TUSSEN BEOORDELEN EN BESLISSEN

Beoordelen en beslissen zijn twee activiteiten die elkaar opvolgen. Het is essentieel dat bij het ontwerpen van het onderwijs expliciet wordt nagedacht over het type en de hoeveelheid informatie die nodig is om een zorgvuldige beslissing te nemen. Een beslissing kan worden genomen op basis van één beoordeling. Voor beslissingen waarbij voor de leerling veel op het spel staat, wordt bij voorkeur gebruik gemaakt van meerdere beoordelingen.

Het lectoraat hanteert bij het beantwoorden van de centrale lectoraatsvraag naast de omschrijving van eigentijds beoordelen en beslissen het toetsweb als theoretische basis. In deze rede wordt een nieuwe versie van het in 2016 ontwikkelde web gepresenteerd. Dit web bestaat uit één voorwaardelijk aspect, de toetsvisie. Daarnaast omvat het web vijf toetsentiteiten die gezamenlijk van belang zijn bij het ontwerpen van toetsing: (1) het toetsbeleid, (2) de toetsorganisatie, (3) het toetsprogramma, (4) de toetsen en toetstaken en (5) de toetsbekwaamheid. Voor deze zes aspecten geldt dat deze niet los van elkaar

gezien kunnen worden en onderling verweven zijn: is de kwaliteit op een van de aspecten niet op orde, dan heeft dat direct implicaties voor de kwaliteit van de andere aspecten.

In aansluiting op recente discussies over de kwaliteit van toetsing en de huidige toetscultuur, en in aansluiting op de recente constatering dat zowel de toetsbekwaamheid van docenten als de kwaliteit van de toetsprogramma's te vaak te wensen overlaat (Onderwijsraad, 2018; Commissie Kwaliteit Schoolexaminering, 2018), richt het lectoraat zich bij het beantwoorden van de centrale lectoraatsvraag op het ontwikkelen van kennis over en het achterhalen van werkende factoren ten aanzien van:

- de ontwerpeisen voor een toetsprogramma, zodanig dat dit programma niet alleen leidt tot het realiseren van de beoogde leerresultaten, maar ook bijdraagt aan de drie functies van toetsing;
 - de (toets)bekwaamheid van docenten die nodig is om eigentijds beoordelen en beslissen te ontwerpen, te implementeren en uit te voeren;
 - de (toets)bekwaamheid van leerlingen die nodig is om met informatie uit het toetsprogramma daadwerkelijk invulling te kunnen geven aan het (nemen van regie over het) eigen leerproces;
 - de voorwaarden die aan een schoolorganisatie gesteld moeten worden om eigentijds beoordelen en beslissen te implementeren,
- en op het toegankelijk en toepasbaar maken van deze kennis voor de onderwijspraktijk.

Om invulling te kunnen geven aan *eigentijds beoordelen en beslissen* worden in deze rede zes handvatten gepresenteerd. Deze handvatten ondersteunen het pleidooi om beoordelen weloverwogen en bewust in te zetten als integraal onderdeel van onderwijs zodat zorgvuldige beslissingen over het leren van leerlingen genomen kunnen worden. De handvatten kunnen helpen om de drie onderscheiden functies van toetsen die de kern vormen van eigentijds beoordelen en beslissen te realiseren, ongeacht het toetsstelsel dat binnen een school/opleiding wordt gehanteerd. De zes handvatten zijn:

1. Maak bewuste en geïnformeerde ontwerpkeuzes
2. Maak onderscheid tussen beoordelen en beslissen
3. Verzamel informatie ten behoeve van de beslissing die je wil nemen
4. Ga in gesprek met de leerling over de waarde van beoordelingen
5. Geef en benut informatierijke beoordelingen
6. Blijf bewust toetsbekwaam

Enkele beoogde opbrengsten van het lectoraat zijn:

- Voorbeelden van de wijze waarop in verschillende onderwijssectoren en binnen verschillende toetsystemen invulling gegeven wordt aan de zes handvatten.
- Nadere uitwerking van de wijze waarop de drie functies van toetsen vorm kunnen krijgen in de verschillende onderwijssectoren, zodanig dat deze elkaar versterken.
- Handreikingen waarmee de ontwikkelde kennis ten aanzien van het ontwerpen, inrichten en uitvoeren van eigentijds beoordelen en beslissen toegepast kan worden binnen de eigen onderwijspraktijk.
- Handreikingen waarmee, expliciet vanuit het perspectief van eigentijds beoordelen en beslissen, en met oog voor de verschillende onderwijssectoren en toetsystemen, de kwaliteit van de vijf toetsentiteiten die een rol spelen bij toetsing kritisch bezien én geoptimaliseerd kan worden.



1. INLEIDEND

Geacht college, lieve familie, beste collega's en studenten,

Wat fijn dat ik deze rede met jullie mag delen. En wat spannend. Wellicht kunnen jullie je voorstellen hoe de periode richting de installatie is geweest. Het nadenken over de boodschap die je in de installatierede wil meegeven, het opschrijven daarvan, het voorbereiden van de workshops rondom de installatie. Hoe je het ook wendt of keert, en hoe blij ik ook ben met de kans die ik krijg om dit mooie lectoraat te vervullen en vorm te gaan geven, het is spannend. Het leverde veel hoofdbreken op: wat wil ik zeggen, wat ga ik opschrijven, hoe zal het ontvangen worden? Want hoe dan ook wil je graag een zinvolle bijdrage leveren aan het gesprek dat over beoordelen en beslissen wordt gevoerd.

Het maakte dat ik me realiseerde dat dit moment voor mij naderde als een toetsmoment: het moment waar je je op voorbereidt door zo goed mogelijk in de stof te zitten en je te verdiepen, het moment waarop je graag wil laten zien dat je weet hoe het allemaal moet. In het besef dat je hoe dan ook beoordeeld zult worden op je prestatie. Dat een ieder die naar dit verhaal luistert daar een oordeel over zal hebben.

Uiteraard wordt dit gevoel vooral ingegeven door de ervaringen die ik zelf heb opgedaan in mijn lange schoolcarrière als leerling en student. Het gevoel dat ik leerde voor een toets en vervolgens hoorde of *ik* de toets had gehaald. Niet of ik de stof al beheerste, of dat ik op de goede weg was qua ontwikkeling, maar dat *ik* de toets wel of niet had gehaald. Dat had uiteraard impact, niet alleen op hoe ik toen als leerling het leren en toetsen ervaarde, maar ook op hoe ik mijn leren en mijn ontwikkeling zelf wilde vormgeven. Het hordesysteem - wat de invulling van de toetsing binnen het onderwijs toch vaak nog is - maakte dat 'beoordeeld worden' snel voelde als iets negatiefs: alsof er een waardeoordeel werd gegeven over hoe jij - als persoon - gepresteerd had. Alsof je daarop werd afgerekend.

Hoe mooi zou het zijn als 'wij', zowel de leerlingen als de docenten, het beoordelen gaan zien als iets positiefs? Het loskoppelen van de beoordeling en de beslissing, ik kom daar straks nog op terug, maakt het naar mijn mening mogelijk het geven en ontvangen van een oordeel vooral te zien en te ervaren als het voeren van een professionele dialoog waarin docent en leerling samen met elkaar kijken naar waar de leerling al staat. En dus ook in alle openheid spreken over wat de leerling nog niet beheerst. Niet met als doel de leerling daarop af te rekenen, maar met als doel de leerling een stap verder te helpen.

Ik nodig jullie dan ook van harte uit om allemaal een oordeel te hebben over deze rede. Laat me weten wat je ervan vindt, waar de belangrijkste vragen liggen die we nog zouden moeten beantwoorden. Alleen als we elkaar oordelen blijven geven en vragen blijven stellen, kunnen we samen volgende stappen zetten en bruikbare, zinvolle kennis blijven ontwikkelen.

Mijn eigen ervaringen met toetsen hebben uiteraard niet alleen impact gehad op mijn gevoel over toetsing, maar ook op hoe ik – tijdens mijn latere werkloopbaan – dacht dat toetsen vormgegeven moest worden. Nu, na 8 jaar associate lector geweest te mogen zijn, is mijn beeld van wat er kan en mogelijk is totaal bijgesteld. Ik heb zelf mogen ervaren dat toetsing ook op een heel andere manier ingericht kan worden, dat het niet alleen gaat over het *vaststellen* van het leren, maar juist ook over het *begeleiden* en *stimuleren* van leren. Ik heb van dichtbij aanwezig mogen zijn bij keuzes die opleidingen maakten, en ook mogen helpen bij het maken van deze keuzes, over de wijze waarop de toetsing ingericht zou gaan worden.

Wat ik uit die 8 jaar vooral meeneem, is dat er geen pasklaar antwoord is op de vraag 'hoe het moet'. De interessante vraag om te stellen is juist 'waarom doe je het zoals je het doet?'. De dialoog over wat je gezamenlijke visie is op onderwijs, op de functie van toetsen daarbinnen, en op de kwaliteit die je daarbij nastreeft, maken dat je gezamenlijk in staat bent het antwoord op het 'hoe' te geven. Pas nadat je het 'waarom' dus eerst met elkaar hebt verkend.

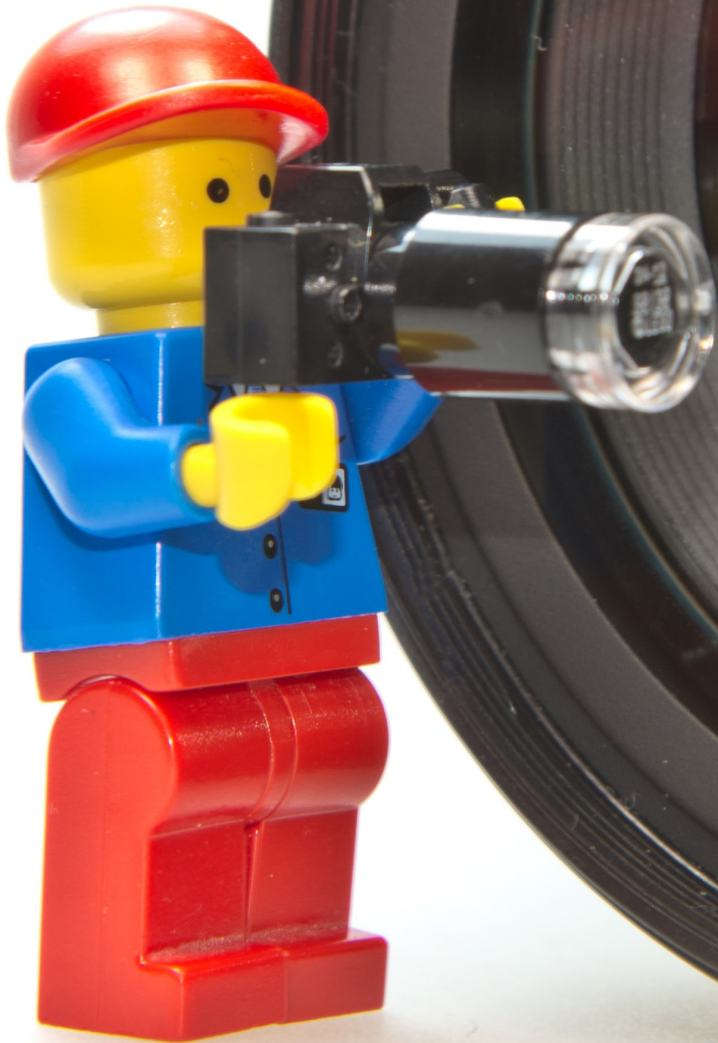
Met deze rede en de invulling van mijn lectoraat hoop ik te bereiken dat docenten, ontwikkelaars van onderwijs en onderwijsmanagers bewuste en geïnformeerde *eigen* keuzes durven te maken die passen bij hun context van onderwijs en toetsing (het 'hoe'), en die zij kunnen beargumenteren (het 'waarom'). Ik hoop daarnaast te bereiken dat docenten zien en ervaren dat het binnen iedere onderwijscontext mogelijk is om het beoordelen en beslissen eigentijds in te richten, zonder dat daarvoor een volledige herziening van het onderwijs nodig is. Zoals we al eerder schreven in een artikel: "*Kwaliteit van toetsing is van essentieel belang voor de kwaliteit van onderwijs (...), maar het is ook essentieel dat opleidingen zich realiseren dat zij zelf verantwoordelijk zijn voor het formuleren van een onderbouwd verhaal: een verhaal dat recht doet aan de eigen visie op opleiden en toetsen en de specifieke opleidingscontext.*" (Van Schilt-Mol et al., 2015, p. 24).

Graag neem ik u in deze rede mee in de mogelijkheden en daag ik u uit om kritisch te durven kijken naar de mate waarin de wijze waarop toetsing is vormgegeven écht aansluit bij jullie visie op onderwijs, én uiteraard naar de mate waarin sprake is van eigentijds beoordelen en beslissen.

Ik zal eerst kort stilstaan bij de noodzaak van het lectoraat: waarom is een lectoraat dat geheel gewijd is aan eigentijds beoordelen en beslissen eigenlijk nodig? Vervolgens zal ik ingaan op wat er al bekend is over de kwaliteit van toetsing. In de afgelopen jaren dat ik werkzaam mocht zijn als associate lector Toetsen en Beoordelen, ontwikkelden wij - onderzoekers van het onderzoeksteam Kwaliteiten van Leraren, collega's van Onderwijs, Onderzoek en Kwaliteit, en collega's Desirée Joosten-ten Brinke en Dominique Sluijsmans - kennis over wat kwaliteit van toetsing is en hoe die geoperationaliseerd kan worden. Dit heeft geresulteerd in het toetsweb. In deze rede presenteer ik u graag de nieuwste versie van het toetsweb, dat als theoretische basis dient voor de onderzoekslijnen die in het lectoraat centraal staan.

Daarna neem ik u graag mee in wat binnen het lectoraat verstaan wordt onder eigentijds beoordelen en beslissen. Ik zal daarbij expliciet stilstaan bij het onderscheid tussen beoordelen en beslissen, en ingaan op de, in mijn optiek, drie belangrijkste functies van toetsing: de leerfunctie, de beslisfunctie en de evaluatiefunctie. Hoewel in de visies van scholen en opleidingen vaak uitspraken worden gedaan die te relateren zijn aan deze drie functies, blijkt er in de praktijk vaak sprake van een discrepantie tussen de visie op papier en de wijze waarop daar in de praktijk invulling aan gegeven wordt. Ik koppel dat aan de verschillende manieren waarop de toetssystemen over het algemeen zijn ingericht, waarbij ik een onderscheid maak naar (1) het toetssysteem waarin de beslisfunctie het uitgangspunt vormt voor het ontwerpen van het onderwijs- en toetsprogramma, (2) het toetssysteem waarin de leerfunctie én de beslisfunctie als twee separate aspecten leidend zijn voor het ontwerp van het onderwijs- en toetsprogramma, en (3) het toetssysteem waarin de leer- en beslisfunctie met elkaar verweven zijn.

Vervolgens benoem ik enkele handvatten om hier, ongeacht het toetssysteem dat gehanteerd wordt, invulling aan te kunnen geven. Het maken van het onderscheid tussen beoordelen en beslissen is hierbij vanzelfsprekend een van de handvatten. Ik sluit deze rede tot slot af door de onderzoeksvragen die de komende jaren centraal staan in het lectoraat aan u te presenteren.



2. WAAROM AANDACHT VOOR EIGENTIJDS TOETSEN EN BEOORDELEN?

Aandacht voor toetsing is uiteraard niet nieuw. Vanzelfsprekend, aangezien toetsen een essentieel onderdeel vormen van het onderwijs. Op basis van toetsresultaten worden belangrijke en veelomvattende beslissingen over leerlingen¹ genomen: selectiebeslissingen, certificeringsbeslissingen en beslissingen over het leren en ontwikkelingspotentieel, allen worden ze gevoed door informatie die via toetsen verkregen wordt. Het is dan ook gerechtvaardigd te zeggen dat de kwaliteit van toetsen boven alle twijfel verheven moet zijn omdat op basis van de resultaten hiervan beslissingen worden genomen over leerlingen (Van Schilt-Mol & Sluijsmans, 2015).

Discussie over het gebruik van toetsen

Hoewel het *belang* van toetsing niet ter discussie staat (Onderwijsraad, 2018), staat de *kwaliteit* van toetsing, en meer specifiek de waarde die aan toetsen wordt gehecht, dat wel. Binnen alle onderwijssectoren is er in toenemende mate sprake van kritiek op het *gebruik* van informatie uit toetsen. Er lijkt sprake van een tendens dat de kwaliteit van onderwijs vooral wordt vastgesteld op basis van slagingspercentages, eindexamenresultaten en doorstroomcijfers. Hierdoor ervaren docenten het gevoel 'afgerekend' te worden op de resultaten die leerlingen behalen op toetsen: het begeleiden van leerlingen in hun ontwikkeling lijkt van minder belang dan het behalen van meetbare resultaten. Deze focus op indicatoren als rendement en resultaten kan leiden tot ongewenste effecten (Kneyber & Evers, 2013): bij docenten kan dit bijvoorbeeld leiden tot het uitsluitend toepassen van "teaching to the test" (Sluijsmans & Kneyber, 2016), bij leerlingen tot een afnemende (intrinsieke) motivatie voor leren.

Daarnaast ervaren docenten in toenemende mate onvrede over de wijze waarop zij informatie uit toetsen (kunnen) gebruiken om het leren van leerlingen te sturen. Zo vinden toetsen bijvoorbeeld vaak pas plaats ná het leren van leerlingen (Simpson-Beck, 2011; Cauley & McMillan, 2009), waardoor informatie uit deze toetsen eigenlijk te laat komt om nog bij te kunnen dragen aan het leerproces van een leerling (Sluijsmans et al., 2013).

1. In deze rede wordt steeds gesproken over leerling, docent en school. Hiermee wordt bedoeld op zowel leerlingen en studenten, op leerkrachten, leraren en docenten, en op scholen binnen alle onderwijssectoren.

Dit kan een negatief effect hebben op de motivatie van leerlingen (Hattie & Timperley, 2007) doordat de toets feitelijk een afsluiting vormt van het leerproces en leerlingen dus eigenlijk niets meer hoeven met de informatie uit deze toets (Sadler, 1989). Om dit gegeven te onderstrepen toonde Malecka (2021) tijdens een presentatie over het belang van feedback een foto van een prullenbak met daarin een toets die zij had voorzien van uitgebreide feedback maar die door de student direct na ontvangst ongelezen was weggegooid: voor de student was met het 'voldoende' afronden van de toets het vak afgerond, ook al had deze student de gegeven feedback goed kunnen gebruiken bij vakken die nog op het programma stonden.

Realiseren van kwalitatief goede toetsing

In de praktijk blijkt dat docenten niet vanzelfsprekend beschikken over de vereiste toetsbekwaamheid om kwalitatief goede toetsing te kunnen ontwerpen, ontwikkelen, uitvoeren en evalueren. In 2018 constateerde zowel de Onderwijsraad als de Commissie Kwaliteit Schoolexaminering dat de toetsbekwaamheid van docenten vaak in onvoldoende mate aanwezig is, en dat niet alleen concrete professionele standaarden specifiek voor toetsing voor docenten (m.u.v. hbo-docenten), leidinggevend en docentenopleiders ontbreken, maar tevens dat binnen docentenopleidingen te weinig aandacht besteed wordt aan toetsdeskundigheid. Ook constateerden zij dat er in alle onderwijssectoren onvoldoende (na)scholingsaanbod is dat is gericht op het vergroten van de toetsbekwaamheid.

De Onderwijsraad (2018) constateerde daarnaast dat de huidige toetsprogramma's in alle onderwijssectoren onevenwichtig zijn. Er is in de programma's zowel onvoldoende aandacht voor het geven van feedback die het leren van leerlingen stuurt als voor het nemen van verantwoorde beslissingen over de leerlingen: het blijkt dat beoogde leerresultaten niet altijd getoetst worden en dat resultaten van toetsen beperkt worden meegenomen in de beslissingen. Bovendien blijkt bij het opstellen van toetsprogramma's dat er vaak onvoldoende is nagedacht over de aansluiting tussen de leerdoelen en de toetsing (constructieve afstemming, ook bekend als 'constructive alignment') en over de relatie tussen het nemen van beslissingen en de ingezette toetsen. Dit heeft uiteraard risico's voor de zorgvuldigheid en betrouwbaarheid van de beslissingen die op basis van toetsen worden genomen. Als leerdoelen en toetsen onvoldoende op elkaar aansluiten, is het lastig om, zeker op basis van een enkele toets, vast te kunnen stellen of de leerdoelen zijn behaald. Als daarnaast ook de toetsen onderling binnen een toetsprogramma niet in evenwicht zijn, is het eveneens onmogelijk om vast te stellen of de beoogde eindniveaus zijn behaald. De toetsen in het programma vormen idealiter immers samen een bewuste en beargumenteerde combinatie van toets(vorm)en, passend bij de doelen en het

onderwijsprogramma (Baartman & Van der Vleuten, 2015), die gezamenlijk de basis vormen waarop beslissingen worden gebaseerd.

Ten aanzien van de leerlingvaardigheden is uit onderzoek weliswaar gebleken dat de mate waarin leerlingen in staat zijn te reflecteren op eigen werk en een actieve rol in het toetsproces hebben van invloed is op een succesvolle implementatie van eigentijdse toetsvormen als formatief evalueren (Schildkamp et al., 2014), er is echter beperkt onderzoek uitgevoerd naar de specifieke (toets)bekwaamheid van leerlingen die nodig is om met informatie uit het programma van beoordelen en beslissen daadwerkelijk invulling te kunnen geven aan het (nemen van regie over het) eigen leerproces.

Naar meer eigentijds beoordelen en beslissen

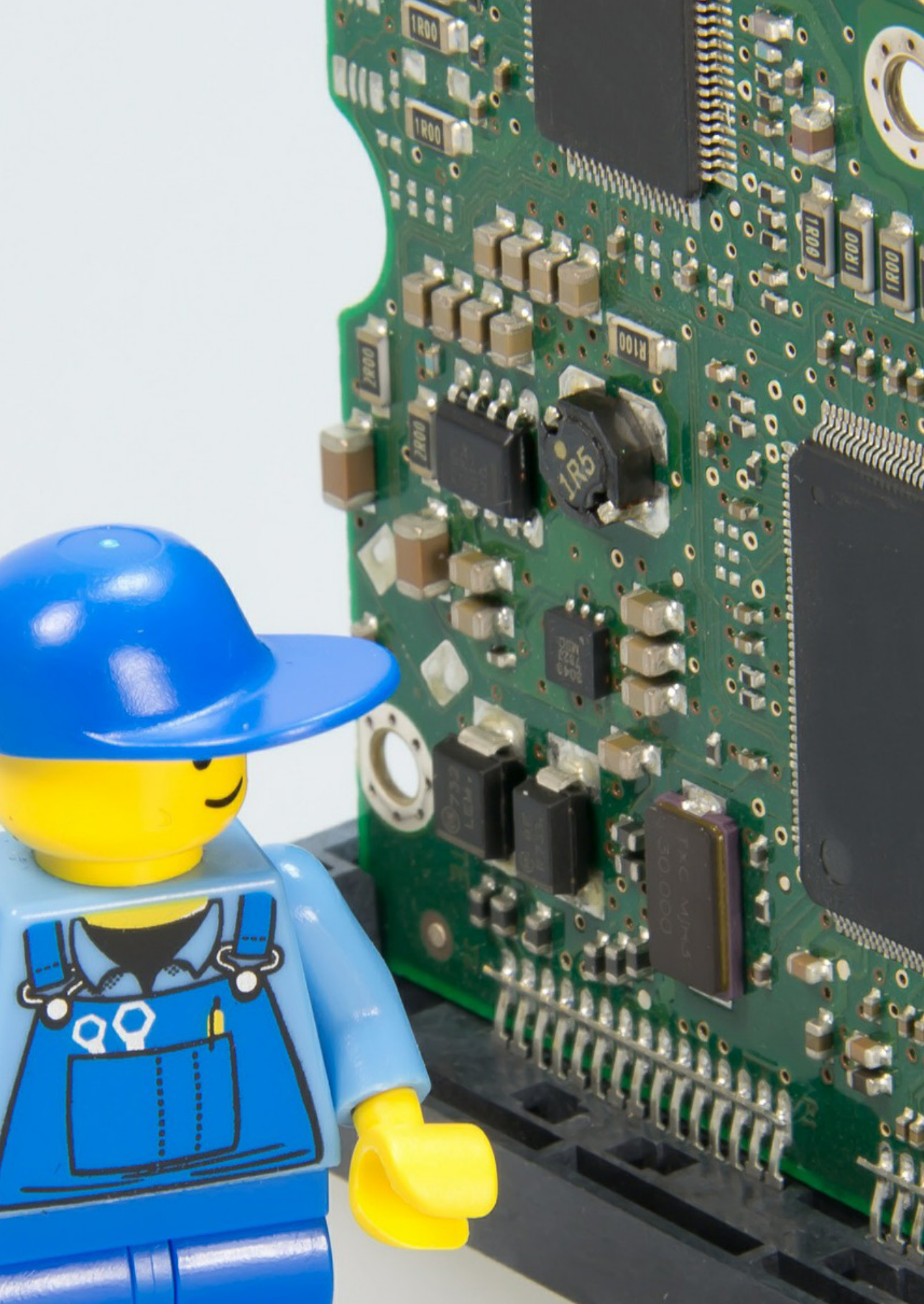
In de dagelijkse onderwijspraktijk hebben deze recente discussies over de kwaliteit, de toetsbekwaamheid en het gebruik van toetsing ertoe geleid dat in alle onderwijssectoren wordt nagedacht over de wijze waarop toetsing vormgegeven zou kunnen worden. Voorbeelden hiervan zijn experimenten met vormen van toetsing als cijferloos lesgeven (po, vo), het afschaffen van alle tussentijdse toetsen (vo, hbo), en de invoering van programatisch toetsen (hbo/wo). Deze voorbeelden hebben gemeen dat hierin heel expliciet gekeken wordt naar de leerfunctie van toetsen: hoe kunnen toetsen niet alleen helpen de voortgang van leerlingen en effectiviteit van lesgeven in kaart te brengen, maar ook als instrument dienen om onderwijs op maat aan te bieden (Schildkamp, e.a., 2014). Uit onderzoek blijkt dat voor een succesvolle implementatie van dergelijke eigentijdse toetsvormen voldaan moet zijn aan vier condities: docenten moeten bezitten over voldoende toetsbekwaamheid van docenten, er moeten effectieve vormen van professionalisering worden georganiseerd, een onderzoekende houding naar de eigen toetspraktijk moet worden gestimuleerd en er moet een leergemeenschap worden gecreëerd (Sluijsmans et al., 2013). Daarnaast moeten toetsen zelf betekenisvol zijn en voorzien in feedback en moeten docenten op basis van toetsresultaten leren en lesgeven kunnen verbeteren en kennis en vaardigheden hebben om toetsgegevens te kunnen verzamelen (Schildkamp et al., 2014). Bovendien moeten leerlingen kunnen reflecteren op eigen werk en een actieve rol in het toetsproces hebben en binnen de school sprake moet zijn van een cultuur waarin verbeteren van onderwijs centraal staat en geïnvesteerd wordt in tijd (samenwerken aan formatieve toetsen) en middelen (professionele ontwikkeling) (Schildkamp et al., 2014).

Zowel de toetsbekwaamheid van docenten én leerlingen als de positieve feedbackcultuur waarin toetsbekwame professionals toetsresultaten benutten om kritisch te reflecteren op hun eigen handelen, het onderwijs te optimaliseren en de ontwikkeling van leerlingen optimaal vorm te geven, blijken met andere woorden essentieel bij de invoering van

eigentijdse vormen van toetsing. Hiermee kan niet alleen invulling worden gegeven aan het personaliseren van leren (d.w.z. dat leerlingen in staat worden gesteld regie te nemen over hun leren/leerproces en dat onderwijs is afgestemd op de individuele onderwijsbehoeften van leerlingen), tevens draagt deze positieve toetscultuur er aan bij dat docenten regie (durven) nemen over de kwaliteit van hun onderwijs, en daarmee over de kwaliteit van de toetsing.

De conclusie dat zowel de toetsbekwaamheid van docenten als de kwaliteit van de toetsprogramma's te vaak te wensen overlaat (Onderwijsraad, 2018; Commissie Kwaliteit Schoolexaminering, 2018) en het feit dat nog onvoldoende duidelijk is welke toetsbekwaamheid leerlingen nodig hebben om regie te kunnen nemen over het eigen leerproces, zijn echter belangrijke aandachtspunten. Juist daar waar geëxperimenteerd wordt met nieuwe(re) vormen van toetsing zijn aspecten als een duidelijke visie op de functies van toetsen, een hoge toetsbekwaamheid en een kwalitatief goed toetsprogramma noodzakelijk om te garanderen dat, ondanks experimenten, wel juiste en zorgvuldige beslissingen over (het leren van) leerlingen worden genomen. De focus van het lectoraat Eigentijds beoordelen en beslissen zal dan ook hierop liggen: wat vraagt het van een school, van docenten en van leerlingen om meer eigentijdse vormen van toetsing op een kwalitatief hoogstaande manier te ontwikkelen en implementeren?

2. Waarom aandacht voor eigentijds toetsen en beoordelen

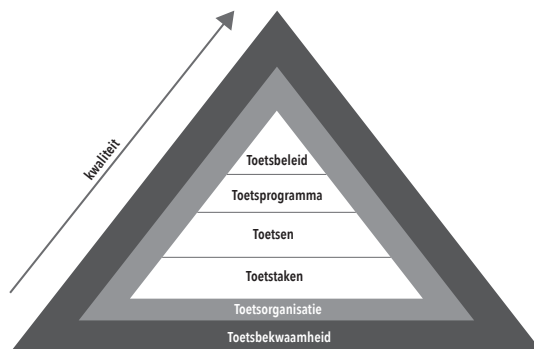


3. KWALITEIT VAN TOETSING

Voordat dieper ingegaan wordt op wat verstaan wordt onder eigentijds beoordelen en beslissen, wordt eerst stilgestaan bij *kwaliteit van toetsing*. Want hoewel het, gezien het belang van toetsen bij het nemen van beslissingen over leerlingen, voor de hand liggend lijkt te streven naar ‘kwalitatief goede toetsing’ bleek de afgelopen jaren dat er nog veel onduidelijk bestond over dit concept; wat verstaan we onder toetsing, wanneer is er sprake van voldoende kwaliteit, en hoe kunnen we zelf regie nemen over deze kwaliteit?

KWALITEIT VAN TOETSING GEOPERATIONALISEERD

In 2015 werden de eerste contouren gepresenteerd voor een methodiek om kwaliteit van toetsing niet alleen te operationaliseren, maar ook te analyseren en te borgen (Sluijsmans et al., 2015, Van Schilt-Mol et al., 2015): Toetsing Getoetst. Dit leidde in 2016 tot de eerste versie van het toetsweb, dat de kern vormt van deze methodiek (Van Schilt-Mol et al., 2015). Hoewel dit web in eerste instantie ontwikkeld is voor het hoger beroepsonderwijs, is inmiddels op basis van ervaringen van gebruikers gebleken dat deze methodiek, soms met enige aanpassing van terminologie, ook goed bruikbaar is binnen de andere onderwijssectoren. Het web kan worden gezien als de doorontwikkeling van de toetspiramide (Joosten-ten Brinke, 2011; Joosten-ten Brinke & Sluijsmans, 2012; Figuur 1) naar een kwaliteitszorginstrument.



Figuur 1. Piramide van eigentijds toetsen en beoordelen (Joosten-ten Brinke & Sluijsmans, 2012)

Bij het ontwerp van de piramide is uitgegaan van het *docent*perspectief: welke entiteiten spelen een rol voor de docent om kwalitatief goede invulling te geven aan toetsing? De kernboodschap van de piramide is dat de kwaliteit van toetsing wordt bepaald door een goed samenspel tussen de zes piramidelagen (Sluijsmans, et al., 2015). De kwaliteit van toetsing is met andere woorden zo sterk als de zwakste schakel.

Bij de doorontwikkeling naar een kwaliteitszorginstrument om de kwaliteit van toetsing te kunnen analyseren en borgen is ingestoken vanuit het *opleidings*perspectief: hoe kan de opleiding duurzaam werken aan het realiseren van een kwalitatief goede invulling aan toetsing? En vooral: hoe kan een opleiding zelf uitspraken doen over de kwaliteit van toetsing en regie nemen over de duurzame ontwikkeling van deze kwaliteit?

Om de ontwikkeling inzichtelijk te maken is inspiratie ontleend aan het Auditing Instrument for Sustainability in Higher Education (AISHE) (Brouwer & Boer, 2013; DHO, 2010). AISHE is gericht op het duurzaam werken aan ontwikkeling en het stimuleren van lange termijn denken: hoe komen we van de huidige situatie naar de gewenste situatie (Van Schilt-Mol, Peeters, Sluijsmans & Jakobs, 2015)? Binnen AISHE worden vijf ontwikkelingsfasen onderscheiden: 1) de activiteiten-georiënteerde fase, 2) de proces-georiënteerde fase, 3) de systeem-georiënteerde fase, 4) de keten-georiënteerde fase en 5) de maatschappij-georiënteerde fase. Kenmerkende aspecten binnen deze fasen zijn een ontwikkeling in de wijze waarop *actoren* ingezet worden (van individuele acties naar acties waarbij alle relevante actoren betrokken zijn), een ontwikkeling in het *langetermijndenken* (van ad hoc naar een langetermijnstrategie) en een ontwikkeling in de *integratie van processen en activiteiten* in het onderwijsproces (van losstaande activiteiten naar activiteiten die bewust in samenhang worden uitgevoerd binnen lopende processen). Deze ontwikkelingsfasen zijn vertaald naar de context van toetsing (Figuur 2) zodat deze gebruikt kunnen worden bij het bepalen van de huidige fase van ontwikkeling en het vaststellen van de gewenste fase van ontwikkeling.

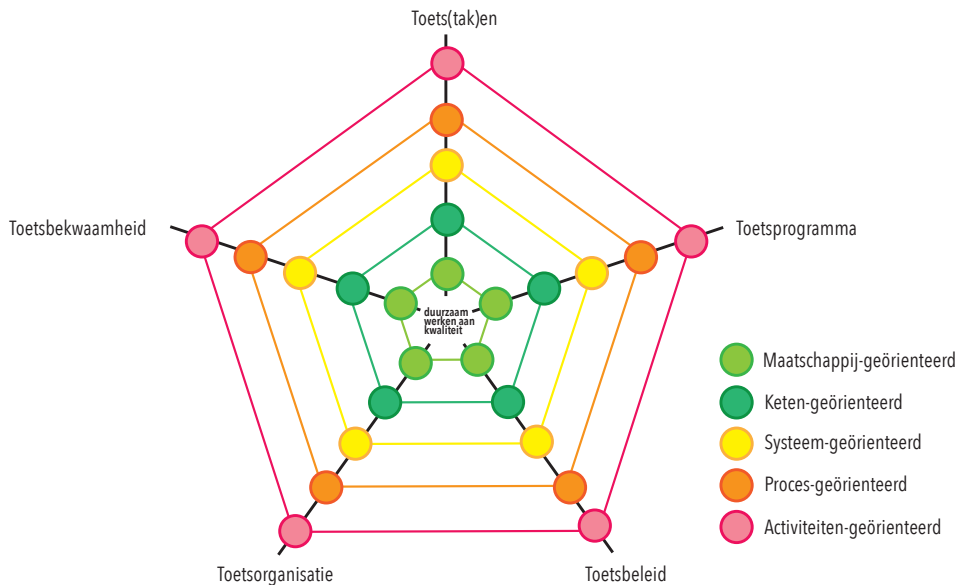
Activiteiten-georiënteerd	Individuele actoren in de opleiding voeren op ad-hoc basis losstaande toetsactiviteiten uit.
Proces-georiënteerd	Relevante actoren in de opleiding voeren vanuit een korte termijnbeleid toetsactiviteiten uit die onderling samenhangen.

Systeem-georiënteerd	Alle relevante actoren in de opleiding voeren vanuit een middellangetermijnbeleid toetsactiviteiten uit die onderling samenhangen en geïntegreerd zijn in lopende processen in de opleiding.
Keten-georiënteerd	Alle relevante actoren in de opleiding voeren vanuit een middellangetermijnbeleid, samen met ketenpartners, toetsactiviteiten uit die onderling samenhangen en geïntegreerd zijn in lopende processen in de opleiding en de keten.
Maatschappij-georiënteerd	Alle relevante actoren in de opleiding voeren vanuit een langetermijnbeleid, samen met ketenpartners en de maatschappij, toetsactiviteiten uit die onderling samenhangen en geïntegreerd zijn in lopende processen in de opleiding en de keten/maatschappij.

Figuur 2. Fasen van ontwikkeling in de weg naar duurzaam werken aan toetsing (gebaseerd op Brouwer & Boer, 2013, beschrijvingen overgenomen uit Van Schilt-Mol et al., 2015)

ONTWIKKELING VAN HET TOETSWEB

Op basis van onderzoek naar de ervaringen naar de methodiek Toetsing Getoetst (Peters, et al., 2017) is besloten bij de vertaling van de piramide naar een kwaliteitszorginstrument de toetsentiteiten toetsen en toetstaken samen te voegen: het onderscheid tussen deze entiteiten blijkt in de dagelijkse onderwijspraktijk niet expliciet te worden gemaakt (Van Schilt-Mol et al., 2016), bovendien gelden voor deze entiteiten dezelfde kwaliteitscriteria (met uitzondering van het kwaliteitscriterium 'discriminerend vermogen' dat alleen op de entiteit toetstaken betrekking heeft). Omdat daarnaast uit het onderzoek bleek dat de vorm van piramide er ten onrechte toe leidde dat de gebruikers van de piramide veronderstelden dat er sprake was van een basis (toetstaken) en een top (toetsbeleid), en omdat zij ten onrechte veronderstelden dat toetsorganisatie en toetsbekwaamheid voorwaardelijk zouden zijn voor de overige lagen in de piramide (in de visuele weergave lijken organisatie en bekwaamheid de andere entiteiten te omvatten), er is voor gekozen om de vorm van een web te kiezen (Figuur 3). Dit maakte het tevens mogelijk om ook de ontwikkelingsfasen zoals hierboven beschreven hierin te verwerken.



Figuur 3. Eerste versie van het toetsweb (Van Schilt-Mol, et al., 2016)

Het web symboliseert de verwevenheid van de verschillende toetsentiteiten, en maakt visueel zichtbaar dat deze allemaal met elkaar verbonden zijn: is de kwaliteit op een bepaalde entiteit niet op orde, dan heeft dat direct implicaties voor de kwaliteit van de andere entiteiten. Als bijvoorbeeld het toetsbeleid niet transparant is, zal dit zijn uitwerking hebben op de kwaliteit van toetsprogramma's, toetsen en de onderliggende toetstaken (Van Schilt-Mol et al., 2020). Zowel de toetsentiteiten als de ontwikkelingsfasen beïnvloeden de stabiliteit van de kern van het toetsweb (Van Schilt-Mol et al., 2020; Van Schilt-Mol & Sluijsmans, 2018), het duurzaam werken aan de kwaliteit van toetsing. Door het ontwikkelen op de ontwikkelingsfasen komt een opleiding steeds dichterbij deze kern van het web.

Op basis van ervaringen bij de begeleiding van de methodiek bij vele opleidingen in verschillende opleidingscontexten binnen het beroepsonderwijs is vastgesteld dat het onderscheid tussen de keten- en maatschappijgeoriënteerde fase voor de specifieke context van toetsing niet zo scherp gesteld kan worden als binnen AISHE is gedaan. Tijdens de uitvoering werd duidelijk dat het verschil tussen de keten en de maatschappij op het kenmerkende aspect 'actoren' vaak niet onderscheidend genoeg was. Voor de meeste mogelijke maatschappelijke partners (zoals collega-opleidingen, ouders, afnemend scholenveld) geldt dat deze binnen het (beroeps)onderwijs ook gezien kunnen worden (of zouden moeten worden!) als ketenpartners. Omdat voor de andere kenmerkende

aspecten per ontwikkelingsfase (termijn, processen en activiteiten) geldt dat deze gelijk zijn in de keten- en maatschappij georiënteerde fase (zie Figuur 2) en het onderscheid tussen de actoren (wel/niet keten/maatschappij) in deze maatschappij-georiënteerde fase dus niet onderscheidend genoeg was, is er in 2018 voor gekozen om de keten- en maatschappij-georiënteerde fase samen te voegen. Dat heeft geleid tot de versie van het web zoals deze ook is opgenomen in de 'Kwaliteiten van de toetsbekwame hbo-docent' (Van Schilt-Mol, et al., 2020).

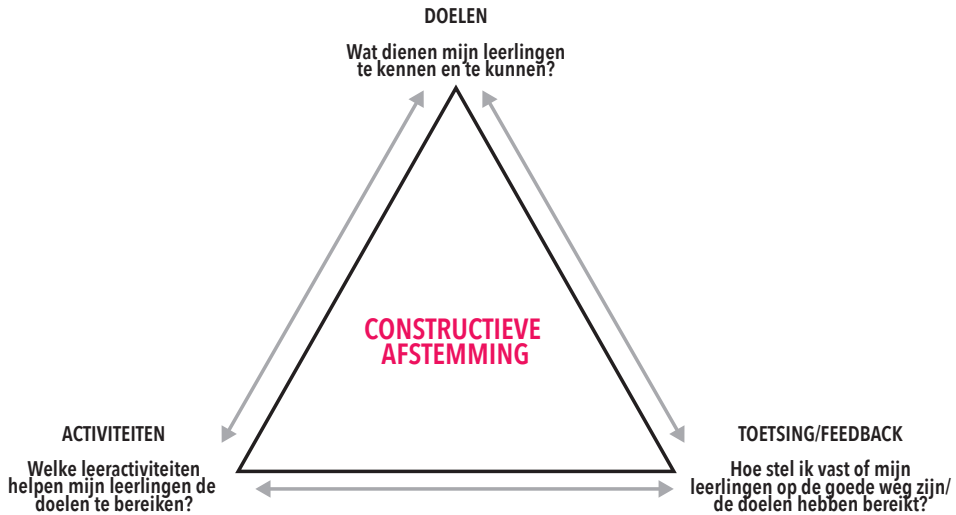
Doorontwikkeling van het toetsweb: toetsweb 2.0

In 2021 is binnen het lectoraat een expertsessie georganiseerd waarin ervaringen met het toetsweb en Toetsing Getoetst zijn geanalyseerd en de consequenties hiervan voor het toetsweb zijn uitgewerkt. Dit heeft geleid tot enkele aanpassingen.

Visie op onderwijs

Op de eerste plaats is de tekst in de kern van het web ('duurzaam werken aan kwaliteit') vervangen door 'visie op onderwijs'. Daardoor is de kern in plaats van een omschrijving van het doel van de methodiek an sich, veranderd in een toegevoegd inhoudelijk aspect van het toetsweb. Steeds meer wordt duidelijk dat onderwijs en toetsing niet kunnen worden gezien als twee aparte aspecten. Door incidenten die in het hoger onderwijs hebben plaatsgevonden, - die er toen toe hebben geleid dat de maatschappelijke waarde van diploma's ter discussie werd gesteld -, is er de afgelopen jaren echter veel expliciete aandacht geweest voor specifiek toetsbekwaamheid en afstuderen. Hierdoor ontstond de tendens een aparte focus te leggen op toetsing. Gelukkig lijkt het tij te keren: steeds meer wordt benadrukt dat toetsbekwaamheid onderdeel is van docentbekwaamheid (Van Schilt-Mol, et al., 2020) en dat onderwijs en toetsing juist onlosmakelijk met elkaar verbonden zijn.

Dit inzicht is uiteraard niet nieuw: al in 1996 introduceerde John Biggs de term *constructive alignment* (Biggs, 1996). Hiermee wordt bedoeld op de samenhang die er moet zijn tussen de doelen die met het onderwijs worden nagestreefd, de leeractiviteiten die worden uitgevoerd om de leerlingen de kans te geven zich te ontwikkelen op deze doelen, en de toetsactiviteiten die worden uitgevoerd om vast te stellen of de leerlingen de doelen hebben behaald. Door een goede samenhang - constructieve afstemming - tussen deze drie elementen te realiseren is er sprake van een samenhangend curriculum (Sluijsmans, 2020).



In de vorige versie van het web veronderstelden we impliciet dat in een toetsbeleid ook aandacht zou moeten zijn voor de visie op toetsing. Echter, gezien de onlosmakelijke verbondenheid van leerdoelen, leeractiviteiten en toetsactiviteiten (zoals weergegeven in Figuur 4), gaat het bij het realiseren van kwalitatief goede toetsing die leidt tot betrouwbare en zorgvuldige beslissingen over leerlingen, dus niet *alleen* om een visie op toetsing. Het gaat om een visie op *onderwijs*, die inzichtelijk maakt waar je als school voor staat, waar je voor gaat, en wat je kernkwaliteiten en speerpunten zijn (SPV, 2020). De visie geeft met andere woorden antwoord op de vraag wat jullie verstaan onder ‘goed onderwijs’. De visie op onderwijs wordt gevormd door de legitimering van het onderwijs/de opleiding, een visie op het doel van het onderwijs/het beroep waartoe wordt opgeleid, een visie op de opdracht van het onderwijs, een visie op leren en een visie op toetsen. Daarmee wordt in een visie op onderwijs respectievelijk antwoord gegeven op ten minste de volgende vragen:

- wat dragen wij, vanuit onze professionele identiteit, bij aan maatschappelijke ontwikkelingen,
- waartoe leiden wij op,
- hoe geven wij invulling aan zowel de vormende (socialisatie en persoonsvorming) als kwalificerende doelstelling van het onderwijs,
- hoe willen wij dat onze leerlingen leren, en
- hoe stellen wij onze leerlingen in staat te laten zien wat ze hebben geleerd?

De manier waarop de toetsing wordt ingericht wordt daarmee een middel om de visie op onderwijs te realiseren, in plaats van een doel op zich (Baartman, Van Schilt-Mol & Van

der Vleuten, 2020). Een visie op toetsen die samenhangt met een visie op onderwijs zal er daarnaast voor zorgen dat toetsing onderdeel wordt van het onderwijsconcept en zal daardoor versterkend werken op het realiseren van de onderwijsdoelen die de opleiding/school voor ogen heeft (VO-Raad, 2019). De visie op toetsing vormt daarmee het fundament voor het borgen van de kwaliteit van de toetsing. Dit veronderstelt uiteraard dat het hebben van een visie op onderwijs voorwaardelijk is om op een zinvolle wijze invulling te kunnen geven aan toetsing.

Gewenste fase van ontwikkeling

In eerdere publicaties daagden we opleidingen uit te streven naar het bereiken van ten minste de systeem-georiënteerde fase: in deze fase voeren alle relevante actoren in de opleiding vanuit een middellangetermijnbeleid toetsactiviteiten uit die onderling samenhangen en geïntegreerd zijn in lopende processen in de opleiding. De kans dat de kwaliteit van bijvoorbeeld toetsen of het toetsprogramma op toeval berust of op de kwaliteiten van die ene professionele docent die er individueel aan de slag gaat is hiermee klein (Van Schilt-Mol et al., 2016). Echter: gezien het maatschappelijk belang dat gehecht wordt aan betrouwbare en zorgvuldige beslissingen over leerlingen, en gezien het belang van een onderwijssysteem waarin leerlingen voorbereid worden op ofwel vervolgonderwijs, ofwel een plaats op de arbeidsmarkt, zouden onderwijssectoren/opleidingen onderling, en opleidingen en het beroepenveld in het (hoger) beroepsonderwijs zich gezamenlijk verantwoordelijk moeten voelen voor het vaststellen en bereiken van de leerdoelen die we aan leerlingen stellen: zij maken met elkaar onderdeel uit van de onderwijsketen. Vanuit dat perspectief is het niet meer dan vanzelfsprekend dat opleidingen streven naar het bereiken van de keten-georiënteerde fase.

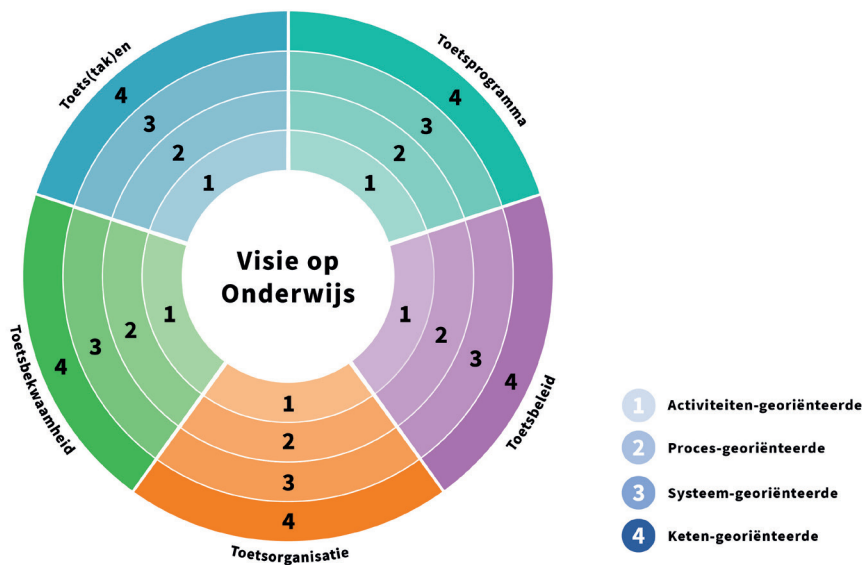
Vormgeving

Uit de ervaringen met het werken met de oude versie van het web bleek visualisatie van het web enkele ongewenste effecten te sorteren. Op de eerste plaats leidde het 'stoplicht-kleurgebruik' tot verkeerde interpretaties. Het web is uitdrukkelijk bedoeld als zelfevaluatie-instrument waarmee de ontwikkeling in kaart gebracht kan worden en op basis waarvan ambities en vervolgstappen kunnen worden geformuleerd. Door het kleurgebruik, waarbij de activiteiten-georiënteerde fase rood is gekleurd en de ketengeoriënteerde fase groen is gekleurd, wordt de indruk gewekt dat de activiteitengeoriënteerde fase 'onvoldoende' is en de ketengeoriënteerde fase 'goed'. Dit doet geen recht aan het ontwikkelingsgerichte karakter van het web en eventuele ontwikkelingen die opleidingen ingezet hebben.

Op de tweede plaats bleek de wijze waarop het web werd ingevuld voor verwarring te zorgen. In deze versie symboliseerde de buitenste ring de 'laagste' fase van ontwikkeling

aan, hoe meer een opleiding zich ontwikkelde hoe dichter de opleiding bij de kern van het web kwam. Deze wijze van werken met een web is afwijkend van hoe dergelijke ontwikkelingswebben, zoals bijvoorbeeld het curriculaire spinnenweb (Van den Akker, 2003), gangbaar worden gebruikt. Daarbij is de ontwikkeling zichtbaar door de groei vanuit de kern naar de randen van het web: hoe groter de oppervlakte is die wordt gedekt, hoe groter de mate van ontwikkeling. Door de kern 'duurzaam werken aan kwaliteit' te vervangen door het voorwaardelijke aspect 'visie op onderwijs' lijkt het logisch om te starten vanuit deze kern: deze is immers voorwaardelijk om vervolgens de huidige fase van ontwikkeling in te kunnen schatten en om ambities te formuleren ten aanzien van de gewenste ontwikkeling. Door de 'hoogste' fase van ontwikkeling (keten-georiënteerd) aan te duiden als de streeffase als buitenste ring op te nemen, wordt bovendien de groei in de ontwikkeling ook visueel zichtbaar. Tot slot is de vormgeving van het web zelf aangepast: bij het invullen bleek onduidelijk of de bolletjes in het web, of juist de ruimtes tussen de lijnen leidend waren.

Alle aanpassingen maken dat het nieuwe toetsweb er als volgt uit ziet (Figuur 5).



Figuur 5. Toetsweb²

2 Na het uitspreken van de rede zijn de laatste wijzigingen doorgevoerd door Bureau 302 van de Hogeschool van Arnhem en Nijmegen. Hier werken studenten als onderdeel van hun opleiding aan opdrachten van binnen en buiten de HAN. De versie die hier is opgenomen is begin 2022 opgeleverd.

In deze nieuwe versie van het web is de buitenste ring de keten-georiënteerde fase, de binnenste ring de activiteiten-georiënteerde fase. Vergelijkbaar met de vorige versie van het web, zijn de ontwikkelingsfasen als volgt te karakteriseren:

KENMERKENDE ASPECTEN FASE	ACTOREN	TERMIJN	PROCESSEN	ACTIVITEITEN
Activiteit-georiënteerd	Individuele actoren in opleiding	Ad hoc	Onderdelen van processen	Los van elkaar uitgevoerd
Proces-georiënteerd	Relevante actoren in opleiding	Korte termijn	Op zich staande processen	In samenhang uitgevoerd
Systeem-georiënteerd	Alle relevante actoren in opleiding	Middellange termijn	Toetsprocessen geïntegreerd in processen binnen de opleiding	Systematisch en structureel in samenhang uitgevoerd
Keten-georiënteerd	Alle relevante actoren in keten	Lange termijn	Toetsprocessen geïntegreerd in processen binnen de keten	Systematisch en structureel in samenhang uitgevoerd

Een opleiding kan door zich verder te ontwikkelen als het ware omhoog klimmen in het web: hoe verder de opleiding 'klimt', hoe groter de oppervlakte die wordt gevuld. Voor ieder vlakje in het toetsweb is een beknopte omschrijving beschikbaar waarin geschetst wordt op welke wijze de kwaliteitsprocessen ingericht zijn voor die betreffende toetsentiteit³. De kern blijft de verbindende schakel in het web: op de eerste plaats is het hebben van een visie op onderwijs voor iedere ontwikkelingsfase noodzakelijk. Op de tweede plaats beïnvloeden zowel de toetsentiteiten als de ontwikkelingsfasen de stabiliteit van de kern.

3 De uitgebreide beschrijvingen van zowel de entiteiten als de wijze waarop deze in de verschillende ontwikkelingsfasen vormgegeven zijn, zijn te vinden op: <https://www.han.nl/onderzoek/samenwerken/kwaliteit-van-leren/producten/>. Hier is ook een overzicht opgenomen van mogelijke kwaliteitscriteria die gesteld kunnen worden aan de verschillende entiteiten.

Het web symboliseert daarmee de kwetsbaarheid van het geheel: te veel of te weinig aandacht voor één entiteit of te weinig aandacht voor de onderlinge verbondenheid van de entiteiten kan de kern van het web verstoren.

Het web kan worden gebruikt om met elkaar het gesprek aan te gaan over de huidige fase van ontwikkeling op de toetsentiteiten, en om gezamenlijk ambities te bepalen voor de gewenste fase van ontwikkeling. Door te werken met het web, en meer specifiek met de ontwikkelingsfase per toetsentiteit, wordt, zoals eerder benoemd, voorkomen dat de kwaliteit van toetsing op toeval berust: de ontwikkelingsfasen bieden handvatten om het werken aan kwaliteit van toetsing duurzaam in te richten, waarbij gezamenlijke afspraken over wat onder kwaliteit wordt verstaan, de kwaliteitscriteria, leidend zijn voor het daadwerkelijke ontwerpen en borgen volgens de gezamenlijk afgesproken samenhangende processen. Een opleiding is daarbij altijd *zelf* verantwoordelijk voor het vaststellen van de kwaliteitscriteria die op opleidingsniveau worden gehanteerd en voor het verantwoorden van deze keuze (Staf en Services OOK & Lectoraat Eigentijds beoordelen en beslissen, 2021). Deze moeten immers aansluiten bij de visie van de opleiding op onderwijs (en dus ook op toetsen), en bij het beroep waartoe wordt opgeleid (en de eisen die dit vanuit constructieve afstemming stelt aan de toetsvormen die daarbij worden gebruikt).



4. EIGENTIJDS BEOORDELEN EN BESLISSEN

Weten hoe voor afzonderlijke toetsen, een toetsbeleid of een toetsprogramma duurzaam gewerkt kan worden aan kwaliteit is één, het hier op een eigentijdse wijze invulling aan kunnen geven is een volgende stap. Om deze stap te kunnen zetten is inzicht in wat verstaan wordt onder 'eigentijds beoordelen en beslissen' noodzakelijk. Om dit te kunnen duiden wordt eerst ingegaan op de definitie die binnen het lectoraat geformuleerd is. Vervolgens worden twee belangrijke kernbegrippen uit deze definitie, beoordelen en beslissen, nader toegelicht en wordt beargumenteerd waarom het zinvol is hier een onderscheid in te maken.

EIGENTIJDS BEOORDELEN EN BESLISSEN: EEN DEFINITIE

Onder eigentijds beoordelen en beslissen wordt het volgende verstaan:

“Het bewust en weloverwogen inzetten van toetsen als integraal onderdeel van het onderwijs, zodat op basis van informatierijke beoordelingen zorgvuldige beslissingen genomen worden over (het leren van) leerlingen en de kwaliteit van het onderwijs wordt geoptimaliseerd.”

In de definitie wordt allereerst gesproken over het op *bewust inzetten van toetsen*. In navolging op de aanpassingen in het toetsweb 2.0 staat uiteraard ook binnen het eigentijds beoordelen en beslissen de visie op onderwijs centraal. Zoals eerder beschreven bevat deze visie niet alleen een visie op waartoe wordt opgeleid, maar ook op de wijze waarop het leren hierbij wordt vormgegeven: hoe willen wij dat onze leerlingen leren en

hoe stellen we hen in staat regie te nemen over hun eigen ontwikkeling en hen te laten zien wat ze geleerd hebben? De antwoorden op deze vragen vormen de basis voor het onderwijsontwerp, toetsing maakt daar op deze manier als vanzelf *integraal onderdeel* van uit. Door vanuit de visie op onderwijs eerst vast te stellen waartoe wordt opgeleid ('wat zijn de doelen'), hoe dat wordt gedaan ('welke leeractiviteiten zetten we in om leerlingen de kans te geven zich te ontwikkelen op deze doelen') en hoe wordt geborgd dat de uitgangspunten worden gerealiseerd ('welke toetsactiviteiten zetten we in om vast te stellen of de leerlingen de doelen hebben behaald') ontstaat constructieve afstemming.

In de definitie wordt niet alleen gesproken over het bewust inzetten van toetsen, maar ook over het *weloverwogen inzetten* ervan. Hier doelen we op het maken van ontwerpkeuzes die zijn gebaseerd op actuele kennis en inzichten en die zijn vertaald naar de eigen onderwijscontext. Het gaat, met andere woorden, om het evidence informed handelen: het benutten van zowel praktijkkennis als kennis uit onderzoek om handelen in de praktijk te verbeteren of te verrijken, waarbij kritisch wordt gekeken welke onderdelen uit onderzoek of praktijk ingezet kunnen worden passend bij de praktijkcontext (Wubbels en Van Tartwijk, 2017).

Daarnaast wordt in de definitie gesproken over het *optimaliseren van de kwaliteit van het onderwijs*. Dit vereist dat docenten - op basis van de informatie uit toetsen - kunnen en willen reflecteren op het eigen handelen. Dit vereist uiteraard dat docenten niet alleen (mede) verantwoordelijk zijn het realiseren van constructieve afstemming en voor het ontwerpen en uitvoeren van toetsing, maar ook in staat moeten zijn toetsen te verzamelen, analyseren en hier betekenis aan te geven, maar ook het onderwijs moeten kunnen verbeteren op basis van de analyses van de toetsresultaten (Meijer et al., 2020, Schildkamp et al., 2014).

ONDERSCHIED TUSSEN BEOORDELEN EN BESLISSEN

In de definitie van eigentijds beoordelen en beslissen wordt een expliciet onderscheid gemaakt tussen 'informatierijk beoordelen' en het 'nemen van zorgvuldige beslissingen'. Het onderscheid tussen beoordelen en beslissen kan het beste als volgt worden gedeut:

Bij beoordelen wordt vastgesteld waar de leerling zich bevindt ten opzichte van zijn leerdoel/de beoogde leerresultaten, op basis daarvan wordt gerichte feedback en feedforward gegeven (waar staat de leerling nu en hoe kan hij het beoogde resultaat behalen). Bij beslissen wordt een consequentie verbonden aan één of meerdere beoordelingen.

Beoordelen en beslissen zijn daarmee twee activiteiten die elkaar opvolgen. Bij het beoordelen wordt – door de leerling, door de docent, of door hen samen – gekeken naar een prestatie van een leerling. Beoordelingen kunnen dus gegeven worden op basis van leer-/toetsactiviteiten als kennistoetsen, stage, beroepsproducten, peerfeedback, selfassessment, mentorgesprek, etc.

De prestatie wordt gerelateerd aan het beoogde leerdoel. Op basis hiervan wordt een informatierijke beoordeling gegeven die handvatten bevat voor de ontwikkeling en het leerproces. Er wordt, met andere woorden, op basis van een beoordeling gerichte feedback en feedforward gegeven. Bij het beslissen wordt een consequentie verbonden aan een of meerdere beoordelingen. Deze beslissingen kunnen gericht zijn op verschillende niveaus; op de (verdere) ontwikkeling van de leerling, op de voortgang ten aanzien van de leerdoelen en op het vaststellen van wat geleerd is (certificeren, kwalificeren, selecteren, etc.). In onderstaande wordt eerst dieper ingegaan op het informatierijk beoordelen, vervolgens worden de verschillende niveaus van beslissen verder uitgewerkt.

INFORMATIERIJK BEOORDELEN

Het vaststellen van waar iemand staat ten opzichte van de leerdoelen om daar vervolgens ook gerichte ontwikkeling in mogelijk te maken, vraagt uiteraard dat er sprake is van ‘informatierijke’ beoordelingen. Het lijkt heel vanzelfsprekend dat een beoordeling pas een waarde krijgt als de leerling – al dan niet met begeleiding van de docent – op basis van die beoordeling richting kan geven aan zijn verdere leerproces. Toch blijkt dat in de praktijk niet zo vanzelfsprekend. Over het algemeen wordt een beoordeling namelijk nog (te) vaak uitgedrukt in uitsluitend een cijfer. Hoewel cijfers leerlingen kunnen stimuleren als zij het idee hebben dat een hoog cijfer haalbaar is als zij er hun best voor doen, kan het geven van cijfers, vooral bij leerlingen met vermijdend gedrag, ook averechts werken doordat ze de leerling angstig of onzeker maken (Kirschner et al., 2018). Daarnaast blijkt dat beoordelingen die uitsluitend gericht zijn op cijfers kunnen leiden tot wat aangeduid wordt met ‘aangeleerde afhankelijkheid’ (Sluijsmans, 2013; Yorke, 2003): leerlingen leren

dat de docent bepaalt wat van hen wordt verwacht en dat zij dat niet zelf hoeven te doen. Bovendien geeft uitsluitend een cijfer de leerling geen informatie of concrete handvatten over het 'hoe nu verder'.

Een treffend citaat dat direct de kern van dit laatste probleem samenvat, is het volgende: *"Stel je geeft me voor dit schrijfsel een cijfer. Zeg een 7. Wat moet ik nu de volgende keer anders doen? Ja, een 8 halen! Maar hoe? Het cijfer is betekenisloos. Ik denk dat cijfers prima werken in gestandaardiseerde beoordelingen. (...) ... voor het beoordelen van complexe vaardigheden - bijvoorbeeld gedragsvaardigheden - zijn scores en cijfers betekenisloos. Gelukkig beschikken we over een prachtig middel: taal! Dus laten we woorden gebruiken om complexe vaardigheden te beoordelen."* (Van der Vleuten, 2017, p. 32).

Wat dit citaat duidelijk maakt is dat een cijfer of de indicatie (on)voldoende alléén niet de informatie geeft waarmee de leerling richting kan geven aan zijn verdere ontwikkeling. In haar lectorale rede geeft ook Sluijsmans (2013) hiervan een treffend voorbeeld: *"Vergelijk het met het krijgen van de uitslag van een bloedonderzoek. Als blijkt dat je cholesterol hoger is dan 8,0 mmol/l, weet je dat deze sterk verhoogd is maar nog niet a) waardoor dat komt en b) hoe je deze waarde kunt verlagen."* (Sluijsmans, 2013, p. 23-24).

Om een beoordeling waardevol en zinvol te maken, is het met andere woorden noodzakelijk dat het gaat om een informatierijke beoordeling. Dat is overigens niet per definitie een pleidooi om cijfers volledig af te schaffen, maar wel om ze altijd te voorzien van een narratieve onderbouwing die de leerling niet alleen informatie geeft over de geleverde prestatie in relatie tot de beoogde doelen, maar ook handvatten biedt om toe te werken naar deze beoogde doelen.

ZORGVULDIGE BESLISSINGEN

Voortbordurend op het geven van informatierijke beoordelingen is het ook bij het nemen van beslissingen over de ontwikkeling en voortgang van leerlingen essentieel om daarbij weloverwogen inhoudelijke afwegingen te maken. Opmerkelijk is dat bij het inzetten van toetsen in het onderwijs, bij het beoordelen hiervan en bij het verbinden van consequenties hieraan, de nadruk echter vaak uitsluitend ligt op het afleggen van verantwoording over de voortgang en niet op het kunnen doen van uitspraken of het nemen van beslissingen over de ontwikkeling van leerlingen. Als een willekeurige docent gevraagd wordt waarom hij toetsen inzet, is het antwoord bedroevend vaak: "om punten te verzamelen voor op het rapport", of "om te kunnen laten zien dat iemand alle vakken voldoende beheerst zodat we een diploma kunnen geven".

Voor het vaststellen van de ontwikkeling van individuele leerlingen worden toetsen daarbij vooral modulair gebruikt: in dit traditionele model van toetsen is er meestal sprake van losstaande toetsen (waarbij het, vooral voor de leerling), vaak onduidelijk is hoe deze zich tot elkaar verhouden. In dit model leidt iedere toets tot een zwaarwegende beslissing: voldoende of onvoldoende, zakken of slagen, overgaan of blijven zitten. Er hangt, met andere woorden, voor de leerlingen steeds veel af van een toets.

De beslissingen die uiteindelijk genomen worden komen vaak tot stand door middel van een berekening van een rekenkundig gemiddelde op de behaalde toetsen die in een specifiek onderdeel van het toetsprogramma zijn afgenomen. De kritische vraag kan gesteld worden of op basis van een dergelijke gemiddelde score een zorgvuldige beslissing genomen kan worden. Niet alleen geeft de opbrengst van een rekenkundige formule over het algemeen geen goed beeld van de daadwerkelijke beheersing van een vaardigheid, ook worden er vaak cijfers bij elkaar opgeteld en gedeeld die niet met elkaar vergelijkbaar zijn. Bijvoorbeeld doordat de toetsen die tot die cijfers hebben geleid verschillende onderwijskundige doelen hadden, er verschillende beoordelingssystematieken zijn gebruikt, of de toetsen niet van dezelfde kwaliteit waren. Door de *schijn*veiligheid van cijfers en het rekenen daarmee hebben we, het woord zegt het al., wel het idee dat we met deze cijfers tot een betrouwbare en zorgvuldige beslissing komen, terwijl dat dus in feite niet het geval is.

Deze rede bevat dan ook nadrukkelijk het pleidooi om beslissingen niet (enkel) te baseren op rekenkundige gemiddeldes, maar op de rijke informatie uit de, mogelijk aan cijfers toegevoegde, narratieve onderbouwingen van deze cijfers. Zodat er daadwerkelijk zorgvuldige beslissingen genomen kunnen worden die recht doen aan de leerling.

Het nemen van een beslissing op basis van informatierijke beoordelingen vereist dat docenten in staat zijn moeten om op grond van leeractiviteiten als toetsen, mentorgesprekken, stages en toetsdialogen, leerlingen betekenisvolle feed-up, feedback en feedforward te geven die bijdraagt aan hun ontwikkeling en die input kan geven voor de beslissingen die genomen moeten worden. Dit impliceert enerzijds dat docenten toetsing (moeten) zien als onlosmakelijk onderdeel van onderwijs, anderzijds dat zij toetsing (moeten) beschouwen als instrument om het leren te sturen (assessment as learning). Vanuit het perspectief van eigentijds beoordelen en beslissen is het hierbij essentieel dat het resultaat van één leeractiviteit (bijvoorbeeld een toets) *echt* benut moet kunnen worden om het leerproces te sturen, en dat op basis van één leeractiviteit met andere woorden geen zwaarwegende beslissingen, zoals zak-/slagbeslissingen, genomen zouden moeten worden. Zo krijgt de leerling, zoals ook eerder benoemd, de gelegenheid fouten te maken en de mogelijkheid zich écht te ontwikkelen. Deze visie op informatierijk beoordelen en

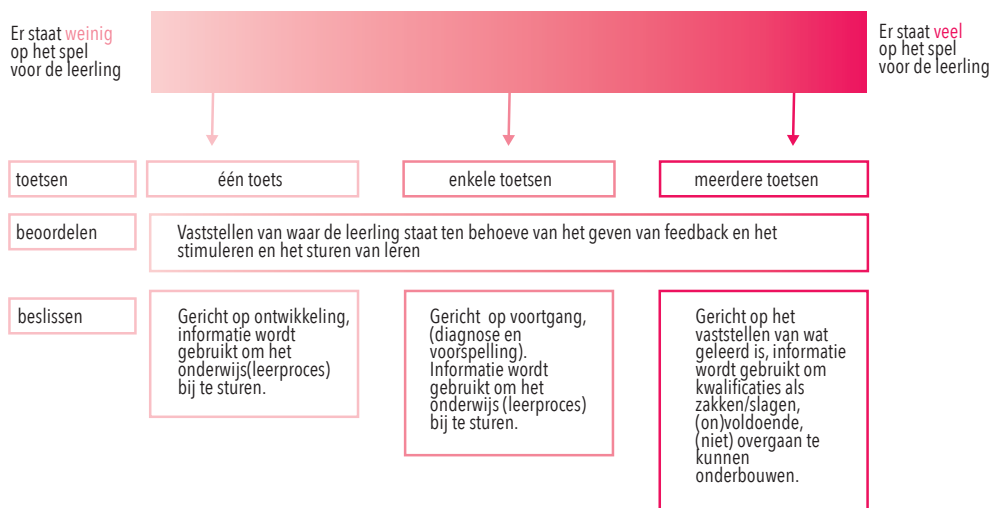
zorgvuldig beslissen sluit nauw aan bij de uitgangspunten van onder meer programmatisch toetsen (Van der Vleuten, et al., 2015; Van der Vleuten, et al., 2018).

Deze visie op beoordelen en beslissen betekent ook dat alle onderwijsactiviteiten (waar- onder toetsen) informatie leveren die het leren stuurt en informatie leveren die gebruikt kan worden om beslissingen te onderbouwen. De beoordelingen van deze activiteiten krijgen pas een summatieve waarde als hieraan een consequentie wordt verbonden. Het onderscheid tussen formatief en summatief als functie van één afzonderlijke toets wordt hiermee minder relevant en feitelijk vervangen door *de mate waarin iets op het spel staat* (Van der Vleuten, et al., 2018).

Er kunnen daarbij drie niveaus, ofwel drie typen beslissingen, onderscheiden worden, die in literatuur vaak aangeduid worden met de termen low stake, intermediate stake of medium stake, en high stake (Van der Vleuten, et al., 2015). Bij het eerste niveau staat er voor de leerling weinig op het spel, en is er met andere woorden sprake van 'weinig belang'. Dit is het geval als de informatie van een enkele toets- of leeractiviteit wordt gebruikt om zicht te krijgen op de ontwikkeling van de leerling. Deze informatie kan worden gebruikt om het onderwijs 'van morgen' invulling te geven of bij te stellen en om het onderwijsleerproces van de leerling te sturen.

Het tweede niveau is het niveau waarbij sprake is van een 'redelijk belang': er staat hier voor de leerling dus iets meer op het spel. De informatie uit toetsen wordt in dit niveau gebruikt om de ontwikkeling van de leerling vast te stellen om (tussentijdse) voortgangs- beslissingen te kunnen nemen. Deze tussentijdse beslissingen kunnen diagnostisch van aard zijn (wat gaat goed, wat kan beter, wat moet er nog gebeuren om de leerling de beoogde doelen te laten behalen?) en voorspellend (waar gaat dit naar toe, gaat de leerling met de huidige ontwikkeling de beoogde doelen behalen?) (Baartman et al., 2020).

Bij het derde niveau, tot slot, gaat het om zwaarwegende beslissingen. Er staat daarbij voor de leerling veel op het spel, er is dan ook sprake van 'veel belang'. Bij dit type beslissingen, dat bij voorkeur onderbouwd wordt vanuit een veelheid van beoordelingen, richt zich vooral op het vaststellen van wat er geleerd is. De verzamelde informatie wordt gebruikt om kwalificaties toe te kunnen kennen als 'gezakt/geslaagd', 'wel/niet toegelaten' of 'wel/niet bevorderd naar het volgende leerjaar/volgende onderwijsperiode'.



Figuur 6. Continuüm van de mate waarin er voor de leerling iets op het spel staat bij beoordelen en beslissen, geïnspireerd op Van der Vleuten et al., 2015 en Van der Steen et al., 2020).

Bewust inzetten van leer- en toetsactiviteiten

Bewustzijn van het onderscheid tussen beoordelen en beslissen is essentieel omdat het het mogelijk maakt om toets- en leeractiviteiten bewust met een specifiek doel in te zetten. Het geeft handvatten om heel bewust het onderscheid te maken naar de leer- en de beslisfunctie, waar in het volgende hoofdstuk uitgebreid op in zal worden gaan. Bij de leerfunctie van toetsen staat uiteraard het sturen en stimuleren van leerlingen centraal, bij de beslisfunctie het nemen van zorgvuldige beslissingen over het leren van leerlingen en over wat geleerd is (vaststellen van het leren). Om écht te mogen ontwikkelen, om echt fouten te mogen maken, en om recht te doen aan uitspraken als 'van toetsen kun je leren' is het van essentieel belang dat aan een (enkele) toets geen zwaarwegende beslissing wordt gekoppeld: feitelijk is daarmee immers de boodschap dat je géén fouten mag maken. Het is daarmee voor de leerling, met andere woorden, ook van groot belang dat hij weet wanneer een toets, of welke combinatie van toetsen, wél gebruikt wordt om een zwaarwegende beslissing op te baseren; het doel van de toets, het vaststellen van wat geleerd *is*, is dan expliciet anders dan vaststellen van wat nog geleerd *moet worden*.



5. FUNCTIES VAN TOETSEN

In de definitie van eigentijds beoordelen en beslissen nemen de drie functies die binnen toetsen worden onderscheiden een centrale positie in. In dit hoofdstuk wordt eerst toegelicht waarom er binnen het Lectoraat Eigentijds beoordelen en beslissen bewust voor gekozen is juist deze functies te onderscheiden. Vervolgens wordt beknopt ingegaan op ieder van de functies. Tot slot wordt stilgestaan bij hoe een bewuste balans gezocht kan en moet worden gezocht bij het implementeren van deze functies binnen de gangbare toetssystemen.

LEREN, BESLISSEN EN EVALUEREN

Gebruikelijk wordt binnen toetsen het onderscheid gemaakt tussen toetsen met een summatieve en formatieve functie. Informatie uit toetsen met een summatieve functie wordt gebruikt om leerlingen te certificeren of de effectiviteit van een curriculum te evalueren (Sluijsmans, et al., 2013). Toetsen met een formatieve functie daarentegen worden vooral ingezet om het proces van curriculumontwerp, onderwijzen en leren te verbeteren (Pryor & Crossouard, 2008). Belangrijk aandachtspunt is dat de aanduidingen 'formatief' en 'summatief' hierbij iets zeggen over de wijze waarop *omgegaan* wordt met de informatie die de toets oplevert: die wordt gebruikt om het leren vast te stellen (summatief) of het leren te sturen en stimuleren (formatief). Zoals onder meer Sadler (1989) en Hattie (2009) stellen, is niet de toets *zelf* formatief of summatief; deze begrippen zeggen iets over de beslissingen, de consequenties, die aan de informatie uit de toetsen verbonden worden. Doordat in de dagelijkse praktijk veelal gesproken wordt over 'formatieve toetsen' en 'summatieve toetsen', wordt hieraan vaak, ten onrechte, de conclusie gekoppeld dat formatief en summatief kenmerken of eigenschappen zijn van een toets.

Bovendien wekt de formulering 'formatieve toetsen' en 'summatieve toetsen' de suggestie dat een toets óf formatief is, óf

summatief. Ook dit is een onterechte aanname: toetsen omvatten – in meer of mindere mate – bij voorkeur steeds beide functies: zo geeft een toets met primair een summatieve functie informatie over waar de student staat en waarmee het proces van onderwijzen en leren kan worden verbeterd, deze toets heeft daarmee dus ook een formatieve functie (Van Schilt-Mol, et al., 2020). Om deze onjuiste connotaties te voorkomen wordt steeds vaker gesproken over formatief evalueren of formatief handelen ter vervanging van formatief toetsen.

Zoals bij de toelichting op de definitie besproken is het onderscheid tussen formatief en summatief minder relevant in de context van het eigentijds beoordelen en beslissen: de mate waarin iets voor de leerling op het spel staat (Van der Vleuten, et al., 2018) als indicatie voor de ‘zwaarte’ of de waarde van een toets lijkt meer passend. Binnen het Lectoraat Eigentijds beoordelen en beslissen is er dan ook bewust voor gekozen om bij de beschrijving van de functies van toetsen geen onderscheid te maken op grond van de terminologie formatief en summatief, maar om de functies te omschrijven vanuit de drie *doelen* die er mee worden nagestreefd: (1) het stimuleren en sturen van leren van leerlingen, (2) het nemen van zorgvuldige beslissingen over leerlingen en (3) het optimaliseren van de kwaliteit van het gegeven onderwijs (Brown, 2008; Van der Vleuten et al., 2015). Binnen het lectoraat staan, met andere woorden de *leerfunctie*, de *beslisfunctie* en de *evaluatiefunctie* van toetsen centraal.

Uitgangspunt hierbij is dat voor alle toetsen geldt dat alle drie de functies aan bod (kunnen) komen; de mate waarin dat gebeurt en de balans die daarbij wordt gezocht tussen de functies is uiteraard afhankelijk van onder meer de visie op onderwijs, en daarmee op leren en ontwikkelen.

Bij de leerfunctie staat het stimuleren en het sturen van het leren van leerlingen centraal, een functie die de laatste tijd veel aandacht krijgt. De tweede functie richt zich op het nemen van zorgvuldige beslissingen over leerlingen. Het gaat daarbij zowel om bijvoorbeeld beslissingen over het leerproces, als om het type beslissingen dat vaak wordt aangeduid als ‘zwaarwegende’ beslissingen, bijvoorbeeld het toekennen van studiepunten, het geven van een schooladvies of het uitreiken van een diploma. De derde functie is de evaluatiefunctie. In de literatuur wordt deze vooral omschreven als het evalueren op de kwaliteit van het onderwijs, gezien vanuit eigentijds beoordelen en beslissen is dat onlosmakelijk verbonden met het reflecteren op het eigen handelen. Van de drie functies is deze functie vaak nog het meest onderbelicht. Als docenten erkennen we dat informatie uit toetsen ons veel zegt over de leerling, maar kijken we nog te weinig naar wat deze informatie ons zegt over ons eigen handelen en over de kwaliteit van het onderwijs dat we gegeven hebben.

In de praktijk worden deze functies over het algemeen door iedereen onderschreven. In onderwijsvisies of toetsbeleidplannen worden hier vaak termen aan gekoppeld als 'ontwikkelingsgericht toetsen', 'eigenaarschap bij leerlingen', 'kansrijk adviseren' en uitingen als 'fouten maken mag'. Hoewel dit termen zijn die aansluiten bij de functies van toetsen, staat de daadwerkelijke uitvoering in de praktijk hier vaak haaks op.

In de volgende paragrafen worden, voorbeeldmatig, per functie een aantal voorbeelden weergegeven van hoe in de praktijk vaak ontwerpkeuzes gemaakt worden die er - onbewust - voor zorgen dat de visie van waaruit deze keuzes gemaakt worden, niet wordt gerealiseerd. Ook wordt heel beknopt ingegaan op de achtergronden van de verschillende functies. Beknopt, omdat er over iedere functie veel meer te zeggen is dan in deze rede mogelijk is. Een van de beoogde opbrengsten van het lectoraat is dan ook het uitwerken van de wijze waarop de drie functies van toetsen vorm kunnen krijgen in de verschillende onderwijssectoren, zodanig dat deze elkaar versterken.

Belangrijk is de kanttekening dat bij de voorbeelden die volgen gechargeerd wordt. De voorbeelden dienen in deze rede dan ook vooral als spiegel om nog eens kritisch naar ons eigen handelen te kijken als het gaat om het geven van een passende invulling aan de leerfunctie, de beslisfunctie en de evaluatiefunctie.

LEERFUNCTIE: TOETSEN OM HET LEREN VAN LEERLINGEN TE STIMULEREN EN TE STUREN

Uit veel onderzoek is bekend dat er een relatie bestaat tussen toetsen en leren. Zelfs als toetsen niet bewust worden ingezet vanuit de leerfunctie leveren toetsen an sich vrijwel altijd wel een bijdrage aan het leren: toetsen, en de wijze waarop wordt getoetst, stuurt immers het leren (onder meer: Dochy et al., 2015; Sambell et al., 2013; Cilliers et al., 2012). En hoewel ook zonder expliciete aandacht voor de leerfunctie van toetsen, en meer specifiek voor het geven van feedback, leerlingen leren van 'het leren voor een toets' (Waitling & Ginsburg, 2019), wordt het leereffect van toetsen verhoogd als er wél informatierijke feedback wordt gegeven (bijv. Malecka et al., 2020; Waitling & Ginsburg, 2019; Winstone et al., 2017; Gulikers & Baartman, 2017; Sluijsmans et al., 2013; Wiliam, 2013a, 2013b).

Als toetsen heel bewust worden ingezet om het leren te stimuleren en te sturen, kan dit grote effecten hebben op zowel de leerlingen zelf als op de leerresultaten. Zo komt uit veel onderzoek naar voren dat de leerfunctie een belangrijke positieve bijdrage kan leveren aan onder meer de motivatie, het eigenaarschap en de zelfregulatie bij leerlingen (Van der Ploeg & Van Engen, 2017; Stobart, 2008). Ook kan de leerfunctie een effect hebben op het realiseren van betere leerprestaties (Briggs, et al., 2012; Kingston & Nash,

2011; Ruiz-Primo & Furtak, 2007; Van der Ploeg & Van Engen, 2017). Het is dan ook niet verwonderlijk dat er de afgelopen jaren in toenemende mate aandacht is voor de wijze waarop specifiek invulling gegeven kan worden aan het sturen en het stimuleren van het leren, en aan wat dat vraagt van docenten (bijv. Van der Steen et al., 2020; Sluijsmans, 2020; Gulikers & Baartman, 2017; Sluijsmans & Seegers, 2017; Christodoulou, 2016; Furtak et al., 2016; Sluijsmans et al., 2013; Brookhart, 2011) en leerlingen (bijv. Dobbelear et al., 2021; Molloy et al., 2020; Winstone & Carless, 2020; Carless & Boud, 2018; Tai et al., 2018).

Ondanks deze toegenomen aandacht voor de leerfunctie blijkt dit in de praktijk niet eenvoudig. De leerfunctie betekent op de eerste plaats heel expliciet dat leerlingen fouten moeten kunnen en mogen maken zonder dat dit direct consequenties heeft voor een beslissing die over, met of voor hen genomen wordt. In de praktijk is het echter vaak zo dat aan iedere toets een cijfer wordt gekoppeld. Dit cijfer wordt vervolgens gebruikt om over een langere periode een gemiddelde te berekenen. Zo kan het bijvoorbeeld in de context van het voortgezet onderwijs heel goed voorkomen dat een leerling aan het begin van een periode nog wat lagere cijfers haalt, terwijl hij aan het eind van de periode de stof beter snapt en beheerst en dit ook zichtbaar wordt in hogere cijfers. Een beslissing over de uiteindelijke beheersing wordt desondanks vrijwel altijd genomen op basis van alle cijfers die de leerling heeft gehaald. Ondanks dat in veel visiestukken en beleidsdocumenten zinnen staan als *"leerlingen krijgen de ruimte om te leren en zich te ontwikkelen"*, worden leerlingen daar in de praktijk dus eigenlijk op afgerekend.

Ruimte voor ontwikkeling of afrekenen op ontwikkeling?

Maarten heeft gedurende periode 4 van 4 havo drie toetsen gehad voor Aardrijkskunde. Achtereenvolgens haalde hij een 5, een 6,4 en een 8. In de laatste toetsweek sloot Maarten de periode af met een toets waarin alle aangeboden stof uit periode 4 getoetst werd. Voor deze toets haalde Maarten een 9. Op zijn rapport staat *'dus'* een 7,1.

Wat zeggen de cijfers op deze toetsen en op het rapport nu eigenlijk? Zeggen deze dat de gemiddelde beheersing van Maarten een 7,1 'is'? Of laten de afzonderlijke cijfers zien dat Maarten aan het begin van periode 4 nog niet alle vaardigheden beheerste en zich gedurende de onderwijsperiode op deze vaardigheden heeft ontwikkeld? En dat die '9' dus wellicht een beter beeld geeft van wat Maarten nu kan?

Hoewel bovenstaand voorbeeld bij vrijwel alle docenten, en ook bij ouders en bij leerlingen, het gevoel oproept dat hier iets 'misgaat', is het voorbeeld wel een getrouwe weergave van wat er op grote schaal in de praktijk gebeurt. Als we écht de leerfunctie willen omarmen en écht op waardevolle wijze invulling willen geven aan de in de visiedocumenten opgenomen termen als 'ontwikkelingsgericht toetsen', 'gepersonaliseerd leren' en 'fouten maken mag', betekent dit dat we heel kritisch moeten kijken naar hoe we de leerling daartoe écht kunnen ruimte kunnen geven en moeten we ons handelen daarop aanpassen.

Dat geldt dan niet alleen voor de wijze waarop we de leerlingen beoordelen en hier beslissingen aan verbinden, dit geldt per definitie ook voor het *ontwerp* van ons onderwijs. Hoewel het onderstaande plaatje wellicht in eerste instantie vooral een lach zal ontlokken, is het daarnaast waarschijnlijk pijnlijk herkenbaar.



Figuur 7. Illustratie over het 'hordelopen-gevoel' in het onderwijs

Want hoewel we weten dat toetsen het leren stuurt (onder meer: Dochy et al., 2015; Sambell et al., 2013; Cilliers et al., 2012), en dat het inzetten van toetsen ons daardoor ook bij uitstek mogelijkheden biedt om het leren te sturen (onder meer: Waitling & Ginsburg, 2019; Winstone et al., 2017; Gulikers & Baartman, 2017), én hoewel we weten dat toetsen die worden ingezet na het leren niet alleen over het algemeen geen bijdrage meer leveren aan het leerproces maar daarnaast ook een negatief effect kunnen hebben op de motivatie van leerlingen (Hattie & Timperley, 2007, Sadler, 1989), worden toetsen veelal toch vooral ingezet ter afsluiting van onderwijs. Hierdoor ontstaat voor de leerlingen een 'hordeloop-gevoel': 'we volgen onderwijs, we sluiten dat af met een toets, en we gaan door met het volgende vak'.

Door in ons onderwijsontwerp toetsen bewust onderdeel te laten zijn van het onderwijs(leerproces) kunnen we ervoor zorgen dat de informatie uit deze toetsen door ons als docenten gebruikt kan worden om het onderwijs bij stellen – zodanig dat de leerlingen zich kunnen ontwikkelen richting de beoogde leerdoelen – en dat de informatie uit de toetsen door de leerlingen gebruikt kan worden zelf zicht te krijgen op en richting te geven aan hun leerproces.

Het ontwerpen van onderwijs waarbij recht gedaan wordt aan de leerfunctie van toetsen stelt eisen aan de schoolorganisatie én vraagt uiteraard specifieke bekwaamheden van zowel docenten als leerlingen. Schildkamp et al (2014) concluderen op basis van een uitgevoerde reviewstudie naar formatief handelen, dat in deze rede uiteraard onder de scope van de leerfunctie valt, onder meer dat leerlingen moeten kunnen reflecteren op eigen werk en een actieve rol in het toetsproces moeten hebben, en dat toetsen betekenisvol moeten zijn en moeten voorzien in feedback. Gulikers en Baartman (2017) hebben onderzoek gedaan naar wat dit concreet betekent voor het docent-handelen en voor de effecten die dit handelen heeft. Op basis van een reviewstudie hebben zij de formatieve toetscyclus ontwikkeld die bestaat uit 5 stappen: (1) verwachtingen verhelder, (2) reacties ontlokken en verzamelen, (3) reacties analyseren en interpreteren, (4) communiceren over resultaten en (5) ondernemen vervolgacties en aanpassen van leren. Zij stellen vast dat om deze stappen goed te kunnen doorlopen docenten vakinhoudelijke kennis moeten kunnen toepassen, dat zij moeten beschikken over een vakdidactische handelingsrepertoire en dat zij activiteiten moeten kunnen ontwikkelen om leerlingen actief in te zetten in de verschillende fasen van de cyclus.

Om meer zicht te krijgen op wat het waardevol invulling geven aan de leerfunctie door de leerlingen zelf van leerlingen vraagt, is binnen het lectoraat eigentijds beoordelen en beslissen onderzoek gedaan naar de concrete vaardigheden die formatief handelen stelt aan leerlingen (Dobbelaer et al., 2021b). Het onderzoek van Gulikers & Baartman (2017) is

hierbij als uitgangspunt gehanteerd, aanvullend is een literatuuronderzoek uitgevoerd en zijn experts bevestigd. Dobbelaer et al (2021b) stellen vast dat er (ten minste) vijf belangrijke vaardigheden zijn die leerlingen moeten beheersen. Op de eerste plaats is dat *leren verhelderen van de verwachtingen*: hoe beter de leerling weet wat er van hem verwacht wordt, hoe beter hij in staat zal zijn zijn eigen kennis en kunde te beoordelen en na te denken over wat er nog nodig is om de doelen te behalen (Carless & Boud, 2018). Op de tweede plaats moeten leerlingen in staat zijn zelf *feedback te vragen én te ontvangen*. Dit vereist dat leerlingen de waarde van de feedback inzien voor hun eigen leerproces. Om de feedback ook te accepteren en er iets mee te kunnen doen, moeten de leerlingen om kunnen gaan met de effecten die zowel positieve als negatieve feedback kan hebben op hun emoties (Carless & Boud, 2018; Molloy et al., 2020). Een derde vaardigheid is het *kunnen geven van peerfeedback*: het kunnen geven van peerfeedback veronderstelt immers dat de leerlingen goed zicht hebben op de eisen die aan hen gesteld worden om een afgesproken doel te behalen. De peerfeedback heeft niet alleen voordelen voor degene die de feedback ontvangt - bijvoorbeeld een grotere variëteit aan feedback (Lindler & Visser, 2018) -, maar ook voor de feedbackgever: zo bevordert het geven van peerfeedback de eigen kritische blik van de leerling (Pearce et al., 2009). De vierde vaardigheid richt zich op *het ondernemen van vervolgacties*: leerlingen moeten beschikken over een repertoire van strategieën om productief te kunnen handelen om effectief gebruik te maken van de ontvangen feedback (Carless & Boud, 2018). Met behulp van de feedbackvragen (waar ga ik heen, waar sta ik nu, en wat ga ik doen, Hattie & Timperly, 2007), moeten zij in staat zijn zelf doelen stellen, een plan van aanpak te maken om de feedback te verwerken en hun ontwikkeling te monitoren (Molloy et al., 2020), om zo het verschil tussen de huidige en gewenste situatie te verkleinen (Dobbelaer et al., 2021b). Tot slot moeten leerlingen de *waarde van een feedbackcultuur inzien* en hier een actieve bijdrage aan kunnen leveren. Zij moeten op een constructieve manier het gesprek over ontwikkeling van zichzelf en anderen kunnen voeren, en fouten durven maken. Om leerlingen te ondersteunen bij het ontwikkelen van deze vaardigheden is vanuit het lectoraat een toolkit ontwikkeld (Dobbelear et al., 2021a).

Ook in het promotieonderzoek van Janneke van der Steen, dat zich richt op het ontwerpen van programma's van samenhangende formatieve leeractiviteiten, is onderzoek gedaan wat het invulling geven aan de leerfunctie vraagt van schoolleiders, docenten en leerlingen (Van der Steen et al., in voorbereiding). Janneke ontwikkelde zogenoemde vignetten die voor ieder van de drie groepen inzicht geven in het vereiste gedrag, de houding en de vaardigheden. Ook voor de schoolorganisatie ontwikkelde zij een dergelijk vignet (dat inzicht geeft in de eisen die aan de schoolorganisatie gesteld moeten worden). In onderstaand tekstkader is voorbeeldmatig het vignet voor docenten opgenomen.

In deze klas is er, dankzij de docent, sprake van een veilig en vertrouwd leerklimaat. De docent zorgt ervoor dat de leerlingen een actieve rol hebben, vragen durven te stellen, feedback geven en fouten mogen maken. De docent heeft goede leerling-docent relaties waarbij de leerling zelf aan het roer staat van zijn eigen leerproces en mede-leerlingen gebruikt om dit leerproces te ondersteunen. De docent heeft hierbij een coachende rol en maakt effectief gebruik van feed-up, feedforward en feedback. De docent is overtuigd van het nut en de noodzaak van formatief evalueren voor de leerprogressie van leerlingen. Hij heeft het vertrouwen dat formatief evalueren het nemen van beslissingen in het onderwijs en het leren van de leerlingen kan ondersteunen. Het is echt onderdeel van het leerproces. Hierdoor durft de docent toetsen met een summatief karakter meer los te laten. In de programma's van formatieve leeractiviteiten, die de docent voor elke periode maakt, worden zowel zijn vakinhoudelijke en vakdidactische expertise, als ook zijn ontwerpvaardigheden voor deze programma's zichtbaar. Waar mogelijk en passend gebruikt de docent technologie om het proces van formatieve evaluatie te ondersteunen. Hij is goed in staat om leerdoelen te vertalen naar een lesprogramma waarbinnen op een formatieve manier wordt gewerkt aan de leerdoelen. De docent durft deze geplande lessen echter ook los te laten wanneer informatie uit formatieve leeractiviteiten laat zien dat er op dat moment iets anders nodig is. Hij heeft voldoende kennis van de kern van het vak en lange leerlijnen en beschikt over vakdidactisch handelingsrepertoire om dit te kunnen en durven doen. De docent heeft vertrouwen in het kunnen ontwerpen, uitvoeren en evalueren van haar programma's van formatieve leeractiviteiten maar erkent dat hij hierbij ook zelf nog lerende is en fouten mag maken. In dit leerproces maakt de docent gebruik van eigen zelfreflecties, kennisdeling met anderen, maar vraagt ook feedback van leerlingen en collega's.

BESLISFUNCTIE: TOETSEN OM ZORGVULDIGE BESLISSINGEN TE KUNNEN NEMEN

De tweede functie die centraal staat in de definitie van eigentijds beoordelen en beslissen is de beslisfunctie. Met grote regelmaat wordt gesteld dat de kwaliteit van toetsen niet ter discussie mag staan, omdat op basis daarvan belangrijke beslissingen genomen worden over (het leren van) leerlingen (Van Schilt-Mol & Sluijsmans, 2015). In het voorgaande, bij de toelichting op de definitie van eigentijds beoordelen en beslissen, werd al beknopt ingegaan op de waarde van cijfers bij het nemen van beslissingen. In deze paragraaf wordt daar meer op ingezoomd.

Iedere docent is ervan doordrongen dat hij verantwoordelijk is voor het nemen van zorgvuldige beslissingen, en van de consequenties die deze beslissingen kunnen hebben voor leerlingen. Hoewel docenten zich er meestal van bewust zijn dat het niet altijd lukt om daadwerkelijk zorgvuldige beslissingen te nemen, gaat het in veel gevallen om onbewuste onzorgvuldige beslissingen. Bekende 'valkuilen' waar docenten zich over het algemeen van bewust zijn maar die soms toch voorkomen, zijn bijvoorbeeld het halo-effect (een gunstige indruk wordt doorgezet in een gunstige beoordeling: 'hij heeft zo zijn best gedaan, dat zit wel goed'), het horn-effect (een ongunstige indruk wordt omgezet in een ongunstige beoordeling: 'waarom heeft hij niets met mijn feedback gedaan?') en centrale tendentie (neiging om 'in het midden' te beoordelen: 'een tien halen kan niet, het gaat om een 7 of een 8') (Joosten-ten Brinke & Draaier, 2015; Van Berkel et al., 2014).

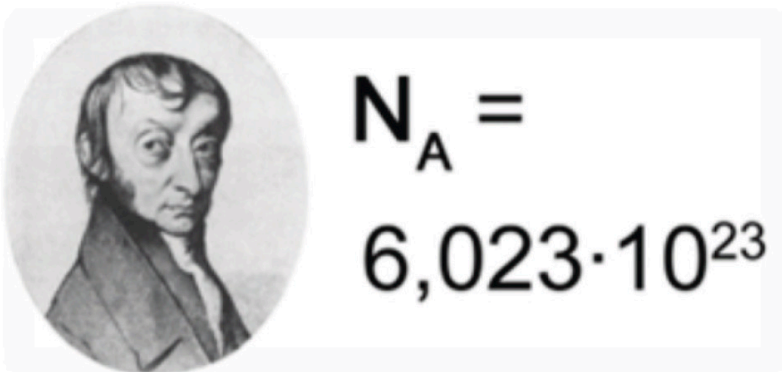
Van onbewuste zorgvuldige beslissingen bijvoorbeeld ook sprake als docenten onbewust andere factoren dan prestaties laten meewegen bij het beoordelen en het vervolgens koppelen van een beslissing aan deze beoordeling. Dit zijn bijvoorbeeld factoren als het talent van de leerling, de geleverde inspanning, het leergedrag in de klas en de mate waarin een leerling aanmoediging nodig heeft (Resh & Sabbagh, 2016). Daarnaast zijn er indicaties dat naast deze factoren vooroordelen van docenten een rol spelen bij de verwachtingen die docenten van leerlingen hebben (Van Vijfeijken & Van Schilt-Mol, 2020) en dat de sociaaleconomische status van een leerling kan zorgen voor onbewuste beoordelingsfouten (Autin, et al., 2015): hoewel meta-analyses hebben aangetoond dat docenten over het algemeen accurate verwachtingen hebben van hun leerlingen, is dit voor specifieke groepen als leerlingen met een lagere sociaaleconomische status anders (Jussim & Harber, 2005). Deze inaccuraten inschattingen kunnen voor leerlingen uit deze laag(re) sociale groepen vervolgens ook leiden tot lagere cijfers, en daarmee onrechtvaardige beslissingen (Malouff & Thornsteinsson, 2016). Een voorbeeld hiervan zijn de adviezen die docenten aan leerlingen geven bij de overgang van basis- naar voortgezet onderwijs: uit onderzoek blijkt dat de vooroordelen ten aanzien van leerlingen uit lagere sociale milieus kunnen leiden tot lagere schooladviezen (De Boer, et al., 2016; Timmermans, et al., 2015), waardoor leerlingen met lager opgeleide ouders niet vanzelfsprekend dezelfde leerkanalen krijgen als leerlingen met hoger opgeleide ouders (De Wolf, 2017).

Het is dan ook niet verwonderlijk dat veel leerlingen weleens het gevoel hebben gehad 'oneerlijk' beoordeeld zijn voor hun (school)werk: dit gevoel wordt veroorzaakt door de aanname dat mogelijk andere aspecten dan uitsluitend prestatie hebben geleid tot een te laag cijfer of een te laag schooladvies (Van Vijfeijken et al., in press). Deze gevoelens van onrechtvaardige behandeling komen in alle lagen van de bevolking voor (Resh & Sabbagh, 2016). Het is van groot belang om deze gevoelens zo veel mogelijk te voorkomen omdat deze negatieve consequenties kunnen hebben op de mate waarin een leerling zich

verbonden voelt met school, op hoe de leerling zich op school gedraagt en zelfs op de houding die de leerling ontwikkelt ten aanzien van de maatschappij (Van Vijfeijken, et al., in press; Resh & Sabbagh, 2016). In het boek "Wat is eerlijk? Werken aan kansengelijkheid in het onderwijs." dat momenteel onder redactie van Marijke van Vijfeijken tot stand komt, wordt dieper ingegaan op de relatie tussen rechtvaardig beoordelen en kansengelijkheid in het onderwijs (Van Vijfeijken, in press).

Hoewel, zoals eerder gezegd, de meeste docenten zich er dus wel van bewust zijn dat het, bijvoorbeeld als gevolg van (on)bewuste vooroordelen of als gevolg van beoordelaarsvalkuilen, niet altijd lukt om een zorgvuldige en rechtvaardige beslissing te nemen, zijn te weinig docenten zich expliciet bewust van de kanttekens die geplaatst zouden moeten worden bij het gebruik van cijfers bij het nemen van zorgvuldige en rechtvaardige beslissingen. Onderstaande voorbeelden illustreren de kanttekeningen die bij het huidige gebruik van cijfers geplaatst kunnen worden.

Het eerste voorbeeld betreft het getal van Avogadro, ook wel de constante van Avogadro genoemd. Amedeo Avogadro is de natuur- en scheikundige die de grondlegger is van deze constante. Deze constante geeft de verhouding aan tussen het aantal deeltjes van de atomaire en moleculaire wereld en de elementen van de door de mens waarneembare macroscopische wereld. Om het getal van Avogadro te kunnen duiden, is het belang te begrijpen wat deze verhouding precies betekent. Voor iemand zonder enige kennis van de onderliggende natuurkundige en scheikundige principes is het getal van Avogadro een getal zonder betekenis. Doordat de noodzakelijke kennis ontbreekt, heeft het getal voor een persoon zonder deze kennis geen eenduidige waarde.



Figuur 8. Het getal van Avogadro

Toch is juist die échte waarde van een cijfer essentieel bij het gebruik van een cijfer. Desondanks geldt voor veel cijfers en de wijze waarop we die gebruiken eigenlijk hetzelfde

als wat voor veel mensen geldt voor het getal van Avagadro: de échte waarde is onduidelijk. Zo is het de vraag of een docent het verschil kan uitleggen tussen een 6,2 en een 6,3: wat maakt nu werkelijk dat de vaardigheid of kennis van de ene leerling beoordeeld wordt met een 6,2 terwijl die bij een andere leerling beoordeeld wordt met een 6,3? Wat kan de ene leerling wel dat de andere niet kan? Voor leerlingen daarentegen is het verschil tussen een 5,49 en een 5,51 overduidelijk: de conclusie die door hen aan specifiek de decimalen achter de komma verbonden wordt, maakt voor hen het verschil tussen een voldoende of onvoldoende.

Het écht snappen en écht kunnen duiden van de *waarde* van een cijfer is zoals al eerder benadrukt cruciaal als je er belangrijke consequenties aan gaat verbinden. Het tweede voorbeeld dat dit illustreert is het werk van de Zwitserse kunstenaar en komiek Ursus Wehrli. Wehrli is bij het grote publiek vooral bekend van boeken en presentaties waarin hij kunstwerken en voorwerpen 'opruimt' (bijv. Wehrli, 2006; Wehrli, 2013). Zo staat er op de kaft van zijn boek 'The Art of Clean Up' (2013) een keurig 'opgeruimd' bakje met fruit. In het boek is de tweeluik van dit fruitbakje te vinden: eerst wordt een foto getoond waarop het bakje gemengd fruit in zijn geheel te zien is, daarnaast de foto waarin alle ingrediënten keurig op een rij naast elkaar zijn gesorteerd: een leeg schoteltje en een lepeltje worden vergezeld van rijtjes met stukjes appel, kiwi, mandarijn, banaan en een toefje slagroom.

Dit bakje fruit in de twee verschijningsvormen is een mooie metafoor voor de wijze waarop in het onderwijs vaak omgegaan wordt met cijfers: we gaan vaak voorbij aan wat de inhoud van dat bakje met gemengd fruit nu echt is. Bij het nemen van een beslissing baseren we ons vaak op een rekenkundig gemiddelde. Om dit gemiddelde te berekenen stoppen we alle cijfers die we hebben in een rekenkundige formule om er één getal van te maken. We maken een fruitshake zagezegd. We gaan daarbij, om in de fruitmetafoor te blijven, volledig voorbij aan de verschillende smaken en structuren van het fruit, en gaan voorbij aan de rijkheid van alle informatie. Een shake is lekker, maar veel minder geschikt om een uitspraak te doen over beheersing van verschillende aspecten dan de afzonderlijke onderdelen uit dat bakje fruit.

Bovendien wordt, zoals eerder ook al benoemd, bij het vaststellen van een gemiddelde heel vaak gerekend met cijfers die eigenlijk niet met elkaar of onderling vergelijkbaar zijn. Bijvoorbeeld doordat de toetsen die tot die cijfers hebben geleid verschillende onderwijskundige doelen hadden of omdat er verschillende beoordelingssystematieken zijn gebruikt. Of doordat de toetsen niet vergelijkbaar waren qua toets-technische kwaliteit. Of, doordat een cijfer niet toegekend werd aan een prestatie, maar bijvoorbeeld aan het aanwezig zijn tijdens een les, aan het spieken bij een medeleerling (voor deze toets krijg je een '1') of doordat een cijfer wordt toegekend aan het uitblijven van een prestatie (zie tekstkader).

“Vorig schooljaar haalde Joris in periode 4 voor alle wiskundetoetsen een hele krappe voldoende: een 6,0, een 5,9, een 6,1 en nog een 6,0 waren het resultaat van heel hard werken aan wiskundevaardigheden. Door ziekte had hij 1 proefwerk gemist en niet ingehaald. De dag van de rapportvergaderingen werd Joris door zijn mentor gebeld: “Niet schrikken Joris, maar op je rapport heb je een 5 voor wiskunde. Die toets die je niet had gemaakt, is in het volgsysteem ingevuld met een beoordeling ‘1’. Gemiddeld heb je deze periode dus een 4,75.”. Wat zegt dit rapportcijfer? Waarom is het rekenkundig gemiddelde vastgesteld op een 5, terwijl de andere resultaten de docent toch duidelijk hadden moeten kunnen maken dat Joris er steeds opnieuw net in slaagt een voldoende te behalen?”

(Van Schilt-Mol, 2018. In: Examens 2018-4. P. 26)

Het feit dat gemiddeldes tot stand komen op basis van een formule wekt heel vaak ten onterechte de suggestie dat we kunnen ‘vertrouwen’ op de uitkomst van de rekensom: maar wat hebben we eigenlijk berekend als de waardes in de rekenformule niet vergelijkbaar zijn? Bovendien is uit onderzoek bekend dat de cijfers die op basis van de formule berekend zijn niet altijd ‘hetzelfde zeggen’: de ene 6 is de andere niet doordat een cijfer alléén niet altijd vanzelfsprekend inzicht geeft in de daadwerkelijke beheersing van specifieke vaardigheden. Zo toont Jeroen van der Linden in zijn, nog in uitvoering zijnde onderzoek (in het hoger beroepsonderwijs) aan, dat er sprake is van een complexe relatie tussen het behalen van een cijfer en de inzet van zelfregulatie-strategieën: deze relatie wordt vooral beïnvloed door de mate waarin studenten in staat zijn bewuste keuzes te maken in zelfregulatie-strategieën en door het doel dat de studenten zichzelf hebben gesteld (performance goal/halen van de toets of mastery goal/professionele ontwikkeling). Een ‘net voldoende’ laat dus niet per definitie zien dat iemand de stof ook ‘net voldoende’ beheerst, maar kan ook verklaard worden door het effectieve gebruik van exact op de eisen van de toets afgestemde zelfregulatie-strategieën (Van der Linden, et al., 2021). Of zoals een student het verwoordde: *“Als ik een 8 haal voor een tentamen en een 5 voor ander tentamen dan betekent dat dat ik te veel tijd besteed heb aan het leren voor het ene vak, en te weinig aan het andere.”*

De twee voorbeelden maken duidelijk dat we met elkaar de schijnveiligheid van cijfers in stand houden. Deze rede is dan vooral ook een oproep om eens kritisch te kijken naar de wijze waarop u zelf cijfers gebruikt bij de totstandkoming van een beslissing. Zoals toege-

licht bij zowel de definitie van eigentijds beoordelen en beslissen als bij de leerfunctie, is ook bij het nemen van een beslissing de waarde van een beoordeling essentieel. Alleen door informatierijke beoordelingen als basis te gebruiken kan een zorgvuldige, rechtvaardige en betrouwbare beslissing worden genomen, waarbij de informatie kan worden gebruikt om de beslissing te onderbouwen.

De afgelopen tijd wordt, gelukkig, binnen alle onderwijssectoren in toenemende mate gesproken over het nemen van beslissingen op basis van narratieve beoordelingen in plaats van op uitsluitend (informatie-arme) cijfers. Binnen het toetsconcept programma-tisch toetsen, bijvoorbeeld, is een leidend principe dat er aan één toetsmoment nooit een zwaarwegende beslissing gekoppeld mag worden (op basis van één informatiepunt kan er immers geen volledig beeld gevormd worden van de beheersing van complexe vaardigheden). Een zwaarwegende beslissing wordt pas genomen als er voldoende rijke informatie verzameld is (bij voorkeur gegeven door beoordelaars vanuit verschillende contexten of perspectieven) (onder meer Van der Vleuten et al., 2015; Van der Vleuten et al., 2018; Baartman, Van et al., 2020). Hoewel de beoordeling hierbij bij voorkeur wordt uitgedrukt in kwalificaties als onder/volgens/boven beoogd niveau, wordt bij opleidingen in het hoger (beroeps)onderwijs, ten gevolge van vigerende onderwijs- en examenrege-lingen, aan de beoordeling uiteindelijk aanvullend vaak toch ook een cijfer gekoppeld (Baartman et al., 2020).

Ook bij pass/fail-beoordelingen wordt veelal gebruik gemaakt van narratieve beoorde-lingen als basis voor de beoordeling. Onderzoek naar meerwaarde van het gebruik van pass/fail-beoordelingen ten opzichte van cijfers laat zien dat dit type beoordelingen in potentie de waarde heeft de motivatie van studenten te verhogen en studenten in staat te stellen effectiever de door de student gestelde doelen te behalen (Ramaswamy et al., 2020). Wel moet worden opgemerkt dat deze wijze van beoordelen uitdagingen met zich mee brengt. Het vereist onder meer een heel helder beeld van de beoogde leerresulta-ten en de 'lat' die hierbij wordt geformuleerd (Schut et al., 2020; Ramaswamy et al., 2020), waarbij het van belang is dat de uitwerking van de leerresultaten niet te gefragmenteerd of analytisch zijn beschreven, omdat dan het holistische einddoel - de beroepsbekwa-me student - uit het oog wordt verloren (Baartman et al., 2020; Baartman & Prins, 2018; Sadler, 2009). Het vereist daarnaast ook een bepaalde mate van beoordelaarsexpertise: beoordelaars moeten in staat zijn te komen tot een professioneel holistisch oordeel. Bij het beoordelen en beslissen kan bij deze manier van beoordelen immers niet meer worden 'vertrouwd' op het rekenen met cijfers, maar moeten de beoordelaars in staat zijn (op basis van goed uitgewerkte beschrijvingen van het te behalen niveau) de narratieve informatie zorgvuldig te interpreteren om te komen tot een betrouwbare beoordeling en beslissing (Baartman et al., 2020; Oudkerk Pool et al., 2018; Van der Vleuten, 2017).

DECISION-DRIVEN DATA-COLLECTION

Deze visie op het nemen van beslissingen op basis van (bij voorkeur meerdere) informatieve beoordelingen, vereist behalve het kunnen interpreteren van de informatie ook dat docenten zich bewust zijn van het belang van samenhang tussen de toets-/leeractiviteiten in relatie tot het onderwijs en de leerdoelen, en van het doel waarmee zij de informatie verzamelen (geven van feedback, nemen van een beslissing).

In het onderwijs is het vaak gebruikelijk om beslissingen – over leerlingen, over de kwaliteit van het onderwijs, etc. – te baseren op de informatie die voorhanden is. In onderwijssectoren waar veel gewerkt wordt met methodegebonden toetsen (toetsen die onderdeel uitmaken van de lesmethode die door de docent wordt gebruikt), zoals in het basis- en voortgezet onderwijs, lijkt dit in eerste instantie ook vanzelfsprekend. Het op deze manier gebruiken van informatie uit toetsen, ook wel ‘opbrengsten’ of ‘data’ genoemd, wordt aangeduid met termen als opbrengstgericht werken en data-based decision-making. Bij beide worden data verzameld en geanalyseerd om beslissingen over leerlingen te nemen die gericht zijn op bijvoorbeeld het verbeteren van schoolprestaties, het leerproces en de kwaliteit van het onderwijs (Heitink, et al., 2016; Schildkamp & Kuiper, 2010)

Het gebruiken van data die zonder meer beschikbaar zijn omdat ze bijvoorbeeld deel uitmaken van de lesmethode heeft echter mogelijk twee consequenties: op de eerste plaats is het mogelijk dat bestaande data niet geschikt zijn of simpelweg niet ‘just in time’ beschikbaar zijn voor bijvoorbeeld lesvoorbereidingen of last-minute aanpassingen in instructie (William, 2013a). Een andere consequentie is dat de informatie uit de bestaande data mogelijk niet exact aansluit op leerdoelen die je als docent zelf, al dan niet aanvullend op bestaande leerdoelen, hebt geformuleerd. Er dan is, met andere woorden, niet alleen geen sprake van constructieve afstemming, het is daarnaast daardoor ook niet mogelijk zorgvuldige beslissingen te nemen.

William (2013a) pleit er dan ook voor om heel *bewust* gegevens over leerlingen te verzamelen om daarmee beter onderbouwde beslissingen te kunnen nemen. Sterker nog: hij betoogt om pas data te verzamelen als duidelijk is waartoe en op welke manier deze gebruikt worden om beslissingen te nemen. Het gaat daarbij dan zowel om beslissingen over het leren en leerproces, als om het vaststellen van wat geleerd is. Als tegenhanger van data-based decision-making noemt hij dit decision-driven data-collection (William, 2013). Bij decision-driven data-collection ligt de focus op, zoals de term al aangeeft, het bewust verzamelen van informatie ten behoeve van een beslissing die genomen moet worden. Zoals eerder beschreven kan het hierbij zowel gaan om beslissingen die gericht zijn op de ontwikkeling van de leerling (waarbij de informatie wordt gebruikt om het onderwijs(leerproces) bij te sturen) als om beslissingen die gericht zijn op de voortgang

(diagnose en voorspelling) of op het vaststellen van wat er geleerd is (waarbij informatie wordt gebruikt om kwalificaties als zakken/slagen, (on)voldoende, (niet) overgaan te kunnen onderbouwen). Afhankelijk van het type beslissing bepaalt de docent wat voor soort informatie en hoeveel informatie hij nodig heeft om de beslissing onderbouwd te kunnen nemen. Wanneer er meer informatie nodig is, is het van essentieel belang dat de combinatie van de activiteiten die wordt ingezet om te informatie te verzamelen (toetsen, leeractiviteiten, etc.) passend zijn bij het leerdoel dat centraal staat en dat de onderlinge samenhang tussen deze activiteiten duidelijk is. Het is, met andere woorden, essentieel dat er sprake is van een samenhangend onderwijs- en toetsprogramma waarbij de in dit programma opgenomen activiteiten niet alleen de duidelijke relatie hebben met de rug-gengraat van de opleiding (waartoe wordt opgeleid) en de onderliggende doelen, maar ook onderling (o.a. Van der Steen, 2020; Van der Steen et al., 2020; Baartman et al., 2020).

EVALUATIEFUNCTIE: TOETSEN OM DE KWALITEIT VAN HET ONDERWIJS TE OPTIMALISEREN

De laatste functie waar op ingegaan wordt, is de evaluatiefunctie. Bij deze functie staat het gebruiken van informatie uit toetsen om de kwaliteit van het onderwijs te verbeteren centraal. Hoewel onderstaand citaat (Figuur 9) vanzelfsprekend zou moeten zijn, is het dat helaas nog lang niet altijd het geval. Zo blijkt bijvoorbeeld uit een recente studie dat docenten in het voortgezet onderwijs leeractiviteiten vooral inzetten om verwachtingen ten aanzien van leerdoelen te verhelderen en leerlingreacties te ontlocken en verzamelen (de eerste stappen uit de formatieve toetscyclus van Gulikers en Baartman, 2017), waarmee leerlingen en docenten zicht krijgen op waar ze staan ten opzichte van de leerdoelen, maar nog minder geneigd zijn om op basis van deze opbrengsten hiervan het onderwijs ("teaching and learning") bij te stellen (Veugen et al., 2021).



Figuur 9. Evaluatiefunctie volgens een leerling

Uit onderzoek blijkt dat het benutten van de informatie uit toetsen daadwerkelijk kan worden gezien als een effectieve interventie om de kwaliteit van het onderwijs te verbeteren (Briggs, et al., 2012; Offerdahl, et al., 2018). Dit veronderstelt uiteraard wel dat de toetsen niet alleen betekenisvol moeten zijn en moeten voorzien in de juiste informatie, maar vooral ook dat docenten de kennis en vaardigheden moeten hebben om toetsgegevens te kunnen verzamelen, dat zij in staat moeten zijn de informatie uit de toetsen te analyseren en interpreteren én dat zij in staat moeten zijn op basis van deze interpretaties hun lesgeven te verbeteren (Schildkamp, et al., 2014). Schildkamp et al. stellen dat er, om de informatie uit toetsen te kunnen benutten voor het verbeteren van onderwijskwaliteit, daarnaast binnen de school sprake moet zijn van een cultuur waarin dit verbeteren centraal staat en waar geïnvesteerd wordt in tijd (bijvoorbeeld om te kunnen samenwerken aan formatieve toetsen) en middelen (professionele ontwikkeling).

De evaluatiefunctie vraagt uiteraard van docenten dat zij bereid en in staat zijn te reflecteren op hun eigen handelen en als kritisch professional vervolgens ook dit handelen aan durven en kunnen passen. Om daadwerkelijk goede invulling te kunnen geven aan de evaluatiefunctie, dit geldt overigens vanzelfsprekend ook voor de leer- en beslisfunctie, is met andere woorden een voldoende mate van toetsbekwaamheid essentieel. In de nieuwe bekwaamheidseisen die in het Besluit bekwaamheidseisen onderwijspersoneel (Ministerie van Onderwijs, Cultuur en Wetenschap, 2017) zijn vastgelegd voor docenten en docenten in het primair en voortgezet onderwijs en het middelbaar beroepsonderwijs, is opgenomen dat docenten doelen en passende vormen van evalueren en toetsen moeten kennen, toetsen moeten kunnen kiezen, ontwikkelen of samenstellen, toetsresultaten moeten kunnen beoordelen, analyseren en interpreteren, en de kwaliteit van toetsen en examens moeten kunnen beoordelen. Daarnaast is benoemd dat docenten de voortgang van leerlingen moeten kunnen volgen op basis van (analyses) van toetsresultaten. Een verdere concretisering van wat dit vraagt van het handelen van docenten is (vooralsnog) niet uitgewerkt.

Zowel de OnderwijsRaad als de Commissie Kwaliteit Schoolexaminering constateerden in 2018 echter dat deze toetsbekwaamheid vaak in onvoldoende mate aanwezig is. Daarnaast constateerden zij dat, met uitzondering van het hoger beroepsonderwijs, *concrete* professionele standaarden voor toetsen voor docenten, leidinggevend en docentenop-leiders ontbreken en dat binnen docentenopleidingen te weinig aandacht besteed wordt aan toetsdeskundigheid. Ook constateerden zij dat er in alle onderwijssectoren onvoldoende (na)scholingsaanbod is dat is gericht op het vergroten van de toetsbekwaamheid. Ten aanzien van schoolexaminering stelde de Commissie Kwaliteit Schoolexaminering daarnaast vast dat er sprake is van kwetsbaarheden in de kwaliteitsborging van de schoolexamens. Naar aanleiding van deze bevindingen zijn door de VO Raad in 2018 vier

aanbevelingen geformuleerd, die in 2019 zijn uitgewerkt in negen actielijnen (VO Raad, 2019). Een van deze actiepunten gaat specifiek in op de toetsbekwaamheid:

“7. Elke school werkt ernaartoe dat er voldoende expertise op het terrein van examinering aanwezig is, bij de schoolleider, examensecretaris en onder docenten. Alle scholen hebben deze expertise in huis vanaf het schooljaar 2020-2021.”

VO-Raad (2019a)

Uit een monitoronderzoek naar de voortgang op de actielijnen (VO-Raad, 2019b; VO-Raad, 2020) blijkt dat er “nog behoorlijk wat winst” (VO-Raad, 2020, p. 2) te behalen is ten aanzien van de toetsbekwaamheid. In de monitor konden respondenten (290 schoolleiders in de monitor van 2019 en 262 schoolleiders in de monitor van 2020) voor de doelgroepen schoolleider, examensecretaris en docent aangeven of er nog ‘tamelijk veel inspanning’, ‘enigszins inspanning’ of ‘nauwelijks inspanning’ nodig is om te komen tot “voldoende expertise”. De respondenten waren in beide monitor-onderzoeken het meest positief over de expertise van schoolleider en de examensecretaris, al geven de resultaten gelijktijdig aan dat voor een grote groep geldt dat professionalisering noodzakelijk is: in 2020 gaven de respondenten aan dat 29% van de schoolleiders en 33% van de examensecretarissen zijn of haar expertise nauwelijks verder hoeft te ontwikkelen. De meeste ontwikkeling lijkt volgens de schoolleiders nog nodig bij de docenten: in 2019 gaven de respondenten aan dat voor ruim 60% van de docenten geldt dat enigszins inspanning geleverd moet worden om te komen tot voldoende expertise. Aanvullend gaven de respondenten in 2020 aan dat voor de groep docenten geldt dat slechts 4% de expertise nauwelijks hoeft te ontwikkelen en dat een derde van de groep tamelijk veel expertise te ontwikkelen heeft. Uiteraard moet bij deze resultaten de kritische vraag gesteld worden in hoeverre bij alle respondenten sprake is van een vergelijkbaar beeld van ‘voldoende expertise’ gezien de abstracte formulering van de bekwaamheidseisen hieromtrent.

In het hoger beroepsonderwijs is er de afgelopen jaren niet alleen veel aandacht geweest voor het belang van toetsbekwaamheid, maar specifiek ook voor uitwerken van het concept ‘voortdurende toetsbekwaamheid’ en het professionaliseren van docenten. Naar aanleiding van het rapport *Vreemde Ogen Dwingen* van de Commissie externe validering examenkwaliteit hoger beroepsonderwijs (2012) zijn in 2013 specifieke leeruitkomsten voor toetsbekwaamheid van hbo-docenten op basis- en seniorniveau (Expertgroep bke/ske, 2013), de zogenoemde basis- en seniorkwalificatie examinering (bke/ske), geformuleerd. Deze maken onderdeel uit van het ‘protocol Basiskwalificatie Didactische Bekwaamheid (BDB)’ (Vereniging Hogescholen, 2013) dat alle bij de Vereniging Hogescholen aangesloten hogescholen hebben ondertekend. In dit protocol is opgenomen dat alle docenten in het hbo vanaf januari 2015 een BDB-traject aan moeten bieden voor alle

startende docenten en vanaf medio 2015 na- of bijscholing moeten aanbieden voor alle zittende docenten op de onderdelen uit de BDB. In de praktijk heeft deze afspraak ertoe geleid dat vrijwel alle hogescholen zogenoemde bke/ske-professionaliserings- en certificeringstrajecten hebben ingericht. Onderzoek naar de implementatie van deze trajecten (Sluijsmans et al., 2017) laat zien dat de leeruitkomsten bke/ske hebben geholpen om de toetsbekwaamheid van startende en zittende hbo-docenten bespreekbaar te maken en te vergroten, en om de toetsbekwaamheid op vergelijkbare wijze te toetsen. Ook blijkt uit het onderzoek dat zowel ervaren als startende hbo-docenten enthousiast zijn over de inhoud van de bke/ske en het samenwerken met collega's op het gebied van toetsing (Sluijsmans et al., 2017).

Recentelijk is als onderdeel van het project Je Ogen Uitkijken, dat door het netwerk Toetsbekwaamheid op verzoek van de Vereniging Hogescholen wordt uitgevoerd, een literatuuranalyse uitgevoerd naar het concept 'voortdurende toetsbekwaamheid voor h(b)o-docenten'. Uit deze analyse (waarvan de opbrengsten uitgebreid beschreven zijn in Meijer et al., 2020), komt naar voren dat de taken van de h(b)o-docent op het gebied van toetsing meer omvatten dan het afnemen en beoordelen van toetsen. In het h(b)o is de docent ook degene die de toets ontwerpt, die verantwoordelijk is voor de alignment met de leeruitkomsten zowel in de toets als in het onderwijs, en die verantwoordelijk is voor het betekenis geven aan de analyses die na afname van de toetsen worden uitgevoerd ten behoeve van verbeteringen van de toets en het onderwijs.

Voor veel docenten in het h(b)o geldt dat zij geen docentenopleiding hebben gevolgd. Aandacht voor deze vaardigheden in de basisdidactische bekwaamheid (hbo) en de basiskwalificatie onderwijs (wo) is dan ook essentieel. Tegelijk zijn die vereiste kennis en vaardigheden veelomvattend: het gaat om zowel de conceptuele kennis (Pastore & Andrade, 2019; Xu & Brown, 2016) als de onderliggende didactische kennis (Herppich et al., 2018). Belangrijk daarbij is het besef dat het didactisch referentiekader waarover docenten (zonder docentenopleiding) beschikken en waarop zij de toetsing baseren voor een aannemelijk deel gebaseerd op hun eigen ervaringen als leerling (Meijer et al., 2020). Dit maakt het voor de ontwikkeling van toetsbekwaamheid belangrijk dat informatie over toetsing beschikbaar is, een docent ervaring opdoet met andere en nieuwe vormen van toetsing en dat er mogelijkheden zijn om nieuwe toetsen uit te proberen. Bovendien blijkt uit de literatuurverkenning die is uitgevoerd dat het voortdurend toetsbekwaam blijven aanvullende kennis en vaardigheden vraagt van docenten dan het toetsbekwaam worden. Bij voortdurende toetsbekwaamheid gaat het naast de kennis van toetsing, de vaardigheden om de toetsing vorm te geven, en de vaardigheden om dit binnen een specifieke context te kunnen doen, ook om het bewustzijn van de eigen toetsidentiteit en toetsconcepties: hoe sta ik zelf in toetsing, wat is mijn visie en wat zijn mijn overwegingen om de toetsing vorm te geven zoals ik dat doe? Daarnaast moeten docenten in staat zijn toetsing

en onderwijs te kunnen integreren, de sociale en emotionele aspecten van beoordelen beheersen (waaronder aandacht voor de eigen rol als examiner, de ethische aspecten, en de impact van toetsen), en moeten zijn het vermogen hebben zich, op basis van reflecties op het eigen handelen, voortdurend te blijven ontwikkelen om met toetsing aan te blijven sluiten bij zowel het beroep als bij de didactiek (Van Schilt-Mol et al., 2020).

Mede op basis van de analyse zijn de bke/ske-leeruitkomsten uit 2013 samen met vertegenwoordigers van 20 hogescholen herijkt (Van Schilt-Mol et al., 2020) en hertaald naar zogenoemde kwaliteiten van de toetsbekwame h(b)o-docent (Van Schilt-Mol et al., 2020). Binnen deze kwaliteiten wordt toetsbekwaamheid, in het licht van bovenstaande, enerzijds beschouwd als de mix van kennis en vaardigheden die nodig is om kwalitatief goede toetsen te kunnen maken en de mix van kennis en vaardigheden die nodig is om op basis van de informatie uit de toetsen zorgvuldige en betrouwbare beslissingen over (het leren van) leerlingen te kunnen nemen (bijv. Brown, 2008; Popham, 2009; Price et al., 2012; Expertgroep bke/ske, 2013). Anderzijds wordt toetsbekwaamheid beschouwd als het vermogen van docenten om onderwijs en toetsing op goede wijze te integreren (realiseren van constructieve alignment), het vermogen om oog te hebben voor de sociale en emotionele aspecten van toetsen, en het bewustzijn van de eigen toetsidentiteit en toetsconcepties (en de invloed daarvan op het eigen handelen binnen toetsing) (Meijer et al., 2020).

Dit heeft geleid tot de formulering van vier kwaliteiten die in 'Kwaliteiten van de toetsbekwame hbo-docent' (Van Schilt-Mol et al., 2020) verder zijn uitgewerkt:

1. je handelt binnen de context van het toetsbeleid en de toetsorganisatie
2. je handelt binnen het onderwijs- en toetsprogramma
3. je past toetskennis en - vaardigheden adequaat toe
4. je blijft toetsbekwaam

In de uitwerking van de kwaliteiten is beschreven wat een toetsbekwame docent weet, doet en kan. Het is aan de opleidingen zelf om binnen een docententeam het gesprek aan te gaan over de huidige en gewenste toetsbekwaamheid van individuele docenten, en om deze te 'vertalen' naar de context van de betreffende docent en naar een passende wijze voor het aantonen van de bekwaamheid. Daarbij staan drie vragen centraal (Van Schilt-Mol et al., 2020):

1. Wat zijn, gezien de rollen, taken en verantwoordelijkheden van de docent, bekwaamheden die relevant zijn voor deze specifieke toetsbekwame onderwijsprofessional?
2. Waar staat de docent, waar zou hij naar toe willen/moeten en hoe kan/wil hij dat bereiken?

3. Op welke wijze kan de docent invulling geven aan het 'voortdurend toetsbekwaam' worden en blijven?

Een mooie ontwikkeling is dat ook docentenopleidingen zich steeds vaker laten inspireren door de bekwaamheidseisen die worden gesteld aan docenten in het hbo. Zo is het onderwijsaanbod van bijvoorbeeld de docentenopleiding Gezondheidszorg en Welzijn (HAN, 2020) en van de master Ontwerpen van Eigentijds Leren zo omvattend als het gaat om het onderwerp toetsen, dat de studenten na afronding van deze opleidingen niet alleen respectievelijk een bachelor- en masterdiploma ontvangen, maar eveneens een certificaat 'Basiskwalificatie Examinering'.

Het spreekt voor zich dat het lectoraat zich de komende jaren onder meer zal gaan richten op wat eigentijds beoordelen en beslissen vraagt aan specifieke bekwaamheden van docenten in de verschillende onderwijssectoren en een bijdrage zal leveren aan de concretisering van deze vaardigheden, bijvoorbeeld ten behoeve van het optimaliseren van de curricula van docentenopleidingen.

BALANS TUSSEN FUNCTIES: TOETSSYSTEMEN

Voordat ingegaan wordt op *hoe* vanuit de functies van toetsing invulling gegeven kan worden aan eigentijds beoordelen en beslissen, wordt eerst kort stilgestaan bij de *haalbaarheid* van eigentijds beoordelen en beslissen.

Uit de gesprekken die ter voorbereiding op het schrijven van de aanvraag voor het lectoraat zijn gevoerd, kwam duidelijk een soort 'koud watervrees' naar voren: docenten hebben vaak ten onrechte de indruk dat meer eigentijds beoordelen en beslissen betekent dat het onderwijs volledig op de schop moet, dat er heel veel professionalisering nodig is, en dat aanpassing van de organisatie van het onderwijs veel aandacht zal vragen. Uiteraard kán er voor gekozen worden om het onderwijsprogramma te vernieuwen en eigentijds beoordelen en beslissen te integreren in een setting waarin bijvoorbeeld gewerkt wordt vanuit de principes van programmatisch toetsen of High Impact Learning (HILL, Dochy & Segers, 2018). Meer eigentijds beoordelen en beslissen is echter óók mogelijk in een meer traditionele onderwijssetting waarin met regelmaat toetsen met een summatieve functie worden afgenomen. Bijvoorbeeld door niet langer zonder meer uit te gaan van een rekenkundig gemiddelde van alle afgenomen toetsen bij het nemen van een overgangsbepaling, of bijvoorbeeld door alle beoordelingen niet alleen van een cijfer, maar ook van een narratieve toelichting te voorzien.

Met deze rede beogen we vanuit het lectoraat docenten er enerzijds van te overtuigen dat het ook met hele kleine aanpassingen al mogelijk is meer eigentijds te beoordelen en

beslissen, en anderzijds dat dat mogelijk is binnen de bestaande onderwijs- en toetsystemen. Het gaat daarbij om het zoeken naar de juiste balans tussen de drie functies van toetsen. Daarbij wordt 'wat juist is' uiteraard bepaald door de context van de opleiding/school: wat is de visie van de opleiding/school op onderwijs en wat betekent dit voor keuzes die gemaakt worden omtrent (het vormgeven, volgen en vaststellen van) het leren en het ontwikkelen van leerlingen?

Het is het zoeken van deze balans tussen de drie functies, en het benutten van de meerwaarde van de verschillende functies in hun onderlinge samenhang, dat de uitdaging vormt (Waitling & Ginsburg, 2019). Gebaseerd op een beschrijving van Allal (2011; verder uitgewerkt in Laveault & Allal, 2016; Sluijsmans, 2020), kunnen er drie onderwijs- en toetsscenario's onderscheiden worden ten aanzien van de wijze waarop deze functies tot elkaar in verhouding staan: (1) het scenario waarin het vaststellen van wat geleerd is (onderdeel van de beslisfunctie) leidend is voor het ontwerp van het onderwijs- en toetsprogramma, (2) het scenario waarin de leerfunctie én de beslisfunctie als twee separate aspecten leidend zijn voor het ontwerp van het onderwijs- en toetsprogramma, en (3) het scenario waarin de leer- en beslisfunctie met elkaar verweven zijn. In deze scenario's is de evaluatiefunctie niet expliciet meegenomen. De drie scenario's zouden, wanneer de evaluatiefunctie daaraan wordt toegevoegd, als volgt visueel weergegeven kunnen worden:

Scenario

Visuele weergave

Scenario 1:

het vaststellen van wat geleerd is (onderdeel van de beslisfunctie) is leidend bij het ontwerpen van het onderwijs- en toetsprogramma



Scenario 2:

de leerfunctie én de beslisfunctie zijn als twee separate aspecten leidend bij het ontwerpen van het onderwijs- en toetsprogramma en worden beide gevoed door de evaluatiefunctie



Scenario 3:
de leer-, beslis-, en evaluatie-
functie zijn met elkaar
verweven



Figuur 10. Balans tussen leer-, beslis- en evaluatiefunctie (geïnspireerd op Allal, 2011; Sluismans, 2020).

In bovenstaande figuur wordt gesproken over 'leren/leeractiviteiten', 'beslissing', 'evalueren'. Met leren/leeractiviteiten wordt geduid op *alle* activiteiten die kunnen worden ingezet om de ontwikkeling en het niveau van de leerling te beoordelen. Dit kunnen zowel formele toetsen zijn als (in)formele activiteiten als mentorgesprekken, formatieve leeractiviteiten, lesopdrachten, beroepstaken, 360-gradenbeoordelingen, etc. Deze activiteiten zijn, in de terminologie van programmatisch toetsen, zogenoemde 'datapunten': betekenisvolle activiteiten die informatie geven over waar de leerling staat en welke stappen nog gezet moeten worden om het gewenste niveau te realiseren. Met beslissing wordt geduid op de consequenties die verbonden worden aan de beoordelingen (die gebaseerd zijn op de leeractiviteiten). Deze beslissingen kunnen gericht zijn op het bijsturen van het onderwijs en het onderwijsleerproces, en op het vaststellen van wat er geleerd is. Met evalueren wordt hier geduid op het proces van analyseren en interpreteren van de opbrengsten van de beoordelingen en de beslissingen in relatie tot het eigen handelen en het gegeven onderwijs, en het verbinden van consequenties hieraan voor het eigen handelen binnen en/of de inrichting en uitvoering van het beoordelingsproces, het onderwijs, het beslisproces of de toetsing.

In ieder van de hierboven genoemde scenario's is meer eigentijds beoordelen en beslissen mogelijk. Hoewel het in het tweede en derde scenario wellicht meer voor de hand ligt hoe dat kan, kan het uitdrukkelijk ook in het eerste scenario. Van het eerste scenario is veelal sprake in de meer traditionele en/of op beheersingsleren ingerichte onderwijscontexten. In dit scenario wordt iedere onderwijsperiode afgesloten met een of meerdere toetsen met summatieve functie. Met deze toets(en) wordt vastgesteld of de leerling de stof voldoende beheerst en wordt, met andere woorden, vastgesteld wat is geleerd (beslisfunctie). Dit wordt ook wel aangeduid als assessment of learning. De informatie uit de toets(en) wordt - wanneer in dit scenario invulling gegeven wordt aan eigentijds beoordelen en beslissen - gebruikt om de kwaliteit van het onderwijs in de volgende onderwijsperiode te optimaliseren (evaluatiefunctie) en waar mogelijk ook om het (verder) leren van de leerlingen te stimuleren en te sturen (leerfunctie). De informatie wordt dus, met andere woorden, gebruikt om het onderwijs- en toetsprogramma zowel inhoudelijk

als qua uitvoering bij te stellen en verder vorm te geven zodanig dat de leerling de ruimte krijgt zich gericht verder te ontwikkelen.

Binnen dit scenario is het nog vaak gebruikelijk dat de leerling ook als de toets(en) aan het eind van de periode met een onvoldoende resultaat zijn afgerond, toch start met een nieuwe onderwijsperiode. Vaak betekent dit voor de leerling dat er ofwel een herkansing wordt gepland, ofwel dat de leerling de behaalde onvoldoende zal moeten compenseren met een hoger cijfer in de volgende onderwijsperiode, ook als in deze periode inhoudelijk andere onderwerpen worden behandeld. In een onderwijsomgeving waar meer eigentijds wordt beoordeeld en beslist, worden de behaalde resultaten van de leerling altijd voorzien van een narratieve toelichting die de leerling informatie geeft over de stappen die hij nog moet zetten om het gewenste niveau te behalen en krijgt de leerling de ruimte en mogelijkheden hier iets mee te doen. Daarnaast worden heel bewust afwegingen gemaakt over hoe bij het nemen van de zwaarwegende beslissingen aan bijvoorbeeld het eind van het schooljaar wordt omgegaan met alle behaalde resultaten. Er wordt niet op basis van een rekenkundige formule een beslissing genomen, maar er wordt inhoudelijk gekeken en beargumenteerd of de leerling zich voldoende ontwikkeld heeft: heeft de leerling bij de herkansing laten zien de stof wel te beheersen? Beheerst de leerling alle leerdoelen die centraal stonden? En, als dat nog niet helemaal het geval is, is het volledig opnieuw volgen van een studiejaar/schooljaar nodig, of kan de leerling met gerichte remediëring voor een specifiek vak toch al starten in het volgende jaar.

In het tweede scenario vormt niet alleen beslisfunctie het uitgangspunt voor het inrichten van het onderwijs, maar vormen de leer- en beslisfunctie gezamenlijk als twee separate aspecten het uitgangspunt. Zij worden daarbij beide gevoed door een voortdurend proces van evaluatie. Een voorbeeld waar dit scenario is toenemende mate zichtbaar is, is het voortgezet onderwijs. Docenten raken daar steeds meer doordrongen van de meerwaarde van de leerfunctie van toetsactiviteiten en zetten deze niet alleen in ter afsluiting van het leerproces en onderwijs, maar juist ook ten dienste hiervan. Dit betekent concreet dat in dit scenario toets- en leeractiviteiten bewust worden ingezet ter voorbereiding op een formele toets die aan het eind van een onderwijsperiode wordt afgenomen. Op basis van de informatie uit deze activiteiten wordt het onderwijs voortdurend bijgesteld en verbeterd, zodanig dat leerlingen optimaal kunnen werken aan de ontwikkeling van de leerdoelen. De informatie uit de activiteiten én de afsluitende formele toets op basis waarvan het leren wordt vastgesteld (beslisfunctie) worden zowel gedurende als na de onderwijsperiode ingezet om zowel het (verder) leren van leerlingen te stimuleren en sturen (leerfunctie) als de kwaliteit van het onderwijs te optimaliseren (evaluatiefunctie). In dit scenario is het goed zichtbaar dat toetsen nooit alléén een formatieve of alléén een summatieve functie hebben: zo wordt ook de toets waarmee de onderwijsperiode wordt

afgesloten voorzien van narratieve feedback en feedforward die zowel gericht kan zijn op inhoudelijke aspecten, als op specifieke vaardigheden als bijvoorbeeld zelfregulatie, leerstrategieën of samenwerken, die de leerling kan gebruiken om zich verdere te ontwikkelen in de volgende onderwijsperiode(s). Het promotieonderzoek van Janneke van der Steen beoogt om docenten meer inzicht te geven in hoe zij dergelijke toets- en leeractiviteiten op een effectieve manier in samenhang kunnen ontwerpen en inzetten, zodanig dat de leerling optimaal wordt voorbereid op de afsluitende toets en dat de resultaten op deze toets geen verrassing maar een bevestiging zijn van het beeld dat zowel de docent als de leerling van de ontwikkeling en beheersing van de leerling heeft. De eerste resultaten, ontwerpprincipes om deze activiteiten in samenhang te ontwerpen, zijn al gepresenteerd tijdens een inspiratiemiddag die als onderdeel van het promotietraject van Janneke worden georganiseerd (Van der Steen, 2020) en zullen op korte termijn ook in zowel een wetenschappelijke publicatie als vakblad worden gepubliceerd.

Bij het laatste scenario zijn de leerfunctie en de beslisfunctie met elkaar verweven. Zwaarwegende beslissingen waarbij het leren wordt vastgesteld worden *niet* genomen op basis van een enkele toets aan het einde van een onderwijsperiode, maar worden onderbouwd op basis van informatie die verkregen is uit meerdere toets- en leeractiviteiten. Dit betekent met andere woorden dat de informatie uit deze activiteiten zowel gebruikt wordt om het leren te sturen en te stimuleren (leerfunctie) als om beslissingen te onderbouwen (beslisfunctie). Daarnaast wordt deze informatie, gedurende het onderwijs, steeds gebruikt om de kwaliteit van het onderwijs te optimaliseren (evaluatiefunctie). In onderwijscontexten die zijn vormgegeven volgens de principes van programmatisch toetsen is sprake van dit derde scenario. Belangrijk bij dit scenario is dat de toets- en leeractiviteiten altijd gericht zijn op het beoordelen – het verkrijgen van informatie over waar de leerling staat – en het geven van feedback om vervolgstappen te kunnen zetten. Dat betekent dat op basis van een enkele beoordeling nooit een zwaarwegende beslissing wordt genomen. De leerling krijgt zo écht de ruimte om fouten te maken en de ruimte om zich te ontwikkelen. Er wordt pas een zwaarwegende beslissing genomen als er sprake is van ‘saturatie’: dit is het geval als er op basis van alle verzamelde beoordelingen een helder beeld is over de beheersing van de leerling en als een extra activiteit hier geen nieuwe informatie meer aan toevoegt. Belangrijk in dit scenario is dat de zwaarwegende beslissingen worden onderbouwd vanuit de beoordelingen: er wordt dus geen formule losgelaten op de beoordelingen om tot een beslissing te komen, maar de beoordelingen worden benut om de beslissing narratief en inhoudelijk te kunnen onderbouwen.

Samenvattend kunnen de scenario's vanuit het perspectief eigentijds beoordelen en beslissen als volgt worden omschreven.



Aan het einde van een onderwijsperiode wordt een formele toets afgenomen die de basis vormt voor een bewuste en onderbouwde zwaarwegende beslissing. De informatie uit de toets wordt daarnaast gebruikt om de kwaliteit van het onderwijs in de volgende onderwijsperiode te optimaliseren en waar mogelijk om het (verder) leren van leerlingen te stimuleren en sturen.



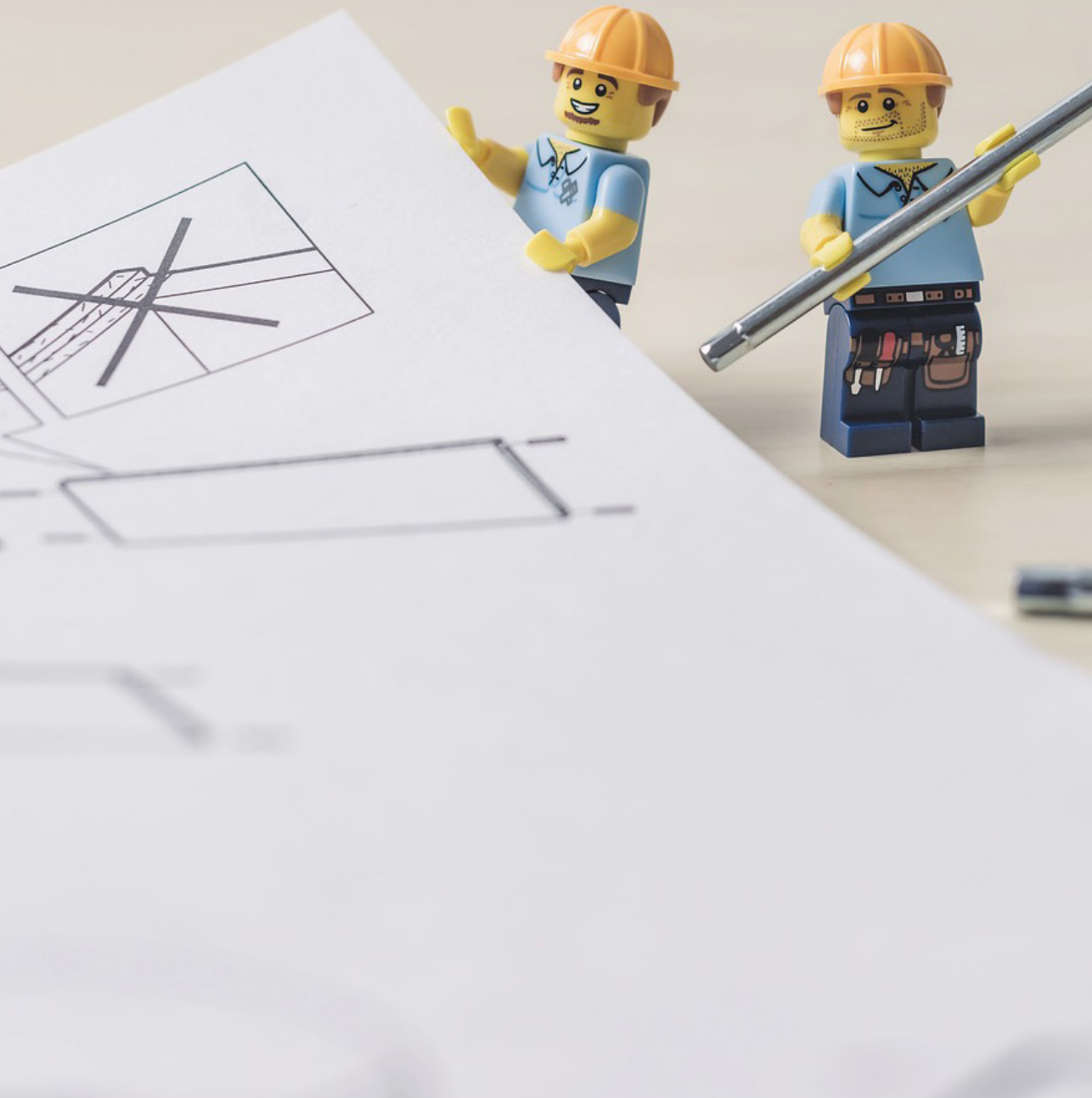
Gedurende een onderwijsperiode worden (toets-/leer)activiteiten bewust ingezet ter voorbereiding op de formele toets die aan het eind van de onderwijsperiode wordt afgenomen en die de basis vormt voor een bewuste en onderbouwde zwaarwegende beslissing. De informatie uit de activiteiten

wordt gedurende de onderwijsperiode ingezet om beslissingen te kunnen nemen over de ontwikkeling en voortgang van de leerling zodat zo nodig bijgestuurd kan worden op het onderwijs(leersproces). De informatie uit de activiteiten én de formele toets worden ook ná de onderwijsperiode ingezet om zowel het (verder) leren van leerlingen te stimuleren en sturen als de kwaliteit van het onderwijs te optimaliseren.



Informatie uit de (toets-/leer)activiteiten wordt enerzijds gebruikt om het leren te sturen en te stimuleren, anderzijds om beslissingen te nemen. Op basis van informatie uit één of enkele activiteiten worden uitsluitend beslissingen genomen met een beperkt belang voor

de leerling (beslissingen over voortgang, noodzakelijke remediëring, etc). Zwaarwegende beslissingen worden genomen op basis van een veelheid van over een langere onderwijsperiode verzamelde informatie (zodat de leerling de tijd heeft zich te ontwikkelen en er een goed beeld ontstaat van de beheersing van de beoogde vaardigheden). Informatie uit (toets-/leer)activiteiten wordt steeds gebruikt om de kwaliteit van het onderwijs te optimaliseren.



6. EIGENTIJDS BEOORDELEN EN BESLISSEN IN DE PRAKTIJK: PRAKTISCHE HANDVATTEN

Na de theoretische verkenning van wat eigentijds beoordelen en beslissen is en een beschouwing op de haalbaarheid hiervan in de verschillende onderwijs- en toetsscenario's, is de vervolgvraag uiteraard hoe daadwerkelijk invulling gegeven kan worden aan eigentijds beoordelen en beslissen. De brede verkenning van het onderscheid tussen beoordelen en beslissen, de drie functies van toetsen (de leerfunctie, de beslisfunctie en de evaluatiefunctie) en het vraagstuk over het vinden van de juiste balans tussen deze functies passend bij de visie van een opleiding/school op onderwijs, leiden 'als vanzelf' naar een schets van hoe eigentijds beoordelen en beslissen er in de praktijk uit kan zien.

Eigentijds beoordelen en beslissen in de praktijk

Bij eigentijds beoordelen en beslissen is de visie op onderwijs leidend bij de ontwerpkeuzes die gemaakt worden bij de inrichting van het onderwijs en de toetsing. Het gaat daarbij om geïnformeerde en onderbouwde keuzes. Toetsen worden weloverwogen ingezet als integraal onderdeel van het onderwijs: er wordt daarbij bewust onderscheid gemaakt tussen beoordelen en beslissen. Er wordt gericht informatie verzameld om zorgvuldige beslissingen te kunnen nemen over het leren van leerlingen. Hierdoor is er als vanzelfsprekend sprake van constructieve afstemming tussen de leerdoelen, leeractiviteiten en toetsactiviteiten. Beoordelingen worden altijd voorzien van rijke informatie: dit geeft de leerlingen handvatten om invulling te geven aan hun leerproces. Doordat de leerlingen de waarde van deze informatie inzien, nemen zij zelf (steeds meer) regie over hun ontwikkeling. De informatierijke beoordelingen bieden ook de docenten handvatten om richting en sturing te geven aan het leerproces. Daarnaast vormen de rijke beoordelingen de basis voor betrouwbare en zorgvuldige beslissingen over (het leren van) leerlingen. De informatie uit toetsen wordt altijd gebruikt om kritisch te reflecteren op het eigen handelen en om de kwaliteit van het onderwijs te optimaliseren.

Op basis van deze schets hebben we binnen het lectoraat een zestal handvatten geformuleerd die kunnen helpen hier invulling aan te geven. Het spreekt voor zich dat het lectoraat zich als doel stelt deze handvatten verder te ontwikkelen en rijke voorbeelden te verzamelen die inzicht geven in hoe de handvatten in de praktijk vorm kunnen krijgen. Op die manier zullen we, samen met docenten en leerlingen, kennis ontwikkelen over het ontwerpen en uitvoeren van eigentijds beoordelen en beslissen en over de werkende principes hierbij.

Handvat 1. Maak bewuste en geïnformeerde ontwerpkeuzes

De visie op onderwijs vormt de basis voor het ontwerp van het onderwijs én de toetsing daarbinnen. Maak bewuste en beargumenteerde ontwerpkeuzes over de inrichting van het toetsstelsel, waarbij je je laat informeren door actuele kennis en inzichten (evidence-informed). Belangrijk is dat je met elkaar in gesprek blijft over waarom je doet wat je doet.

Voer bij het vormgeven van dit handvat met elkaar het gesprek:

- *Wat is onze visie op onderwijs (en daarmee op opleiden en leren)?*
- *Welke functies van toetsen onderscheiden wij?*
- *Wat functie vinden wij essentieel?*
- *Wat betekent dit voor de (ontwerp)keuzes die we maken?*
- *Door welke reeds ontwikkelde kennis/ervaringen kunnen we ons daarbij laten inspireren?*
- ...

Handvat 2. Maak onderscheid tussen beoordelen en beslissen

Wees je ervan bewust dat beoordelen en beslissen twee verschillende activiteiten zijn die elkaar opvolgen. Bij beoordelen stel je vast waar de leerling staat ten opzichte van zijn leerdoelen, bij het beslissen verbind je hier vervolgens een consequentie aan. Beslissingen kunnen betrekking hebben op het (bij)sturen van het onderwijs(leer)proces, op de voortgang van de ontwikkeling van de leerling (diagnose en voorspelling) en op het vaststellen van wat geleerd is (certificeren, kwalificeren, selecteren, etc.). Afhankelijk van het type beslissing kan deze gebaseerd zijn op één beoordeling, maar ook op meerdere beoordelingen.

Voer bij het vormgeven van dit handvat met elkaar het gesprek:

- *Hoe geven wij invulling aan beoordelen en beslissen?*
- *Welke beslissingen nemen wij en op welke manier?*
- ...

Handvat 3. Verzamel informatie ten behoeve van de beslissing die je wil nemen

Ontwikkel je onderwijs vanuit de beslissing die je wil nemen over leerlingen. Stel jezelf bij het ontwerpen steeds de vraag welke informatie je nodig hebt om de beslissing die je wil nemen betrouwbaar en zorgvuldig te kunnen nemen. Dit wordt ook wel aangeduid als decision-driven data-collection. Meer concreet: welke activiteiten moet je inzetten om de juiste informatie te verzamelen zodat je kunt bepalen of vooraf vastgestelde doelen worden en/of zijn behaald? Streef daarbij bewust naar constructieve afstemming (constructive alignment): (leer)doelen, leeractiviteiten en toetsactiviteiten moeten logisch met elkaar samenhangen.

Voer bij het vormgeven van dit handvat met elkaar het gesprek:

- *Welke beslissingen nemen wij op welke momenten?*
- *En waarover nemen we dan een beslissing?*
- *Hoe stellen we de leerling in staat aan te tonen dat hij zich ontwikkelt en dat hij de beoogde leeropbrengsten beheerst?*
- *Welke informatie hebben we nodig om deze beslissing te kunnen nemen?*
- *Hoe komen we tot een zorgvuldig samengesteld toetsprogramma met een bewuste en beargumenteerde keuze van toetsvormen?*
- ...

Handvat 4. Ga in gesprek met de leerling over de waarde van beoordelingen

Het is belangrijk dat de leerling de waarde van informatie uit beoordelingen inziet: dat zorgt ervoor dat hij deze zal accepteren, actief aan de slag zal gaan met deze informatie én ook actief op zoek zal gaan naar extra informatie. Blijf voortdurend met de leerling in gesprek over hoe de leerling informatie uit de beoordelingen kan gebruiken ten behoeve van zijn ontwikkeling. Feedback krijgen is immers geen garantie voor het gebruiken hiervan. Zorg dat de leerling houvast ervaart in zijn leerproces en stimuleer hem om geleidelijk steeds meer regie te nemen over (het gebruik van informatie ten behoeve van) zijn ontwikkeling.

Voer bij het vormgeven van dit handvat met elkaar het gesprek:

- *Hoe willen wij, vanuit onze visie op onderwijs, invulling geven aan het leerproces van de leerling?*
- *Wat heeft de leerling nodig om regie te kunnen nemen over zijn ontwikkeling?*
- *Wat vraagt het van ons als docenten om de leerlingen in hun kracht te zetten?*
- ...

Handvat 5. Geef en benut informatierijke beoordelingen

Een beoordeling krijgt pas een waarde als de leerling op basis van de beoordeling richting kan geven aan zijn verdere leerproces. Het geven van alleen een cijfer of de indicatie (on)voldoende geeft niet de informatie waarmee de leerling dit kan doen. Voorzie iedere beoordeling daarom van een narratieve onderbouwing. Maak ook bij het nemen van een beslissing een weloverwogen inhoudelijke afweging. Baseer je beslissing, met andere woorden, niet op een rekenkundig gemiddelde, maar op de rijke informatie uit de beoordelingen.

Voer bij het vormgeven van dit handvat met elkaar het gesprek:

- *Hebben we allemaal hetzelfde beeld bij het beroepsprofiel/beoogde leerresultaten?*
- *Hoe aggregeren wij de informatie zodanig dat die informatie geeft over de beoogde leerresultaten?*
- *Hoe geven wij beoordelingen die én richting kunnen geven aan het leren én benut kunnen worden bij het onderbouwen van een beslissing?*
- *Hoe borgen wij de kwaliteit van onze beoordelingen en beslissingen?*
- ...

Handvat 6. Blijf bewust toetsbekwaam

De kwaliteit van een beoordeling en een beslissing is niet afhankelijk van de (afzonderlijke) toetsactiviteit(en), maar van degene die de beoordeling geeft en beslissing neemt. Ben je ervan bewust dat toetsbekwaamheid expliciet onderdeel uitmaakt van je bekwaamheid als docent. Als toetsbekwame docent heb je niet alleen kennis van toetsing en vaardigheden om toetsing als onderdeel van onderwijs vorm te geven, maar ben je je ook bewust van de impact van toetsen op (het leren van) leerlingen en heb je de bereidheid en vaardigheid om te reflecteren op je eigen handelen, en op basis daarvan het onderwijs en de toetsing daarbinnen te optimaliseren.

Voer bij het vormgeven van dit handvat met elkaar het gesprek:

- *Welke bekwaamheid verwachten we van wie?*
- *Hoe gaan we om met toetsbekwaam worden en blijven?*
- *Hoe benutten we bewust de reflectie op ons eigen handelen om het onderwijs te verbeteren?*
- ...

6. Eigentijds beoordelen en beslissen in de praktijk: praktische handvatten



7. ONDERZOEKSAGENDA VAN HET LECTORAAT

In aanloop naar de omzetting van het associate lectoraat Toetsen en Beslissen naar het lectoraat Eigentijds Beoordelen en Beslissen heb ik vele rondetafelgesprekken gevoerd met docenten, met toetsexperts en met studenten. Samen met hen heb ik verkend welke actuele vraagstukken er leven op het gebied van eigentijds beoordelen en beslissen.

Uit deze gesprekken kwamen een aantal belangrijke aandachtspunten naar voren. Op de eerste plaats kwam de behoefte naar voren om binnen het toetsen, beoordelen en beslissen veel meer expliciete aandacht te hebben voor *de leerfunctie van toetsen*. Do-centprofessionalisering op het gebied van toetsing richt zich, gebaseerd op de ervaring van de docenten, veelal op het ontwikkelen van toetsen op basis waarvan zwaarwegende beslissingen genomen worden zoals het toekennen van studiepunten of een diploma. Ook bij het ontwerpen van een toetsprogramma gaat volgens de docenten de meeste aandacht uit toetsen met een summatieve functie.

Dit linkt aan het tweede aandachtspunt: behoefte aan meer aandacht voor en erkenning van het *dilemma* dat veel docenten ervaren. Docenten voelen in de dagelijkse praktijk een frictie tussen enerzijds de focus op ontwikkeling en anderzijds op verantwoording van gerealiseerd niveau. Docenten voelen immers niet alleen een verantwoordelijkheid naar hun leerlingen, maar ook naar de vervolgopleiding en/of het beroep waartoe zij opleiden en de maatschappij: de maatschappelijke waarde van bijvoorbeeld diploma's, en nu in deze (corona-)tijd ook examens, staat niet voor niets met regelmaat ter discussie.

Een derde aandachtspunt dat ter sprake kwam, was het *kunnen en durven nemen van keuzes* ten aanzien van de inrichting de toetsing. Uit de gesprekken kwam naar voren dat docenten zich bij de inrichting van het toetsysteem vaak laten leiden door hun eigen ervaringen, en door 'wat gebruikelijk is binnen de opleiding/school'. Hoewel docenten op de hoogte zijn en blijven van actuele ontwikkelingen, en deze ook met interesse volgen, leiden deze vaak ook tot terughoudendheid: de indruk ontstaat dat het hele onderwijs op de schop moet, en dat er op zo veel aspecten ontwikkeling nodig is dat er al koudwater-vrees ontstaat zonder dat het water überhaupt in zicht is. Hieraan gerelateerd benoemden de gesprekspartners dat er een grote behoefte bestaat aan meer inzicht in werkende principes rondom toetsing, en meer specifiek ten aanzien van het beoordelen en beslissen. Wat werkt in welke context wel en niet, en waarom is dat zo? Daarbij werd de uitdrukkelijke wens uitgesproken voor praktische handreikingen en tools, die niet voorschrijvend

moeten zijn, maar handvatten bieden om binnen de eigen context bewuste (ontwerp) keuzes te kunnen maken.

Een vierde aandachtspunt, tot slot, is de behoefte aan meer *inzicht* in wat de verschillende functies van toetsing nu eigenlijk vragen van niet alleen de docenten, maar ook de leerlingen. Welke toetsbekwaamheid hebben leerlingen nodig om echt richting te kunnen geven aan hun eigen leren, om keuzes te kunnen maken over de volgende stappen in hun ontwikkeling? En daaraan gekoppeld: wat vraagt het van de schoolorganisatie om daadwerkelijk te kunnen doen aan alle drie de functies?

Op basis van de gesprekken is een onderzoeksprogramma geformuleerd voor het lectoraat Eigentijds Beoordelen en Beslissen. De focus daarbij is dat het onderzoeksprogramma zich richt op het ontwerpen, inrichten en uitvoeren van eigentijds beoordelen en beslissen. We hebben daarbij aandacht voor de verschillende onderwijssectoren, waarbij we steeds vanuit de specifieke onderwijs- en opleidingscontext met elkaar in gesprek zullen gaan over hoe vanuit de visie van de school of opleiding op onderwijs invulling gegeven kan worden aan eigentijds beoordelen en beslissen. Belangrijk uitgangspunt hierbij is de praktijkgerichte benadering: onderzoek naar wat werkt doen we samen met leerlingen, docenten, opleidingen en scholen.

De vraag die centraal staat in het lectoraat is: **hoe kan eigentijds beoordelen en beslissen zo worden ontworpen, ingericht en uitgevoerd worden dat het bijdraagt aan de leerfunctie, de beslisfunctie en de evaluatiefunctie van toetsing?**

Met andere woorden: hoe kunnen we ervoor zorgen dat toetsen niet alleen het leren van leerlingen stimuleren en sturen, maar dat de resultaten op deze toetsen ons helpen zorgvuldige en betrouwbare beslissingen te nemen over (het leren van) leerlingen en dat deze resultaten ons helpen de kwaliteit van het onderwijs te verbeteren? En hoe doen we dat specifiek binnen de verschillende onderwijs- en opleidingscontexten, én passend bij de eigen visie op onderwijs?

Om antwoord te kunnen geven op deze vraag wordt binnen het lectoraat onderzoek gedaan naar:

- de ontwerpeisen voor een toetsprogramma, zodanig dat dit programma niet alleen leidt tot het realiseren van de beoogde leerresultaten, maar ook bijdraagt aan de drie functies van toetsing,
- de voorwaarden die aan een schoolorganisatie gesteld moeten worden om eigentijds te beoordelen en beslissen,

- de (toets)bekwaamheid van docenten die nodig is om eigentijds beoordelen en beslissen te ontwerpen en uit te voeren, en
- de (toets)bekwaamheid van leerlingen die nodig is om met informatie uit het toetsprogramma daadwerkelijk invulling te kunnen geven aan het (nemen van regie over het) eigen leerproces.

Bij het onderzoeken hiervan richten we ons specifiek op het ontwikkelen van kennis hierover en het achterhalen van werkende principes. Op basis hiervan worden binnen het lectoraat praktische handvatten ontwikkeld die docenten kunnen toepassen in de dagelijkse onderwijspraktijk, en zullen handvatten ontwikkeld worden voor de vormgeving van professionaliseringstrajecten binnen scholen en (docenten)opleidingen om de toetsbekwaamheid van (toekomstige) docenten én leerlingen te versterken.

Enkele beoogde opbrengsten van het lectoraat zijn dan ook:

- Voorbeelden van de wijze waarop in verschillende onderwijssectoren en binnen verschillende toetssystemen invulling gegeven wordt aan de zes handvatten.
- Nadere uitwerking van de wijze waarop de drie functies van toetsen vorm kunnen krijgen in de verschillende onderwijssectoren, zodanig dat deze elkaar versterken.
- Handreikingen waarmee de ontwikkelde kennis ten aanzien van het ontwerpen, inrichten en uitvoeren van eigentijds beoordelen en beslissen toegepast kan worden binnen de eigen onderwijspraktijk.
- Handreikingen waarmee, expliciet vanuit het perspectief van eigentijds beoordelen en beslissen, en met oog voor de verschillende onderwijssectoren en toetssystemen, de kwaliteit van de vijf toetsentiteiten die een rol spelen bij toetsing kritisch bezien én geoptimaliseerd kan worden.

Vertaald naar de verschillende domeinen die onderscheiden worden (onderwijs, beroepspraktijk en kennisontwikkeling) betekent deze doelstelling het volgende:

Onderwijs van de HAN en de scholing van docenten van de HAN

- Opbrengsten van onderzoek worden gedeeld met de docentenopleidingen binnen de HAN, zodat hiervan gebruik gemaakt kan worden bij het onderwijs aan docenten in opleiding;
- Docenten in opleiding (studenten) worden waar mogelijk betrokken bij onderzoek;
- Opbrengsten van onderzoek worden vertaald naar praktische toepassingen en handreikingen die HAN-docenten kunnen gebruiken bij het ontwerpen en ontwikkelen van hun onderwijs;

- Opbrengsten van onderzoek worden gedeeld met de curriculumcommissie van de Basiskwalificatie Didactische Bekwaamheid, zodat deze opgenomen kunnen worden in het opleidingsaanbod voor startende HAN-docenten;
- HAN-opleidingen worden ondersteund bij het onderzoeken en verbeteren van de kwaliteit van toetsing;
- HAN-opleidingen worden ondersteund bij het ontwerpen, ontwikkelen en implementeren van eigentijdse vormen en invullingen van toetsing, en bij het onderzoeken van de effecten hiervan.

Beroepspraktijk en samenleving

- Het lectoraat vervult een leidende rol in het formuleren van bekwaamheidseisen voor docenten ten aanzien van toetsing en het onderzoeken van duurzame toetsbekwaamheid;
- Het lectoraat vervult een leidende rol bij het uitwisselen van kennis en ervaringen en het ontwikkelen van kennis van toetsing in het hoger onderwijs middels de coördinatie van het netwerk toetsbekwaamheid h(b)o (waarin vrijwel alle hogescholen en een zevental universiteiten participeren);
- Het lectoraat vervult een leidende rol bij het uitwisselen van kennis en ervaringen over programmatisch toetsen middels het leernetwerk programmatisch toetsen dat samen met het bijzonder lectoraat Beoordelvingsvraagstukken in het beroepsonderwijs wordt geleid;
- Het lectoraat levert een bijdrage aan de landelijke discussies ten aanzien van toetsingsvraagstukken door participatie in samenwerkingsverbanden als het netwerk formatief evalueren docentenopleidingen, de kwaliteitsgroep schoolexaminering vmbo en het platform leren van toetsen.

Kennisontwikkeling in het beroepsdomein

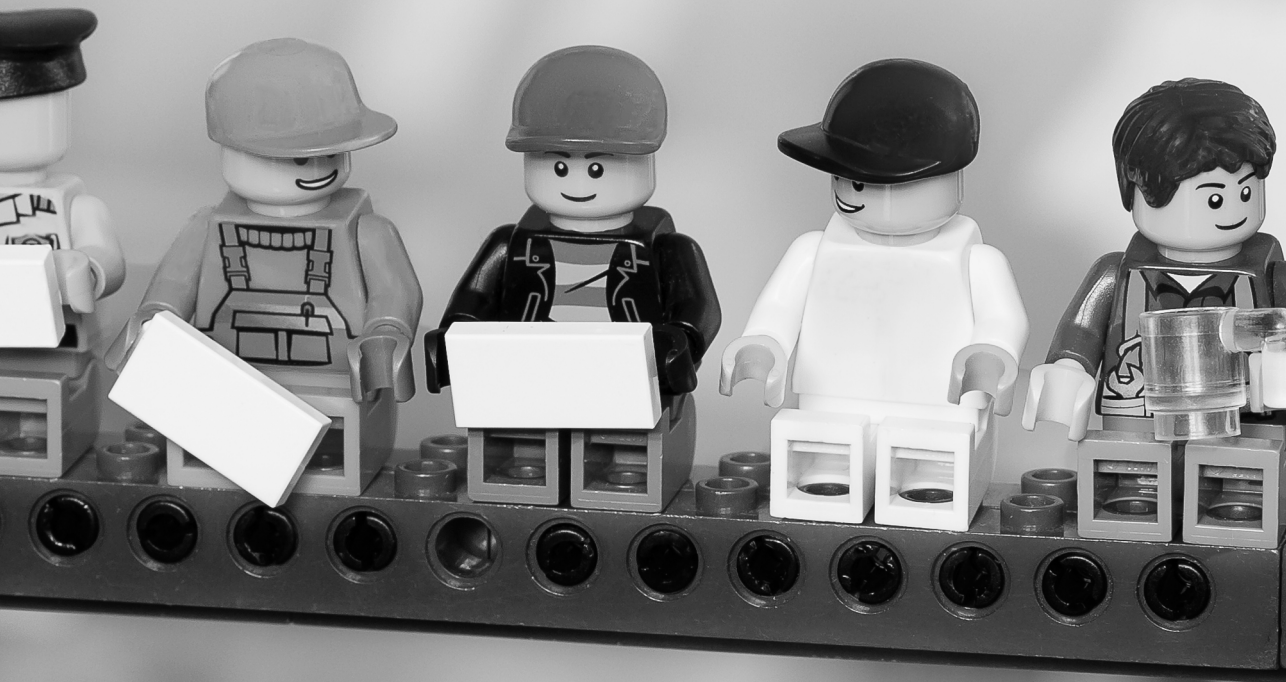
- Middels onderzoek wordt inzicht verkregen in de werkende factoren rondom beoordelen en beslissen in alle onderwijssectoren;
- Middels onderzoek worden de werkzame factoren onderbouwd, waarmee uitspraken mogelijk zijn over vereiste toetsbekwaamheid voor docenten én leerlingen, over voorwaarden die aan eigentijds beoordelen en beslissen moeten worden gesteld, en waarmee de ontwikkeling van praktische handreikingen mogelijk is;
- Vanuit het lectoraat zal zowel in wetenschappelijke tijdschriften als in vakbladen worden gepubliceerd over opbrengsten van onderzoek;

Vanuit het lectoraat worden verschillende promotietrajecten begeleid die naadloos aansluiten op de focus van het lectoraat én bij actuele vraagstukken binnen de betreffende

onderwijssector. Het promotietraject van Marijke van Vijfeijken is gericht op het ontwikkelen van kennis over opvattingen van leraren over differentiatie in de klas in relatie tot sociale rechtvaardigheid en kansen(on)gelijkheid in het *primair onderwijs*. De centrale onderzoeksvraag is: welke opvattingen over rechtvaardigheid gebruiken leraren om de keuzes die ze maken voor differentiatie in de klas te legitimeren? De legitimaties van leraren worden geanalyseerd aan de hand distributieve rechtvaardigheidsprincipes (verdiens-te-principe, gelijkheid en behoefte). De inzichten van het onderzoek van Marijke zijn van belang voor de reflecties van leraren op het probleem van kansenongelijkheid en voor het ontwikkelen van interventies om ongelijkheden en ongelijke kansen tussen leerlingen van verschillende sociale achtergronden te verminderen.

Het promotietraject van Janneke van der Steen richt zich op het nemen van beslissingen op grond van formatieve programma's in het *voortgezet onderwijs*. De centrale vraag is: Hoe ontwerp je een programma van formatieve leeractiviteiten dat ondersteunt bij het nemen van beslissingen over vervolgstappen in onderwijs en leren? Daarbij wordt eveneens gekeken naar welke voorwaarden in de school aanwezig moeten zijn om te kunnen werken met formatieve programma's en naar hoe docenten het werken met en het nemen van beslissingen op basis van programma's van formatieve leeractiviteiten in hun eigen klas ervaren. Dit promotietraject is gekoppeld aan het toegekende RAAK PRO-project naar ontwerpprincipes voor een programma van formatieve leeractiviteiten.

Het promotietraject van Jeroen van der Linden richt zich op het snijvlak van leren en beoordelen binnen het *hoger onderwijs*: welke invloed heeft het toetsen van studenten op de metacognitieve vaardigheden van de student? Hoe beïnvloedt toetsen het leren en het monitoren van het leren? Conclusie uit een deelstudie binnen zijn promotie naar de complexe relatie tussen zelfregulatie en 'slagen voor de toets' is dat deze relatie vooral beïnvloed wordt door de mate waarin studenten in staat zijn bewuste keuzes te maken in zelfregulatie-strategieën en door het doel dat de studenten zichzelf hebben gesteld (performance goal/halen van de toets of mastery goal/professionele ontwikkeling). Met deze inzichten heeft Jeroen een typologie onderscheiden van vier 'typen' studenten. Deze typologie kan docenten helpen om bewuste keuzes te maken bij het begeleiden van en sturen op zelfregulatie. Verder onderzoek richt zich op het ervaren verschil tussen summatieve en formatieve beoordelingen en de verschillen hierbij tussen verschillende toetssystemen (conventioneel versus programmatisch).



8. TOT SLOT

Met deze rede heb ik een pleidooi willen doen om beoordelen weloverwogen en bewust in te zetten als integraal onderdeel van onderwijs. Zodat op basis van die beoordelingen zorgvuldige beslissingen over het leren van leerlingen genomen kunnen worden. Daarnaast heb ik, op basis van de verkenning van de theoretische concepten in de definitie van eigentijds beoordelen en beslissen, een poging gedaan de mogelijkheden te schetsen om in ieder onderwijssysteem en binnen ieder toetsstelsel invulling te geven aan het meer eigentijds beoordelen en beslissen. Meer eigentijds betekent in mijn optiek niet dat alles compleet anders moet, juist met kleine aanpassingen kan er al veel bewuster gebruik gemaakt worden van alle informatie die gedurende de schoolloopbaan van een leerling wordt verzameld. Niet zonder meer vertrouwen op de schijnveiligheid van cijfers en rekenkundige formules, geen zwaarwegende beslissingen koppelen aan een enkel toetsmoment, het zijn kleine aanpassingen die in ieder onderwijssysteem mogelijk zijn. Ik zal de komende lectoraatsperiode mijn best doen om hier praktische handvatten voor te ontwikkelen, zodat een ieder die dat wil, aan de slag kan om het beoordelen en beslissen meer eigentijds vorm te geven.

Met deze rede heb ik ook gepoogd om de vaak negatieve connotatie aan begrippen als toetsen en beoordelen te verkleinen. Zoals ik al schreef in de inleiding zouden we, naar mijn mening, het geven en ontvangen van een oordeel vooral moeten zien én ervaren als het voeren van een professionele dialoog waarin docent en leerling samen met elkaar kijken naar waar de leerling al staat. Waarbij het mogelijk én wenselijk is om in alle openheid samen te spreken over wat de leerling al wel en nog niet beheerst. Niet met als doel de leerling daarop af te rekenen, maar met als doel de leerling een stap verder te helpen.

Tom Sherrington (2021) schreef onlangs in een column: "If there were no public exams and no accountability system, teachers would still assess students and they would still want to assess themselves. Why? Because assessment isn't really for other people, it's for ourselves.

An effective assessment system should reinforce our idea of standards, because they are an inherent part of understanding how we are doing in the learning process and that understanding is something we all value. Standards are not a controversial thing - they only become so when they have high stakes and doors are closed to anyone who doesn't meet them.

So, if not to impose limitations, what purpose do assessments serve? They should provide feedback to students that will help them answer the questions that they should always be asking themselves, like 'have I explained it well?' and 'how could I improve?'"

Ik sluit hier van harte bij aan. Regelmatig hoor ik docenten uitspraken doen als 'we moeten minder toetsen afnemen om de toetsdruk te verlagen' en 'al dat toetsen levert de leerlingen zo veel stress op'. Oplossingen worden dan soms gezocht in terugbrengen van het aantal toetsen met een summatieve functie, wat weliswaar tot een vermindering van toetsdruk zorgt, maar in veel gevallen ook tot een verhoging van de toetsstress: voor de leerlingen hangt er van het in aantal beperkte aantal toetsen ineens nog meer af.

Mijn pleidooi zou dan ook zijn: toets meer! Maar toets wel op een zinvolle manier. Gebruik de informatie die deze toetsen opleveren niet om leerlingen op af te rekenen, maar om ze te begeleiden in hun ontwikkeling. Gebruik de informatie om met de leerling het gesprek aan te gaan en samen met de leerling te kijken welke stappen op welke manier het beste gezet kunnen worden. Gebruik de informatie om waardevolle, betrouwbare en onderbouwde beslissingen te nemen: beslissingen die gericht zijn op de (verdere) ontwikkeling van de leerling, op de voortgang ten aanzien van de leerdoelen en zwaarwegende beslissingen die gericht zijn op het vaststellen van wat geleerd is (certificeren, kwalificeren, selecteren, etc.). En laat toetsen, en dus vooral specifiek de informatie die deze toetsen opleveren, op deze manier een integraal onderdeel vormen van het onderwijs!

Ik wil deze rede afsluiten met een woord van dank. Aan het college van bestuur van de HAN en aan Aly Smelt als directeur van de Academie Educatie voor het vertrouwen dat ik deze mooie waardevolle functie op een goede manier kan vervullen. Aan collega-lector Helma Oolbakkink-Marchand en alle onderzoekers voor de fijne wijze waarop wij samenwerken binnen ons onderzoeksteam. En een speciaal woord van dank voor Marjoleine Dobbelaer, Linda Jakobs en Ivanka Moerkerken. Marjoleine en Linda, dank voor het voortdurende sparren en jullie waardevolle en kritische blik bij deze rede en het mee nadenken over de invulling van de lectoraatsopdracht. Ivanka, dank voor het vele meedenken over hoe we alle opbrengsten zo toegankelijk en bruikbaar mogelijk kunnen delen met iedereen die er mee aan de slag wil.

REFERENTIES

Allal, L. (2011, June). Assessment for learning culture in the classroom, and beyond. Paper presented at the 4th Assessment for Learning conference, Bergen, Norway.

Autin, F., Batruch, A., & Butera, F. (2015). Social justice in education: How the function of selection in educational institutions predicts support for (non)egalitarian assessment practices. *Frontiers in Psychology*, 6, Article 707. <https://doi.org/10.3389/fpsyg.2015.00707>

Baartman, L. K. J., & Prins, F. J. (2018). Transparency or Stimulating Meaningfulness and Self-Regulation? A Case Study About a Programmatic Approach to Transparency of Assessment Criteria. *Frontiers in Education*, 3. <https://doi.org/10.3389/educ.2018.00104>

Baartman, L., Van der Vleuten, C. (2015). Kwaliteit van een toetsprogramma. In: *Kwaliteit van toetsing onder de loep*. (pp 79-91). Sluijsmans, Joosten-ten Brinke, Van Schilt-Mol. Garant-Uitgevers.

Baartman, L., Van Schilt-Mol. T. & Van der Vleuten, C. (2020). *Programmatisch toetsen. Ervaringen en voorbeelden uit de praktijk*. Amsterdam: Boom uitgevers. Amsterdam.

Biggs, J. (1996). Enhancing Teaching through Constructive Alignment. *Higher Education*, 32(3), 347-364. <http://www.jstor.org/stable/3448076>

Briggs, D., Ruiz-Primo, M., Furtak, E., Shepard, L. & Yin, Y. (2012). Meta-analytic methodology and inferences about the efficacy of formative assessment. *Educational Measurement: Issues and Practice* Winter, 31(4), 13-17.

Brookhart, S. M. (2011). Educational Assessment Knowledge and Skills for Teachers. *Educational Measurement: Issues and Practice*, 30(1), 3-12.

Brouwer, F. M., & Boer, P. (2013). Auditing Instrument for Sustainability in Higher Education - AISHE 2012 [Beoordelingskader duurzame ontwikkeling]. Den Haag: Hobéon.

- Brown, G. T. L. (2008). *Conceptions of assessment: Understanding what assessment means to teachers and students*. New York: Nova Science Publishers.
- Carless, D., & Boud, D. (2018). The development of student feedback literacy: enabling uptake of feedback. *Assessment & Evaluation in Higher Education*, 43(8), 1315-1325.
- Cauley, K. M., & McMillan, J. H. (2009). Formative assessment techniques to support student motivation and achievement. *Clearing House*, 83(1), 1-6.
- Christodoulou, D. (2016). *Making good progress? The future of Assessment for Learning*. Oxford: Oxford University Press
- Cilliers, L., Schuwirth, L. & Van der Vleuten, C. (2012), Modelling the pre-assessment learning effects of assessment: evidence in the validity chain. *Medical Education* 2012: 46, 1087-1098.
- Commissie externe validering examenkwaliteit hoger beroepsonderwijs. (2012). *Vreemde ogen dwingen*. Retrieved from https://www.scienceguide.nl/wp-content/uploads/2018/08/Vreemde_ogen_dwingen.pdf
- Commissie Kwaliteit Schoolexaminering (2018). *Een volwaardig schoolexamen*. Verkregen via <https://bit.ly/2CjqJ5f>
- De Boer, H., Timmermans, A.C., & Van der Werf, M.P.C. (2016). Inaccurate teacher expectations. Relationships with student and class characteristics, and its effect on longterm student performance. In Trusz, S. & Babel, P (Eds). *Interpersonal and intrapersonal expectations*. (157-161). London and New York: Routledge Taylor & Francis Group.
- De Wolf, I. (2017). *Het Nederlandse onderwijs langs de meetlat van gelijke kansen*. In Hoogeveen, K, Jepman, Y, & Studulski, F (Eds). *Kansen bieden in plaats van uitsluiten. Jubileum uitgave 1992-2017*. Utrecht: Sardes.
- DHO (2010). *AISHE Auditing instrument for sustainability in Higher Education*. Den Haag: DHO.

- Dobbelaer, M., Den Tuinder, E., De Beer, F., Buijs, E., Van der Steen, J. & Van Schilt-Mol, T. (2021b). Ontwikkelen van leerlingvaardigheden voor formatief evalueren. Een praktische toolkit voor docenten. *Examens* 2021 (3), p. 15-20.
- Dobbelaer, M., Den Tuinder, E., De Beer, F., Buys, E. & Van Schilt-Mol, T. (2021a). Toolkit leerlingvaardigheden voor formatief evalueren.
- Dochy, F., & Segers, M. (2018). *Creating Impact Through Future Learning*. London: Routledge, <https://doi.org/10.4324/9781351265768>
- Dochy, F., Berghmans, I., Koenen, A. & Segers, M. (2015). *Bouwstenen voor High Impact Learning*. Amsterdam: Boom Uitgevers.
- Expertgroep bke/ske (2013). *Verantwoord toetsen en beslissen in het hoger beroepsonderwijs. Een voorstel voor een programma van eisen voor een basis- en seniorkwalificatie examinering (bke/ske)*. Den Haag: Vereniging Hogescholen.
- Furtak, E. M., Kiemer, K., Circi, R. K., Swanson, R., de León, V., Morrison, D., & Heredia, S. C. (2016). Teachers' formative assessment abilities and their relationship to student learning: findings from a four-year intervention study. *Instructional Science*, 44(3), 267-291. <https://doi.org/10.1007/s11251-016-9371-3>.
- Gulikers, J. & Baartman, L. (2017). *Doelgericht professionaliseren. Formatief toetsen met effect! Wat DOET de docent in de klas?* Den Haag: NRO.
- HAN (2020). *Beter toetsen maakt je een beter docent!* Verkregen via: <https://www.han.nl/artikelen/2020/12/beter-toetsen-maakt-je-een-beter-docent/>.
- Hattie, J. A. (2009). *Visible Learning A synthesis of over 800 meta-analyses*. London: Routledge.
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77, 811-12.
- Heitink, M. C., Van der Kleij, F. M., Veldkamp, B. P., Schildkamp, K., & Kippers, W. B. (2016). A systematic review of prerequisites for implementing assessment for learning in classroom practice. *Educational Research Review*, 17, 50-62. -

Herppich, S., Praetorius, A.-K., Förster, N., Glogger-Frey, I., Karst, K., Leutner, D., Berhman, L., Böhmer, M., Ufer, S., Klug, J., Hetmanek, A., Ohle, A., Böhmer, I., Karing, C., Kaiser, J. & Südkamp, A. (2018). Teachers' assessment competence: Integrating knowledge-, process-, and product-oriented approaches into a competence-oriented conceptual model. *Teaching and Teacher Education*, 76, 181-193

Joosten-ten Brinke, D. & Draaijer, S. (2015). Kwaliteit van toetsen. In D. Sluijsmans, D. Joosten-ten Brinke, & T. Van Schilt-Mol (Red.), *Kwaliteit van toetsing onder de loep*. (pp. 37-54). Antwerpen, België: Garant.

Joosten-ten Brinke, D. (2011). *Eigentijds toetsen en beoordelen (Lectorale rede)*. Tilburg: Fontys Hogescholen.

Joosten-ten Brinke, D., & Sluijsmans, D. M. A. (2012). Tijd voor toetskwaliteit: het borgen van toetsdeskundigheid van examencommissies. *Thema*, 16-21.

Jussim, L., & Harber, K.D. (2005). Teacher expectations and self-fulfilling prophecies: Knowns and unknowns, resolved and unresolved controversies. *Personality and Social Psychology Review*, 9(2), 131-155.

Kingston, N., & Nash, B. (2011). Formative assessment: A meta-analysis and a call for research. *Educational measurement: Issues and practice*, 30(4), 28-37.

Kirschner, P., Claessens, L., & Raaijmakers, S. (2018). *Op de schouders van reuzen*. Ten Brink Uitgevers.

Kneyber, R., & Evers, J. (2013). *Het Alternatief. Weg met de afrekencultuur in het onderwijs*. Amsterdam: Boom.

Laveault, D., & Allal, L. (2016). *Assessment for Learning: Meeting the Challenge of Implementation*. Springer Cham, <https://doi.org/10.1007/978-3-319-39211-0>.

Lindner, J., & Visser, S. (2018). De winst van peerfeedback. *Onderwijsinnovatie* (1).

Malecka, B. (2021). From teacher to feedback intervention designer: how self-reflective enquiry into feedback practices stimulates innovation. Presentatie tijdens Feedback Literacy Symposium. Guildford UK: University of Surrey,

- Malecka, B. (2021). From teacher to feedback intervention designer: how self-reflective enquiry into feedback practices stimulates innovation. Presentation on 13-1-2021, FEEDBACK LITERACY From Education to Professional practice.
- Malecka, B., Boud, D. & Carless, D. (2020). Eliciting, processing and enacting feedback: mechanism for embedding student feedback literacy within the curriculum. *Teaching in higher Education*. DOI: 10.1080/13562517.2020.1754784
- Malouff, J., & Thorsteinsson, E. (2016). Bias in grading: A meta-analysis of experimental research findings. *Australian Journal of Education*, 60(3), 245-256. <https://doi.org/10.1177/0004944116664618>
- Meijer, K., Van Schilt-Mol, T, Baartman, Dobbelaer, M., Van der Linden, J. & Munneke, L. (2020), Voortdurende toetsbekwaamheid in het hoger beroepsonderwijs, een eerste verkenning. *Examens 2020* (2), 24-34.
- Ministerie van Onderwijs, Cultuur en Wetenschap (2017). Besluit van 16 maart 2017 tot wijziging van het Besluit bekwaamheidseisen onderwijspersoneel en het Besluit bekwaamheidseisen onderwijspersoneel BES in verband met de herijking van de bekwaamheidseisen voor docenten en docenten. *Staatsblad 2017*, 148, verkregen via <https://zoek.officielebekendmakingen.nl/stb-2017-148.html>.
- Molloy, E., Boud, D., & Henderson, M. (2020). Developing a learning-centred framework for feedback literacy. *Assessment & Evaluation in Higher Education*, 45(4), 527-540.
- Offerdahl, E. G., McConnell, M., & Boyer, J. (2018). Can I have your Recipe? Using a Fidelity of Implementation (FOI) Framework to Identify the Key Ingredients of Formative Assessment for Learning. *CBE Life Sciences Education*, 17(16), 1-9. <https://doi.org/10.1187/cbe.18-02-0029>
- Onderwijsraad (2018). *Toets wijzer*. Den Haag: Onderwijsraad.
- Oudkerk Pool, A., Govaerts, M.J.B., Jaarsman, D.A.D.C., Driessen, E.W. (2018). From aggregation to interpretation: how assessors judge complex data in a competency-based portfolio. *Health Science Education* (2018) 23, 275-287
- Pastore, S., & Andrade, H. L. (2019). Teacher assessment literacy: A three-dimensional model. *Teaching and Teacher Education*, 84, 128-138.

- Pearce, J., Mulder, R., & Baik, C. (2009). *Involving students in peer review: Case studies and practical strategies for university teaching*. The University of Melbourne: Centre for the Study of Higher Education.
- Peters, M., De Beer, F., Van Schilt-Mol, T., Sluijsmans, D., Jakobs, L. & Van der Linden (2017). *De Toetsing Getoetst onderzocht: herontwerp van de methodiek om de kwaliteit van toetsing te analyseren, verbeteren en borgen*. Nijmegen: Hogeschool van Arnhem en Nijmegen.
- Popham, W. J. (2009). Assessment literacy for teachers: Faddish or fundamental? *Theory into Practice*, 48(1), 4-11.
- Price, M., Rust, C., O'Donovan, B., Hindley, K., & Bryant, R. (2012). *Assessment Literacy: The Foundation for Improving Student Learning*. Oxford: Oxford Centre for Staff and Learning Development.
- Pryor, J., & Crossouard, B. (2008). A socio-cultural theorisation of formative assessment. *Oxford Review Of Education*, 34(1), 1-20.
- Ramaswamy, V., Veremis, B., & Nalliah, R. P. (2020). Making the case for pass-fail grading in dental education. *European Journal of Dental Education*. 2020;24:601-604.
- Resh, N., & Sabbagh, C. (2016). Justice and education. In C. Sabbagh & M. Schmitt (Eds.), *Handbook of justice theory and research* (pp. 349-368). New York: Springer.
- Ruiz-Primo, M. A., & Furtak, E. M. (2007). Exploring teachers' informal formative assessment practices and students' understanding in the context of scientific inquiry. *Journal of Research in Science Teaching*, 44(1), 57-84.
- Sadler, D. (2009). Indeterminacy in the use of preset criteria for assessment and grading, *Assessment & Evaluation in Higher Education*, 34:2, 159-179, DOI: 10.1080/02602930801956059
- Sadler, R. D. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, 18, 119-144.
- Sambell, K., McDowell, L., & Montgomery, C. (2013). *Assessment for Learning in Higher Education*. Abington: Routledge.

- Schildkamp, K., & Kuiper, W. (2010). Data-informed curriculum reform: Which data, what purposes, and promoting and hindering factors. *Teaching and Teacher Education*, 26(3), 482-496. <https://doi.org/10.1016/j.tate.2009.06.007>
- Schildkamp, K., Van der Kleij, F.M., Heitink, M.C., Hoogland, I.L., Kippers, W.B., Dijkstra, A., & Veldkamp, B. (2014). Voorwaarden voor effectieve formatieve toetsing, een praktische review. Den Haag: NRO.
- Schut, S, Maggio, L., Heeneman, S., van Tartwijk, J., van der Vleuten, C., & Driessen, E. (2020). Where the rubber meets the road - An integrative review of programmatic assessment in health care professions education. *Perspect Med Educ*. 2021 Jan;10(1):6-13. doi: 10.1007/s40037-020-00625-w.
- Sherrington, T. (2021). What is the purpose of assessment in education? Verkregen via: <https://www.century.tech/news/what-is-the-purpose-of-assessment-in-education/>
- Simpson-Beck, V. (2011). Assessing classroom assessment techniques. *Active Learning in Higher Education*, 12(2), 125-132.
- Sluijsmans, D. & Seegers, M. (2018). Toetsrevolutie. Naar een feedbackcultuur in het hoger onderwijs. Culemborg: Phronese.
- Sluijsmans, D. (2013). Verankerd in leren. Vijf bouwstenen voor professioneel beoordelen in het hoger beroepsonderwijs. Heerlen: Zuyd Hogeschool, p. 23-24
- Sluijsmans, D. (2020). Toetsing als kans voor leren. Den Haag: NRO.
- Sluijsmans, D., & Kneyber, R. (2016). Toetsrevolutie. Naar een feedbackcultuur in het voortgezet onderwijs. Culemborg: Phronese.
- Sluijsmans, D., Boei, F., de Beer, F., van der Linden, J., & Meijer, K. (2017). Het versterken van de toetsbekwaamheid in het hoger beroepsonderwijs. Den Haag: Vereniging Hogescholen.
- Sluijsmans, D., Joosten-ten Brinke, D. & Vleuten, C. van (2013). Toetsen met leerwaarde. Een reviewstudie naar de effectieve kenmerken van formatief toetsen. Den Haag: NRO.
- Sluijsmans, D., Van Schilt-Mol, T., Peeters, A., & Joosten-ten Brinke, D. (2015). Kwaliteit van toetsing in het hoger onderwijs onder de loep. In D. Sluijsmans, D. Joosten-ten

- Brinke, & T. Van Schilt-Mol (Red.), *Kwaliteit van toetsing onder de loep. Handvatten om de kwaliteit van toetsing in het hoger onderwijs te analyseren, verbeteren en borgen*. Antwerpen-Apeldoorn: Garant, pp. 15-24.
- SPV (2020). *Toolbox visie op toetsing en examinering*. <https://www.platformsvmbbo.nl/themas/schoolexaminering/visie-op-schoolexaminering>.
- Staf en services OOK & Lectoraat Eigentijds beoordelen en beslissen (2021). *HANdreiking toetsbeleid 2021*. Nijmegen: Hogeschool van Arnhem en Nijmegen.
- Stobart, G. (2008). *Testing times: The uses and abuses of assessment*. Abingdon, England: Routledge.
- Straetmans, G. (2006). *Bekwaam beoordelen en beslissen*. Deventer: Saxion Hogescholen.
- Tai, J., Ajjawi, R., Boud, D., Dawson, P., & Panadero, E. (2018). Developing Evaluative Judgement: Enabling Students to Make Decisions about the Quality of Work. *Higher Education* 76(3), 467-481.
- Timmermans, A.C., Kuyper, H., & Van der Werf, M.P.C. (2015). Accurate, inaccurate, or biased teacher expectations: Do Dutch teachers differ in their expectations at the end of primary education? *British Journal of Educational Psychology*, 85, 459-478.
- Van Berkel, H., Bax, A. & Joosten-ten Brinke, D. (2014). Het toetsproces ontleed. In H. van Berkel, A. Bax & D. Joosten-ten Brinke (Red.), *Toetsen in het hoger onderwijs* (pp 1-11). Houten: Bohn Stafleu Van Loghum.
- Van den Akker, J. (2003). Curriculum perspectives: an introduction. In J. Van den Akker, W. Kuiper, & U. Hameyers (Eds.), *Curriculum perspectives: an introduction*. Dordrecht: Kluwer Academic Publishers.
- Van der Linden, J., van Schilt-Mol, T., Nieuwenhuis, L. & Vleuten van der, C. (2021). Learning for a Summative Assessment: The Relationship Between Students' Academic Achievement and Self-Regulated Learning. *Open Journal of Social Sciences*, 2021, 9, 351-367.
- Van der Ploeg, S. & Van Engen, N. (2017). *Wat is er in de wetenschappelijke literatuur bekend over de effecten van formatief evalueren?* Den Haag: Kennisrotonde

- Van der Steen, J. (2020). Effectief formatief evalueren: het ontwerpen van programma's van formatieve leeractiviteiten. Presentatie gegeven op 12 november 2020. Verkregen via: <https://bit.ly/2TyCVKb>.
- Van der Steen, J., Van Schilt-Mol, T., Joosten-ten Brinke, D. & Van der Vleuten, C. (2020). Formatief evalueren: naar een programma van formatieve leeractiviteiten. Tijdschrift voor Docentenopleiders 41(4), 274-285.
- Van der Steen, J., van Schilt-Mol, T., Joosten-ten Brinke, D., & van der Vleuten, C. P. M. (in voorbereiding). Vignetten als gesprekstool voor gesprekken op school over condities voor het implementeren van formatief evalueren. Artikel in voorbereiding.
- Van der Vleuten (2017). Geen cijfers maar woorden. In: D. Sluijsmans (red). In de steigers gezet. Opbrengsten van de eerste periode van het lectoraat Professioneel Beoordelen 2012-2016, p. 31-33. Heerlen: Zuyd Hogeschool.
- Van der Vleuten, C., Schut, S. & Heeneman, S. (2018). Programmatisch toetsen als motor voor professioneel leren in het hoger onderwijs. In D. Sluijsmans & M. Segers. Toetsrevolutie. Naar een feedbackcultuur in het hoger onderwijs (p. 124-133). Culemborg: Phronese.
- Van der Vleuten, C., Schuwirth, L., Driessen, E., Govaerts, M. & Heeneman, S. (2015). 12 Tips for programmatic assessment. *Medical Teacher*, 37, 641-646.
- Van Schilt-Mol, T., Baartman, L., Meijer, K., Dobbelaer, M., Linden van der, J. & Munneke, L. (2020). Kwaliteiten van de toetsbekwame hbo-docent. Herijking van het programma van eisen voor de basis- en seniorkwalificatie examinering zoals opgesteld door de Expertgroep BKE/SKE (2013). Nijmegen/Utrecht: Hogeschool van Arnhem/Hogeschool van Utrecht.
- Van Schilt-Mol, T. & Sluijsmans, D. (2018). De Toetsing Getoetst: een methodiek om als opleidingsteam op formatieve wijze te werken aan duurzame toetskwaliteit. In: D. Sluijsmans & M. Seegers (red.). Toetsrevolutie. Naar een feedbackcultuur in het hoger onderwijs. (p. 194-208). Culemborg: Phronese.
- Van Schilt-Mol, T. & Sluijsmans, D. (2015). Kwaliteit van toetsing geoperationaliseerd. In: D. Sluijsmans, D. Joosten-ten Brinke & T. Van Schilt-Mol, (Red.). Kwaliteit van toetsing onder de loep. Handvatten om de kwaliteit van toetsing in het hoger onderwijs te analyseren, verbeteren en borgen (p. 151-196). Leuven: Garant.

- Van Schilt-Mol, T. (2018). Toets in de geest van toetsen. Examens, 2018-4, p. 26-27.
- Van Schilt-Mol, T., Sluijsmans, D., Peters, M., Beer, F. de & Jakobs, L. (2016). De touwtjes in handen: toetskwaliteit in het hbo. Tijdschrift OnderwijsInnovatie, december 2016, 15-24
- Van Vijfeijken (in press). Wat is eerlijk? Werken aan kansengelijkheid in het onderwijs. Huizen: Uitgeverij Pica.
- Van Vijfeijken, M. & Van Schilt-Mol, T. (2020). Gelijke kansen door gestandaardiseerde toetsen: wat vindt de docent? Tijdschrift voor Docentenopleiders 41(4), 302-312.
- Van Vijfeijken, M., Buijs, E., & Van Schilt-Mol, T. (in press). Rechtvaardig beoordelen en beslissen. In M. van Vijfeijken (Red.). Wat is eerlijk? Werken aan kansengelijkheid in het onderwijs. Huizen: Uitgeverij Pica.
- Vereniging Hogescholen (2013). Protocol inzake wederzijdse erkenning basiskwalificatie didactische bekwaamheid. Verkregen via: <https://bit.ly/2wf7ysX>
- Veugen, M. J., Gulikers, J. T. M., & Den Brok, P. (2021). We agree on what we see: Teacher and student perceptions of formative assessment practice. Studies in Educational Evaluation, 70, 101027. <https://doi.org/10.1016/j.stueduc.2021.101027>
- VO-Raad (2019). De ontwikkeling van een visie op toetsing en schoolexaminering in het VO. Utrecht: VO-Raad
- VO-Raad (2019a). versterking schoolexaminering via negen actielijnen Verkregen via: <https://www.vo-raad.nl/nieuws/vo-raad-versterking-schoolexaminering-via-negen-actielijnen>.
- VO-Raad (2019b). Monitor 'Negen actielijnen ter versterking van de kwaliteitsborging schoolexaminering' - november 2019. Verkregen via: https://www.vo-raad.nl/system/downloads/attachments/000/000/866/original/Monitor_Negen_actielijnen_november_2019_V6_%28002%29.pdf?1578491085
- VO-Raad (2020). Monitor 'Negen actielijnen ter versterking van de kwaliteitsborging schoolexaminering' - november 2020. Verkregen via: https://www.vo-raad.nl/system/downloads/attachments/000/001/156/original/Monitor_Negen_actielijnen_november_2020.pdf?1611240095

- Waitling, C., & Ginsburg, S. (2019). Assessment, feedback and the alchemy of learning. *Medical Education* 2019, 53: 76-85. Doi: 10.1111/medu.13645
- Wehrli, U. (2006). Tidying up art. TedTalk 2006. https://www.ted.com/talks/ursus_wehrli_tidying_up_art.
- Wehrli, U. (2013) *The Art of Clean Up. Life made neat and tidy*. San Francisco: Chronicle Books
- William, D. (2013a). Assessment: The Bridge between Teaching and Learning. *Voices from the Middle*, 21(2), 15-20. Retrieved on 23 June 2021, from <https://library.ncte.org/journals/vm/issues/v21-2/24461>
- William, D. (2013a). *Cijfers geven werkt niet*. Meppel: Ten Brink
- Winstone, N. & Carless, D. (2020). *Designing Effective Feedback Processes in Higher Education: A Learning-Focused Approach*. Routledge.
- Winstone, N., Nash, R., Parker, M. & Rowntree, J. (2017). Supporting learners' agentic engagement with feedback: a systematic review and a taxonomy of reciprocity processes. *Educational Psychologist*, 52(1), 17-37.
- Wubbels T., van Tartwijk J. (2017) Evidence Informed Innovation of Education in the Netherlands: Learning from Reforms. In: Eryaman M., Schneider B. (eds) *Evidence and Public Good in Educational Policy, Research and Practice*. Educational Governance Research, vol 6. Springer, Cham. https://doi.org/10.1007/978-3-319-58850-6_7
- Xu, Y., & Brown, G. T. L. (2016). Teacher assessment literacy in practice: A reconceptualization. *Teaching and Teacher Education*, 58, 149-162.
- Yorke, M. (2003). Formative assessment in higher education: Moves towards theory and the enhancement of pedagogic practice. *Higher Education*, 45, 477-501.

OVER DE AUTEUR



Dr. Tamara van Schilt-Mol (1978) studeerde taalwetenschappen aan de Universiteit van Tilburg. Bij de faculteit Geesteswetenschappen deed zij onderzoek naar onbedoelde moeilijkheden in de Cito Eindtoets Basisonderwijs. Na haar promotie werkte zij als (senior) onderzoeker bij IVA, waar zij onderzoek deed naar onder meer succesfactoren in inburgeringsonderwijs, passend onderwijs, muziekeducatie en opbrengstgericht werken. Van november 2012 tot mei 2020 was Tamara werkzaam als associate lector Toetsen en Beoordelen bij de Hogeschool van Arnhem en Nijmegen. Met de inzet van praktijkgericht onderzoek werden binnen dit associate lectoraat onder meer het toetsweb, het spel Toetsing Getoetst en de kwaliteiten van de toetsbekwame hbo-docent ontwikkeld. Vanaf mei 2020 is Tamara lector Eigentijds Beoordelen en Beslissen.

EIGENTIJDS BEOORDELEN EN BESLISSEN

Binnen dit lectoraat wordt onderzocht hoe eigentijds beoordelen en beslissen zodanig ontworpen, ingericht en uitgevoerd kan worden, dat het bijdraagt aan het stimuleren van het leren, het nemen van zorgvuldige beslissingen en het optimaliseren van de kwaliteit van het onderwijs.

OPEN UP
NEW
HORIZONS.



HAN UNIVERSITY
OF APPLIED SCIENCES