

Beschouwend artikel

Formatief evalueren: naar een programma van formatieve leeractiviteiten

Janneke van der Steen - Hogeschool van Arnhem en Nijmegen, Tamara van Schilt-Mol - Hogeschool van Arnhem en Nijmegen, Desirée Joosten-ten Brinke - Open Universiteit, Cees van der Vleuten - Universiteit Maastricht

“Wat moet ik leren voor de toets?” “Telt deze toets mee?”

Leerlingen lijken als ze leren vooral gericht te zijn op wat er in een toets gevraagd zal gaan worden. Docenten zien het organiseren van toetsen met cijfers vaak als enige manier om leerlingen aan het leren te krijgen. Tegelijkertijd willen zij echter voorkomen dat leerlingen van toets naar toets werken en dat het leren stopt bij het ontvangen van een cijfer. Docenten zijn dan ook op zoek naar manieren om zo te toetsen dat dit het leren stimuleert en het nemen van zelfregie bevordert.

Docenten die daarbij kiezen voor formatief evalueren, ervaren een toenemende motivatie en eigenaarschap bij leerlingen, en hebben zelf meer plezier in het lesgeven. Toch ervaren zij ook de nodige knelpunten. Deze ontstaan doordat docenten veelal losstaande formatieve activiteiten inzetten: er is daarbij geen sprake van een bewust gekozen combinatie van activiteiten die voldoende informatie oplevert om de docent te ondersteunen bij het nemen van een specifieke beslissing, bijvoorbeeld over het bijstellen van het onderwijs of over het al dan niet bereikt hebben van een leerdoel.

Hoewel er veel informatie is over het ontwerpen en inzetten van formatieve leeractiviteiten, is er nog weinig bekend over het ontwerpen van programma's van formatieve leeractiviteiten. In dit artikel wordt beschreven waarom het denken vanuit programma's van formatieve leeractiviteiten van belang is voor de effectiviteit van formatief evalueren. Het ontwerpen van deze programma's kan in alle onderwijssectoren toegepast worden; lerarenopleidingen kunnen het daarnaast opnemen als onderdeel van het curriculum.

Wat is formatief evalueren en waarom is het belangrijk?

Toetsen zijn een essentieel onderdeel van het onderwijs. Toetsen kunnen helpen de voortgang van leerlingen en de effectiviteit van lesgeven in kaart te brengen, en zo als instrument dienen om onderwijs op maat aan te bieden (Schildkamp et al., 2014). Binnen toetsen wordt doorgaans een onderscheid gemaakt tussen toetsen met een summatieve en toetsen met een formatieve functie. Als informatie uit toetsen wordt gebruikt om vast te stellen of de doelen van onderwijs zijn bereikt en om beslissingen te nemen over slagen/zakken/certificeren/toelating, spreken we over de summatieve functie. Als de informatie wordt gebruikt om het proces van onderwijs en leren gedurende het onderwijs te verbeteren, spreken we over de formatieve functie. Idealiter omvatten toetsen – in meer of mindere mate, en afhankelijk van hoe de docent omgaat met de informatie die de toets oplevert – beide functies: een toets die afgenomen wordt met primair een summatieve functie, geeft uiteraard ook informatie over de ontwikkelpunten van de leerling. De informatie uit deze toets kan daarmee ook worden gebruikt om het proces van onderwijzen en leren te verbeteren en te sturen, en heeft daarmee dus – mogelijk – ook een formatieve functie (Van Schilt-Mol et al., 2020).

Binnen het voortgezet onderwijs ligt er een focus op het inzetten van toetsen met een summatieve functie. Hierop is echter in toenemende mate kritiek. Dit heeft vooral te maken met de consequenties die toetsen hebben voor de leerling, docent of school (Taras, 2007). Docenten hebben sterk 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. Dit gevoel is niet verwonderlijk, aangezien de kwaliteit van het voortgezet onderwijs veelal wordt vastgesteld op basis van slagingspercentages, eindexamenresultaten en doorstroomcijfers. Daarnaast merken de docenten dat toetsen met een uitsluitend summatieve functie niet vanzelfsprekend bijdragen aan verder leren: summatieve toetsen zijn vaak retrospectief en vinden plaats na het leren (Simpson-Beck, 2011; Cauley & McMillan, 2009). Hierdoor is feedback vaak beperkt en komt deze te laat voor de leerling (Sluijsmans, Joosten-ten Brinke, & Van der Vleuten, 2013); de leerling hoeft feitelijk niets meer te doen met informatie die een toets oplevert (Sadler, 1989). Bovendien blijkt dat toetsen die gericht zijn op certificering en selectie waarbij geen informatieve feedback wordt gegeven of waarbij er geen sprake is van inbedding van de toets in het leerproces, zelfs een negatief effect kunnen hebben op het leren, doordat ze leerlingen belemmeren om verantwoordelijkheid te nemen zich te motiveren om te leren (Hattie & Timperley, 2007).

Als gevolg van de toenemende weerstand verkennen veel scholen de mogelijkheid om de focus te verleggen naar formatief evalueren. Formatief evalueren is een bewezen aanpak die de ontwikkeling, betrokkenheid en het langetermijnleren van leerlingen versterkt: zo blijkt uit onderzoek dat formatief evalueren kan leiden tot meer motivatie, eigenaarschap en zelfregulatie bij leerlingen (Van der Ploeg & Van Engen, 2017; Stobart, 2008) en tot beter onderwijs en betere leerprestaties (Briggs, Ruiz-Primo, Furtak, Shepard, & Yin, 2012; Kingston & Nash, 2011; Van der Ploeg & Van Engen, 2017; Wiliam, Lee, Harrison, & Black, 2004). In dit artikel wordt de volgende definitie van formatief evalueren gehanteerd: “Activiteiten in een klas zijn formatief als bewijsmateriaal van het niveau wordt verzameld, geïnterpreteerd en gebruikt door leraren, leerlingen en hun klasgenoten om beslissingen te nemen over de volgende stappen in het onderwijs die waarschijnlijk beter, of beter onderbouwd, zijn dan de beslissingen die genomen zouden zijn in de afwezigheid van dergelijk verzameld bewijsmateriaal” (Black & Wiliam, 2009, p. 7).

Binnen het voortgezet onderwijs worden de formatieve (leer)activiteiten waar Black en Wiliam op doelen vaak gezien als voorbereidende activiteiten richting de afname van een toets met een summatieve functie waarin de beheersing van de leerdoelen wordt vastgesteld. Bij het ontwerpen en inzetten van deze activiteiten, het proces van formatief evalueren, worden vijf kernstrategieën onderscheiden (Black & Wiliam, 2009). Dit zijn (1) het verhelderen, bespreken en begrijpen van de leerdoelen en de succescriteria, (2) het uitlokken van bewijs van leren, (3) het geven van inhoudelijke terugkoppeling waarmee het leren wordt bevorderd, (4) het activeren van leerlingen zodat ze een instructiebron worden voor hun medeleerlingen en (5) het activeren van leerlingen als eigenaar van hun eigen leerproces. Hieraan worden gebruikelijk drie kernvragen gekoppeld, die door Hattie & Timperly (2007) zijn gekoppeld aan drie activiteiten:

Feed-up: waar gaat de leerling naartoe?

Feedback: waar staat de leerling nu?

Feed-forward: hoe komt de leerling naar de gewenste situatie / hoe nu verder?

Uit onderzoek blijkt dat er acht effectieve methoden van formatief evalueren te onderscheiden zijn (Sluijsmans et al, 2013). Gekoppeld aan de vijf kernstrategieën en de drie kernvragen, bieden deze methoden de docenten concrete handvatten om in de klas invulling te geven aan formatief evalueren.

	Feed-up: <i>Waar gaat de leerling naartoe?</i>	Feedback: <i>Waar staat de leerling nu?</i>	Feed-forward: <i>Hoe kan de leerling het doel behalen?</i>
Docent	Verhelderen, bespreken en begrijpen van de leerdoelen en de <u>succescriteria</u> : feedback (1), vragen stellen (2), <u>toetsdialogen</u> (3), (<u>beoordelings</u>)rubrieken (4), reflectieve lessen (5), peer-assessment (6) en <u>self-assessment</u> (7)	Uitlokken van bewijs van leren: feedback (1), vragen stellen (2), <u>toetsdialogen</u> (3), beoordelingsrubrieken (4), reflectieve lessen (5), <u>summatieve</u> toetsen (8)	Geven van inhoudelijke terugkoppeling waarmee het leren wordt bevorderd: feedback (1), reflectieve lessen (5)
Medeleerling		Activeren van leerlingen zodat ze een instructiebron worden voor hun medeleerlingen: feedback (1), <u>toetsdialogen</u> (3), beoordelingsrubrieken (4), reflectieve lessen (5), peerassessment (6)	
Leerling		Activeren van leerlingen als eigenaar van hun eigen leerproces: beoordelingsrubrieken (4), reflectieve lessen (5), <u>self-assessment</u> (7)	

Figuur 1. Koppeling van acht bewezen effectieve methoden van formatief evalueren volgens Sluijsmans et al. (2013) aan de kernstrategieën van Black en Wiliam (2009) en de feedbackvragen van Hattie en Timperley (2007).

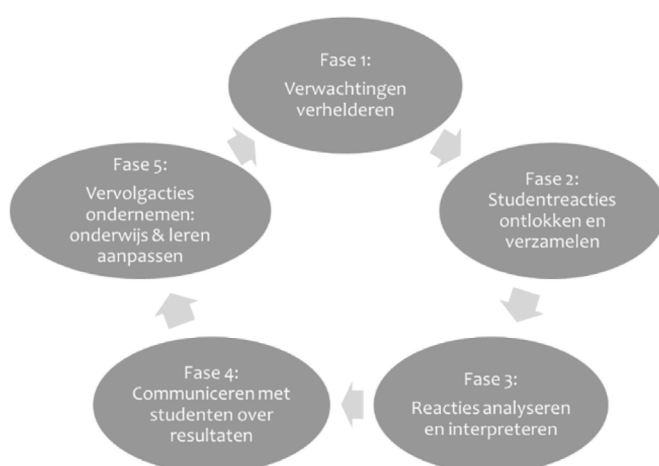
Zoals te zien in Figuur 1 zijn docenten – naast leerlingen – cruciale schakels in het formatieve toetsproces (Sluijsmans et al., 2013; Schildkamp et al., 2014). Het door Gulikers en Baartman (2017) uitgevoerde reviewonderzoek richt zich daarom specifiek op docenten en de effecten die hun formatief handelen heeft. Vragen die hierbij onder meer centraal stonden, waren:

Wat doet de docent die effectieve formatieve toetspraktijken in de klas bereikt?

Welke effecten hebben deze toetspraktijken in termen van leeractiviteiten, leerprocessen en/of leeruitkomsten?

Welke formatieve toetscompetenties heeft een docent nodig?

Om antwoord te kunnen geven op deze vragen gebruikten Gulikers en Baartman de door hen ontwikkelde formatieve toetscyclus als uitgangspunt (Figuur 2).



Figuur 2. Theoretisch en analysekader van de overzichtsstudie van Gulikers en Baartman (2017): de formatieve toetscyclus.

Gulikers en Baartman (2017) concluderen op basis van hun analyse dat, hoewel er geen harde uitspraken gedaan kunnen worden over daadwerkelijk causale relaties tussen docentgedrag en effecten bij leerlingen, het gedrag van docenten afhankelijk lijkt van het effect dat docenten met de formatieve evaluatie willen bereiken. Docenten die formatieve toetsen inzetten om kennisontwikkeling te stimuleren, en om te achterhalen waar de leerlingen staan zodat ze weten wat de volgende stap is om leerlingen verder te brengen, kenmerken zich door het doorlopen van een kleine formatieve toetscyclus (stappen 2-3-5). Docenten die formatieve toetsen inzetten gericht op het stimuleren van zelfregulatie, zelfsturing of autonomie, formuleren expliciete leerdoelen op participatie en samenwerking, en zetten een rijkheid aan strategieën in om leerlingen actief te betrekken bij de verschillende fasen van de formatieve toetscyclus. Om daadwerkelijk formatieve toetspraktijken uit te kunnen voeren moeten de docenten vakinhoudelijke kennis kunnen toepassen, moeten zij beschikken over een vakdidactisch handelingsrepertoire, en moeten zij activiteiten kunnen ontwikkelen die leerlingen een actieve rol geven in de verschillende fasen van de cyclus.

Zowel Sluijsmans et al. (2013) en Schildkamp et al. (2014) als Gulikers en Baartman (2017) stellen vast dat de mate waarin formatief handelen succesvol geïmplementeerd wordt, sterk afhankelijk is van de toetsbekwaamheid van docenten. Docenten moeten kennis en vaardigheden hebben om informatie uit (formatieve) toetsen te kunnen verzamelen en interpreteren, en ze moeten in staat zijn om op basis van deze interpretaties het leren en lesgeven te kunnen verbeteren (Schildkamp et al., 2014).

Wat is de noodzaak van een programma van formatieve leeractiviteiten?

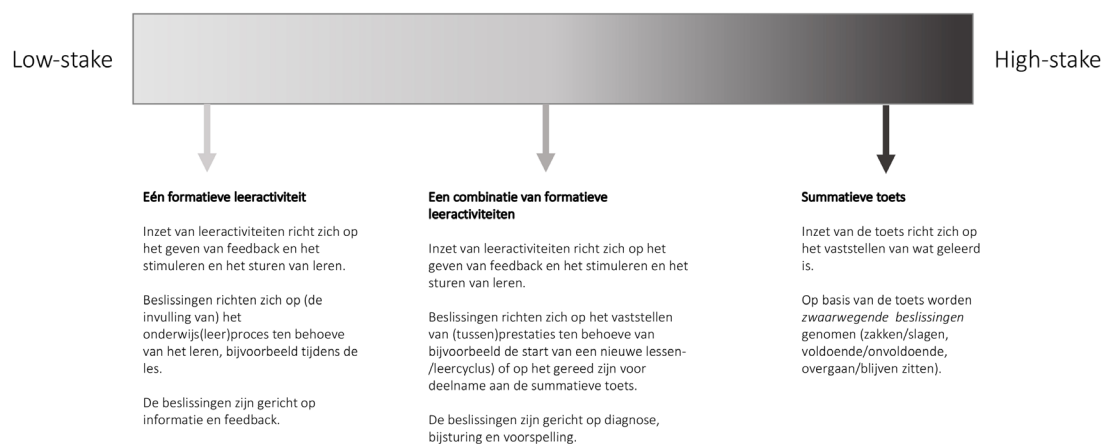
De uitgevoerde reviewstudies hebben, net als andere onderzoeken (William & Leahy, 2018; Christodoulou, 2016; Sluijsmans & Kneyber, 2016) naast kennis over formatief evalueren ook praktische handvatten en richtlijnen opgeleverd. Kenmerkend voor deze handreikingen is dat zij zich richten op praktische toepassing van onderdelen van formatief evalueren, zoals methoden van formatief evalueren (toetsdialoog, feedback, peer-assessment, et cetera) en het volgen van beschreven stappen om specifieke formatieve leeractiviteiten te ontwikkelen (formatieve toetscyclus).

In de praktijk blijkt echter dat docenten over het algemeen weliswaar goed in staat zijn om, op basis van alle kennis en handreikingen die uit onderzoek beschikbaar zijn, losse formatieve leeractiviteiten te ontwikkelen, maar dat het voor hen lastig is om op basis van de informatie uit de verschillende losstaande leeractiviteiten beslissingen te moeten nemen over het leren van leerlingen.

Binnen het traditionele toetssysteem worden beslissingen over de ontwikkeling van individuele leerlingen veelal genomen op basis van afzonderlijke toetsen: ieder onderdeel van een onderwijsprogramma (een hoofdstuk, een onderwijsperiode, et cetera) wordt ofwel afgesloten met een toets met een summatieve functie (Van der Vleuten, Schut, & Heeneman, 2017) ofwel met een berekening van het gemiddelde op de behaalde toetsen die in dit onderdeel van het toetsprogramma zijn afgenomen. Bij deze manier van toetsen worden toetsing en onderwijs of leren vaak los van elkaar gezien, terwijl juist de verbinding van het leren aan onderwijs en toetsing van waarde kan zijn voor de ontwikkeling van de leerling (Baartman, Van Schilt-Mol, & Van der Vleuten, 2020).

Binnen de huidige praktijken van formatief evalueren in het voortgezet onderwijs wordt vaak de beslissing genomen het aantal summatieve toetsen (drastisch) te verminderen, en gebruik te maken van meerdere formatief ingezette toetsen, ook wel formatieve leeractiviteiten, ter voorbereiding op de summatieve toetsen. De informatie uit deze formatieve leeractiviteiten wordt zowel gebruikt om inzicht te krijgen in de ontwikkeling van de leerling (waar staat de leerling nu) als om het leerproces van de leerling te stimuleren en te sturen (hoe kan de leerling een volgende stap zetten in het toewerken naar het beoogde doel). Juist door deze leeractiviteiten te zien als integraal onderdeel van het onderwijs wordt het mogelijk om een leerling te begeleiden in zijn ontwikkeling (Baartman et al., 2020).

Als de functie van formatieve leeractiviteiten en de summatieve toets in relatie gebracht worden met het type beslissing dat binnen de context van het voortgezet onderwijs genomen wordt, kan een continuüm van beslissingen worden geschetst. Hierbij wordt aangesloten bij het gedachtegoed van programmatisch toetsen, waarbij een enkele toets nooit ingezet kan worden om een zwaarwegende beslissing te nemen, en waarbij de ontwikkeling van een leerling centraal staat (Baartman et al., 2020; Van der Vleuten, Schuwirth, Driessen, Govaerts, & Heeneman, 2015; Van der Vleuten et al., 2012). In het continuüm kan een onderscheid worden gemaakt naar zogenoemde low-stake, intermediate-stake en high-stake beslissingen. Vertaald naar de context van het voortgezet onderwijs waarbij het gebruik van formatieve leeractiviteiten is geïntegreerd in het traditionele toetssysteem, ziet het continuüm eruit als in Figuur 3.



Figuur 3. Continuüm van beslissingen binnen het vo, uitgaande van het traditionele toetssysteem (gebaseerd op Van der Vleuten et al., 2015).

Onder low-stake beslissingen worden beslissingen verstaan die gericht zijn op (de invulling van) het onderwijs(leer)proces ten behoeve van het leren, bijvoorbeeld tijdens de les. Deze beslissingen worden low-stake beslissingen genoemd, omdat er geen zwaarwegende consequenties aan verbonden zijn (Van der Vleuten et al., 2017): de informatie die beschikbaar is op basis van één leeractiviteit wordt dan gebruikt om feedback te geven, het onderwijs bij te stellen en het leren te stimuleren, maar wordt niet gebruikt om een zwaarwegende beslissing te nemen (bijvoorbeeld zakken of slagen voor een toets).

Onder intermediate-stake beslissingen worden beslissingen verstaan die gericht zijn op het vaststellen van de beheersing van het leerdoel of (tussen)prestaties bijvoorbeeld ten behoeve van de start van een nieuwe lessen-/leercyclus, of als opmaat naar de afname van een summatieve toets. Deze beslissingen worden intermediate-stake beslissingen genoemd omdat ze wel de basis zijn voor tussentijdse voortgangsbepalingen, maar niet worden gebruikt voor zwaarwegende beslissingen. De resultaten van een aantal leeractiviteiten zouden in samenhang met elkaar voldoende informatie moeten geven om vast te kunnen stellen of de leerling qua ontwikkeling ver genoeg gevorderd is om deel te kunnen nemen aan een toets met een summatieve functie waaraan een high-stake beslissing gekoppeld is: het zakken of slagen voor een toets.

In de praktijk blijkt dat docenten goed in staat zijn invulling te geven aan de low-stake beslissingen, als deze gebaseerd zijn op een enkele formatieve activiteit. Het nemen van een intermediate-stake beslissing is echter vaak niet mogelijk, doordat de formatieve activiteiten op basis waarvan deze beslissing genomen moet worden onvoldoende samenhang vertonen om een uitspraak te kunnen doen. Dit wordt geïllustreerd met het praktijkvoorbeeld in Tekstkader 1.

De schoolleiding van Sjors (docent aardrijkskunde) heeft besloten formatief te gaan evalueren. Sjors staat daar achter: zijn ervaring is dat leerlingen alleen leren voor een cijfer, en dat het móéten halen van goede cijfers veel stress oplevert. Sjors krijgt de volgende instructie: “Per onderwijsperiode mag je één toets afnemen waarvoor je een cijfer geeft. In de voorbereiding naar deze toets toe mogen alleen formatieve leeractiviteiten worden ingezet. Geen cijfers dus!”

Vol goede moed gaat Sjors aan de slag. In een handboek voor formatief evalueren heeft hij allerlei voorbeelden van formatieve leeractiviteiten gezien. Hij kiest een kennistoets waarbij leerlingen feedback krijgen in plaats van een cijfer, een online kennisquiz en hij laat leerlingen geen vingers meer opsteken maar eerst samen nadenken over een vraag voordat hij random iemand aanwijst om antwoord te geven.

Als het proefwerk nadert, stijgt de stress: Sjors heeft verschillende formatieve leeractiviteiten ingezet maar heeft geen goed zicht op of alle leerlingen de leerdoelen ook echt hebben behaald. Zijn stress is nog niets vergeleken bij die van de leerlingen: zij hebben geen idee of ze de stof al beheersen en hebben maar één kans om een goed cijfer te halen. Helaas haalt 45% van de leerlingen een onvoldoende.

Sjors zit met de handen in het haar: wat had hij anders kunnen doen? Hoe had hij de leerdoelen kunnen vertalen naar een set van formatieve leeractiviteiten die hem inzicht zou hebben gegeven in de ontwikkeling van de leerlingen? Hoe had hij die informatie in kunnen zetten om de leerlingen te ondersteunen in hun ontwikkeling? En hoe had hij ervoor kunnen zorgen dat de resultaten op de toets geen verrassing, maar een bevestiging van zijn beeld van de leerlingen zouden zijn?

Tekstkader 1

Dit voorbeeld laat zien dat het inzetten van formatieve leeractiviteiten docenten niet per definitie informatie oplevert die helpt om beter onderbouwde beslissingen te kunnen nemen; in dit geval de beslissing of leerlingen voldoende zijn voorbereid om een toets te kunnen maken of niet. Terwijl je dit, op basis van de definitie van Black & Wiliam (2009), waarin expliciet wordt vermeld dat bij formatief evalueren informatie wordt verzameld om beter onderbouwde beslissingen in het onderwijs te kunnen nemen, wel zou verwachten.

Er zijn twee mogelijke oorzaken aan te geven die zouden kunnen verklaren waarom het onderbouwen van intermediate-stake beslissingen met behulp van formatieve leeractiviteiten nog niet goed lukt én die de aanleiding vormen voor het denken over *programma's* van formatieve leeractiviteiten. Op de eerste plaats is dat de wijze waarop normaliter informatie wordt gebruikt ten behoeve van het nemen van beslissingen in het onderwijs en ten tweede is dat de manier waarop informatie voor het nemen van beslissingen vaak wordt verzameld.

Het gebruiken van informatie bij het nemen van beslissingen

In het onderwijs ligt de nadruk bij het verzamelen van informatie over leerlingen vaak op het afleggen van verantwoording over de voortgang. Zoals eerder beschreven, worden voor het vaststellen van een oordeel over de ontwikkeling van individuele leerlingen toetsen in het voortgezet onderwijs vooral modulair gebruikt: ieder onderdeel van een onderwijsprogramma (een hoofdstuk, een onderwijsperiode, et cetera) wordt afgesloten met een toets. In dit traditionele model van toetsen leidt iedere toets tot een zwaarwegende beslissing: zakken of slagen, voldoende of onvoldoende, overgaan of blijven zitten. Er is, met andere woorden, steeds sprake van een high-stake beslissing. In dit model is er in vrijwel alle gevallen sprake van losstaande toetsen waarbij het, vooral voor de leerling, onduidelijk is hoe de toetsen zich tot elkaar verhouden. Wanneer docenten formatieve leeractiviteiten op dezelfde manier benaderen, verliezen deze hun formatieve functie omdat de informatie over een leeractiviteit niet gebruikt wordt om sturing te geven aan het leerproces – wat is de logische volgende stap in het leerproces? – en de informatie uit verschillende leeractiviteiten niet in samenhang gezien wordt om inzicht te krijgen in de ontwikkeling van de leerling.

In relatie tot het nemen van beslissingen betekent dit dat formatieve leeractiviteiten altijd onderdeel zouden moeten zijn van een groter geheel van formatieve leeractiviteiten. Hierbij worden dan op basis van één formatieve leeractiviteit low-stake beslissingen genomen, zoals: welke instructie hebben mijn leerlingen op dit moment nodig? Beslissingen over of je leerlingen zich voldoende ontwikkelen op een leerlijn, worden genomen op basis van informatie uit meerdere leeractiviteiten.

Het verzamelen van informatie voor het nemen van beslissingen

Een tweede mogelijke oorzaak voor de handelingsverlegenheid is de wijze waarop specifieke informatie ten behoeve van beslissingen in het onderwijs doorgaans wordt verzameld. Hierbij kan onderscheid gemaakt worden naar beslissingen die gebaseerd worden op beschikbare informatie (*data-driven decision-making*, de focus ligt op de data) en beslissingen die genomen worden op basis van specifiek hiertoe verzamelde informatie (*decision-driven data-collection*, de focus ligt op de beslissing).

Over het algemeen is er binnen het traditionele toetsstelsel sprake van data-driven decision-making. Dat betekent dat op basis van beschikbare data, zoals (summatieve) toetsresultaten, beslissingen worden genomen. Als je deze benadering ook zou gebruiken bij formatief evalueren kijk je pas op het moment dat je een beslissing gaat nemen of je alle benodigde informatie hebt om iets over de ontwikkeling van leerlingen te kunnen zeggen. Dit zorgt ervoor dat je op dat moment misschien niet de juiste of niet voldoende informatie hebt om een beslissing te kunnen nemen. De nadruk zou bij formatief evalueren dan ook niet moeten liggen op het gebruiken van beschikbare informatie uit de losstaande formatieve leeractiviteiten, maar op het bewust inzetten van formatieve leeractiviteiten gebaseerd op de beslissing die genomen moet worden: als de docent vooraf weet wat voor beslissing hij of zij wil nemen (bijvoorbeeld het bijstellen van het onderwijs of het vaststellen van de beheersing van een leerdoel) is het noodzakelijk om relevante formatieve leeractiviteiten in te zetten die samen de juiste informatie opleveren om die beslissing daadwerkelijk onderbouwd te kunnen nemen (Wiliam, 2013). Er zou, met andere woorden, niet gewerkt moeten worden vanuit het principe van data-driven decision-making, maar vanuit decision-driven data-collection: *“This is important, because a focus on data-driven decision-making emphasizes the collection of data first without any particular view about the claims they might support, so the claims are therefore accorded secondary importance. By starting with the decisions that need to be made, only data that support the particular inferences that are sought need be collected”* (Wiliam, 2014, p. 6). Op deze manier ligt, met andere woorden, de focus op het bewust verzamelen van informatie ten behoeve van de beslissing die genomen moet worden, en op het in samenhang bezien van alle verkregen informatie.

Denken vanuit programma's van formatieve leeractiviteiten

Beide verklaringen voor de handelingsverlegenheid die docenten ervaren bij formatief evalueren zijn aanleiding om na te denken over programma's van formatieve leeractiviteiten. Samen vragen ze om samenhangende leeractiviteiten die bewust worden gekozen met het oog op beslissingen in onderwijs en leren. We spreken over een programma van formatieve leeractiviteiten als er sprake is van het bewust inzetten van een combinatie van leeractiviteiten die passend is voor het leerdoel dat centraal staat, waarin de onderlinge samenhang tussen de leeractiviteiten duidelijk is en waarmee de juiste informatie verzameld wordt om een zorgvuldige beslissing te kunnen nemen.

In Tekstkader 2 staat beschreven hoe leraar Sjors hier invulling aan zou kunnen geven.

De schoolleiding van Sjors (docent aardrijkskunde) heeft besloten formatief te gaan evalueren. Sjors staat daar achter: zijn ervaring is dat leerlingen alleen leren voor een cijfer, en dat het móéten halen van goede cijfers veel stress oplevert. Sjors krijgt de volgende instructie: “Per onderwijsperiode mag je één toets afnemen waarvoor je een cijfer geeft. In de voorbereiding naar deze toets toe mogen alleen formatieve leeractiviteiten worden ingezet. Geen cijfers dus!”

Vol goede moed gaat Sjors aan de slag. Gebaseerd op de leerdoelen voor het schooljaar bepaalt Sjors welke leerdoelen centraal staan in de komende onderwijsperiode. Hij vertaalt deze leerdoelen naar een aantal tussendoelen, zodat hij ook gedurende de onderwijsperiode kan nagaan of de ontwikkeling van de leerling goed verloopt. Vervolgens bepaalt Sjors welke formatieve activiteiten, gebaseerd op de tussendoelen en de leerdoelen, het beste ingezet kunnen worden.

Sjors besluit iedere les een korte formatieve leeractiviteit in te zetten: soms een exit-kaart, soms een online kennisquiz, soms een opdracht waarbij hij leerlingen samen over een vraag laat nadenken en ze vervolgens random een beurt geeft om antwoord te geven. Doel van deze activiteiten is om in grote lijn zicht te krijgen op de ontwikkeling van de groep en van individuele leerlingen. Hij gebruikt deze informatie om te bepalen of hij iets moet aanpassen aan de volgende geplande les: soms kan hij een onderwerp wat sneller behandelen, soms moet hij een onderwerp opnieuw aan de orde laten komen. Sjors geeft de leerlingen de opdracht voor zichzelf een overzicht te maken van de leerdoelen die ze al goed beheersen en die ze nog niet beheersen. Voor de doelen die ze nog niet goed beheersen gaan ze als huiswerkopdracht in gesprek met een medeleerling die aangeeft het doel al wel te beheersen.

Op driekwart van de onderwijsperiode laat Sjors de leerlingen een formatieve toets maken. Hij bekijkt de antwoorden al even zelf, maar laat de leerlingen in de les erna de antwoorden van elkaar nakijken en onderling bespreken. De leerlingen krijgen de opdracht de door hen gegeven antwoorden te verbeteren waar nodig en om individuele doelen te formuleren die volgens de leerling behaald moeten worden om de toets aan het eind van de onderwijsperiode met een voldoende af te kunnen sluiten.

Aan het eind van de les neemt hij de door de leerlingen verbeterde toetsen weer in en bekijkt deze. Op basis van de antwoorden en de geformuleerde doelen besluit Sjors aan welke onderwerpen hij in het resterende deel van de onderwijsperiode extra aandacht gaat besteden. Ook geeft hij waar nodig leerlingen nog individuele feedback met concrete tips om het gewenste niveau te kunnen behalen.

Als het proefwerk nadert, klopt een enkele individuele leerling zelf nog even aan bij Sjors. “Meneer, de online kennisquiz ging nog niet echt goed. Kunt u dat onderwerp misschien nog een keer uitleggen?” Door alle formatieve leeractiviteiten heeft Sjors een goed beeld gekregen van de ontwikkeling van de leerlingen, de leerlingen hebben dat zelf ook. Het is dan ook voor niemand een verrassing dat 80% van de leerlingen een voldoende haalt.

Tekstkader 2.

Voor docenten betekent het denken vanuit een programma van formatieve leeractiviteiten dat zij bij het ontwerpen daarvan eerst moeten nadenken welke beslissing of beslissingen, low-stake en/of intermediate-stake, ze zouden willen onderbouwen met behulp van informatie uit formatieve leeractiviteiten. Op basis van die toekomstige beslissingen en het moment waarop je die beslissingen wilt nemen, bepaalt de docent vervolgens welke formatieve leeractiviteiten op welk moment een plaats krijgen in het (onderwijs)programma en welke informatie deze activiteiten moeten opleveren. Op deze manier kan recht worden gedaan aan de uitgangspunten van de definitie van Black & Wiliam (2009), en ontstaat een nieuwe definitie: “Er is sprake van een programma van formatieve leeractiviteiten als ten behoeve van een beslissing die genomen moet worden een combinatie van onderling samenhangende leeractiviteiten wordt ingezet. De informatie uit deze

activiteiten wordt verzameld, geïnterpreteerd en gebruikt door leraren, leerlingen en hun klasgenoten om onderbouwde beslissingen te kunnen nemen over de volgende stappen in het onderwijs.”

Vervolg

Zoals benoemd ervaren docenten handelingsverlegenheid bij het ontwerpen en uitvoeren van een programma van formatieve leeractiviteiten. Zij hebben behoefte aan handvatten die hen hierbij ondersteunen. Kennis over voorwaarden waaraan een programma van formatieve leeractiviteiten (1) binnen de scope van het nemen van beslissingen over het leren van leerlingen en (2) binnen de context van het vo moet voldoen, is echter nog niet voorhanden. Vandaar dat in september 2019 een onderzoeksproject is gestart in samenwerking met vier scholen voor voortgezet onderwijs waarin het ontwikkelen van een praktische handreiking centraal staat. Zo'n praktische handreiking voor het ontwerpen van programma's van formatieve leeractiviteiten vraagt een onderzoeksbenadering in twee stappen: allereerst moeten ontwerpprincipes geformuleerd worden waaraan een programma van formatieve leeractiviteiten moet voldoen. Vervolgens moet worden onderzocht welke voorwaarden gesteld moeten worden om de ontwerpprincipes binnen de geldende kaders en de context van het vo te kunnen implementeren. De centrale onderzoeksvraag hierbij is: Wat zijn ontwerpprincipes voor het ontwerpen van een programma van formatieve leeractiviteiten?

Hierbij zijn de volgende deelvragen geformuleerd:

- Welke ontwerpprincipes benoemen experts voor programma's van formatieve leeractiviteiten?
- Hoe waarden toekomstige gebruikers een prototype van de ontwerpprincipes op transparantie, bruikbaarheid, volledigheid en geschiktheid?

Naar verwachting worden de eerste onderzoeksresultaten, waaronder de ontwerpprincipes van de programma's voor formatieve leeractiviteiten, begin 2021 gepubliceerd. Op basis van de onderzoeksresultaten wordt, in samenwerking met de tweedegraads lerarenopleidingen van de Academie Educatie van de Hogeschool van Arnhem en Nijmegen, in studiejaar 2020-2021 een blended cursus 'formatief handelen' ontwikkeld, waarin het werken vanuit de ontwerpprincipes centraal staat. Deze cursus is bedoeld voor zowel leraren-in-opleiding als docenten en lerarenopleiders.

Janneke van der Steen

Janneke van der Steen is onderzoeker bij het Kenniscentrum Kwaliteit van Leren van de Hogeschool van Arnhem en Nijmegen. Haar werkzaamheden concentreren zich op formatief evalueren en onderzoekend vermogen van leraren. Binnen deze thema's voert zij onderzoek- en ontwikkelprojecten uit in scholen en lerarenopleidingen.

janneke.vandersteen@han.nl

Tamara van Schilt-Mol

Tamara van Schilt-Mol is lector Eigentijds Beoordelen en Beslissen bij de hogeschool van Arnhem en Nijmegen. Haar onderzoek focust op de vraag hoe toetsing kan bijdragen aan het stimuleren van leren, het nemen van zorgvuldige beslissingen en het optimaliseren van de kwaliteit van onderwijs.

tamara.vanschiltmol@han.nl

Desirée Joosten-ten Brinke

Desirée Joosten-ten Brinke is hoogleraar Leren van Volwassenen en decaan van de faculteit Onderwijswetenschappen van de Open Universiteit. Haar onderzoeks- en onderwijsfocus is gericht op digitale didactiek, (digitaal) toetsen, het (erkennen van) leren van volwassenen. Zij was tot 2020 lector Technology-enhanced assessment bij Fontys en ontwerper van de masteropleiding Toetsdeskundige.

desiree.joosten-tenbrinke@ou.nl

Cees van der Vleuten

Cees van der Vleuten werkt aan de Universiteit Maastricht, sinds 1996 als hoogleraar in de onderwijskunde. Zijn specialisatie is toetsing van professionele competenties, maar hij is breed betrokken bij onderzoek en ontwikkeling gericht op leren in het gezondheidszorgdomein.

www.ceesvandervleuten.com

Referenties

- Baartman, L., Van Schilt-Mol, T., & Van der Vleuten, C. (2020). *Programmatisch toetsen. Voorbeelden en ervaringen uit de praktijk*. Amsterdam: Boom uitgevers.
- Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability*, 21(1), 1-40.
- 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*, 31(4), 13-17.
- 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.
- Gulikers, J., & Baartman, L. (2017). *Doelgericht professionaliseren. Formatief toetsen met effect! Wat DOET de docent in de klas?* Den Haag: NRO.
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81-112.
- Kingston, N., & Nash, B. (2011). Formative assessment: A meta analysis and a call for research. *Educational measurement: Issues and practice*, 30(4), 28-37.
- Sadler, R. D. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, 18, 119-144.
- 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.
- Simpson-Beck, V. (2011). Assessing classroom assessment techniques. *Active Learning in Higher Education*, 12(2), 125-132.
- Sluijsmans, D., & Kneyber, R. (2016). *Toetsrevolutie. Naar een feedbackcultuur in het voortgezet onderwijs*. Culemborg: Phronese.
- Sluijsmans, D., Joosten-ten Brinke, D., & Van der Vleuten, C. (2013). *Toetsen met leerwaarde. Een reviewstudie naar de effectieve kenmerken van formatief toetsen*. Den Haag: NRO.
- Stobart, G. (2008). *Testing times: The uses and abuses of assessment*. Abingdon, England: Routledge.
- Taras, M. (2007). Assessment for learning: understanding theory to improve practice. *Journal Of Further & Higher Education*, 31(4), 363-371.
- 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 Vleuten, C., Schut, S. & Heeneman, S. (2017). 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., Dijkstra, J., Tigelaer, J., Baartman, L., & Van Tartwijk, J. (2012). A model for programmatic assessment fit for purpose. *Medical Teacher*, 34, 205-214.

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, Meijer, K., Dobbelaer, M., Linden van der, L., & Munneke, L. (2020). *Leeruitkomsten bke/ske. 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.

Wiliam, D. (2013). *Cijfers geven werkt niet*. Meppel: Ten Brink.

Wiliam, D. (2014). *Formative assessment and contingency in the regulation of learning processes*. Paper presented in a Symposium entitled Toward a Theory of Classroom Assessment as the Regulation of Learning at the annual meeting of the American Educational Research Association, Philadelphia, PA, April 2014.

Wiliam, D., Lee, C., Harrison, C., & Black, P. (2004). Teachers developing assessment for learning: Impact on student achievement. *Assessment in Education: Principles, Policy & Practice*, 11(1), 49-65.

Wiliam, D. & Leahy, S. (2018). *Formatieve assessment integreren in de praktijk*. Rotterdam: Bazalt Educatieve Uitgaven.