

Leren van Toetsen

Een cyclisch proces

Fabienne van der Kleij | Jorine Vermeulen | Theo Eggen en Bernard Veldkamp
(Cito | RCEC)

Leren van Toetsen

Een cyclisch proces

Auteurs: Fabienne van der Kleij | Jorine Vermeulen | Theo Eggen en Bernard Veldkamp (Cito/RCEC).

september 2017 (herziene versie), oorspronkelijke uitgave 2013

© ToetsWijzer | Stichting Cito Instituut voor Toetsontwikkeling

info@toetswijzer.nl

www.toetswijzer.nl

Inhoud

INLEIDING	3
1 FORMATIEVE VERSUS SUMMATIEVE TOETSING	4
2 OPBRENGSTGERICHT WERKEN	6
3 ASSESSMENT FOR LEARNING	11
4 DIAGNOSTISCH TOETSEN	14
5 SAMENVATTING VAN DE BELANGRIJKSTE KENMERKEN VAN OGW, AFL EN DT	18
6 OVEREENKOMSTEN EN VERSCHILLEN	19
7 EEN AANZET TOT EEN PROGRAMMA VOOR FORMATIEVE TOETSING	20
8 LITERATUUR	21
9 COLOFON	23

Inleiding

In de afgelopen decennia is er steeds meer aandacht gekomen voor de rol van toetsen in het leerproces. Toetsen zijn een cruciaal onderdeel van het onderwijs, toetsen maken het mogelijk om na te gaan of een bepaalde instructieactiviteit heeft geleid tot het realiseren van de beoogde leeropbrengsten. In de klas worden toetsen veelvuldig gebruikt om informatie te verzamelen over de leeropbrengsten van leerlingen.

Assessment

Onder een toets verstaan we “een instrument voor het meten van iemands kennis en vaardigheden (praktische vaardigheden en houdingen) die door middel van studie en/of onderwijs op een of ander vakgebied zijn verworven” (Cito [toetstechnische begrippenlijst](#), 2010). Naast toetsen kan ook met behulp van andere instrumenten bepaald worden wat leerlingen kennen en kunnen. Zo kan gebruik gemaakt worden van bijvoorbeeld huiswerk, projecten, discussies en observaties. Wanneer een breed spectrum aan technieken wordt gebruikt voor het verzamelen van informatie over leeropbrengsten spreekt men ook wel van assessment.

Formatieve en summatieve toetsing

Traditioneel wordt in de onderwijswereld onderscheid gemaakt tussen formatieve en summatieve functies van toetsing. Kort gezegd hebben summatieve toetsen als functie de leerprestaties van een leerling te beoordelen en vervolgens een beslissing te nemen met betrekking tot selectie, classificatie, plaatsing of certificering ([Sanders](#), 2011). Er zijn echter ook toetsen die als doel hebben inzicht te krijgen in het onderwijsleerproces en dit waar nodig bij te sturen. Deze toetsen worden formatieve toetsen genoemd. Formatieve toetsen hebben als doel informatie te geven aan de leerkracht en leerlingen over de mate waarin de leerstof al wel of niet beheerst wordt.

Over deze ToetsSpecial

Deze special heeft betrekking op het gebruik van toetsen voor formatieve doeleinden. In de afgelopen jaren zijn er veel begrippen ontstaan die samenhangen en deels samenvallen met het begrip formatieve toetsing. Veelgebruikte begrippen zijn opbrengstgericht werken (OGW), assessment for learning (AfL) en diagnostisch toetsen (DT). Het doel van deze special is om de lezer wegwijs te maken in de betekenis van deze begrippen, de overeenkomsten en verschillen tussen deze begrippen te verduidelijken en een beeld te schetsen van de betekenis van deze begrippen voor de onderwijspraktijk in de klas.

Deze ToetsSpecial is geschreven vanuit de context van het primair onderwijs, maar de beschreven benaderingen kunnen in alle vormen van onderwijs worden toegepast.

De special is gebaseerd op het artikel ‘[Integrating data-based decision making, assessment for learning and diagnostic testing in formative assessment](#)’ geschreven door Van der Kleij, Vermeulen, Schildkamp en Eggen (2015).

1.

1. **Formatieve versus summatieve toetsing**

In de **ToetsSpecial [Toetsconstructieproces in 8 stappen](#)** zijn de mogelijke functies van toetsen beschreven

Summatieve toetsen

Toetsen kunnen bijvoorbeeld een selectiefunctie, kwalificerende functie of een prognostische functie hebben. Deze toetsen worden afgenomen na een bepaalde periode om te beoordelen in hoeverre een leerling de beoogde leerdoelen in voldoende mate beheerst. Op basis van de score op een dergelijke toets wordt een (deel van de) zak-/slaagbeslissing genomen. Men zegt in dit geval dat de toets een summatieve functie heeft. Zie voor praktische voorbeelden van summatieve toetsen in verschillende onderwijssectoren hoofdstuk 1 van de publicatie [Toetsen op School](#) (Sanders, 2011).

Formatieve toetsen

Formatieve toetsen daarentegen hebben als doel de leerkracht en leerlingen te informeren over de mate waarin de leerstof beheerst wordt. Op basis van de resultaten op deze toetsen kan worden besloten om het onderwijs aan de groep of de individuele leerling aan te passen. Formatieve toetsen hebben daarom een didactische functie. Wat inhoudt dat ze informatie verschaffen over het onderwijsleerproces en hier sturing aan geven (zie de **ToetsSpecial [Toetsconstructieproces in 8 stappen](#)**). Mede vanwege deze didactische functie wordt van formatieve toetsen beweerd dat deze een positieve invloed kunnen hebben op de leeropbrengst van leerlingen. Er is alleen sprake van formatieve toetsing wanneer de resultaten worden gebruikt voor het geven van feedback en/of het aanpassen van de instructie.

Summatieve of formatieve functie

Een toets is niet per definitie formatief of summatief. Het gaat erom hoe de toetsresultaten gebruikt worden, met andere woorden; de functie van de toetsresultaten. Wanneer het resultaat op een toets meetelt bij het maken van een (deel van de) zak-/slaagbeslissing vervult deze een summatieve functie. Dezelfde toets kan echter ook een formatieve functie vervullen. Dit is bijvoorbeeld het geval wanneer er feedback gegeven wordt aan de leerling, die de leerling kan gebruiken in het vervolgtraject van het onderwijs. Een ander voorbeeld is een leerkracht die toetsresultaten gebruikt om de effectiviteit van zijn of haar instructie te evalueren. Wanneer bijvoorbeeld uit de toetsresultaten blijkt dat het merendeel van de klas een onderdeel van de leerstof niet heeft begrepen, kan de leerkracht concluderen dat hij/zij extra aandacht aan dat deel van de leerstof moet besteden.

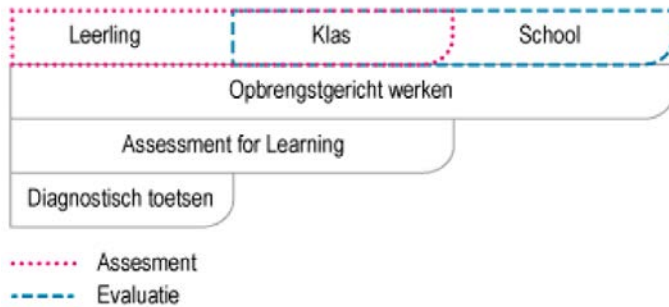
Formatieve toetsing

Formatieve toetsing is een breder begrip dan hierboven is beschreven. De term is geïntroduceerd om de tussentijdse evaluatie van programma's (interventies) te beschrijven (Scriven, 1967). In 1968 is de term formatief voor het eerst gebruikt in de context van instructie. In de loop der jaren zijn er verschillende definities toegekend aan formatieve toetsing.

Er kan onderscheid worden gemaakt tussen formatieve programma-evaluaties en formatieve assessments (Shepard, 2005). Formatieve programma-evaluaties zijn bedoeld om op een hoger aggregatieniveau dan de leerling (klas, school) beslissingen te nemen over de onderwijsbehoeften van groepen leerlingen. Formatief assessment is gericht op het leerling- en klasniveau; het wordt gebruikt om instructie te laten aansluiten op individuele onderwijsbehoeften van leerlingen. In deze special zullen we spreken van formatieve evaluatie als het evalueren van de kwaliteit van het onderwijs wordt bedoeld. Van formatief assessment is sprake als het toetsproces gericht is op leerprocessen binnen een klas en het verbeteren van instructie aan de klas en individuele leerlingen.

Niveaus van assessment en evaluatie

Opbrengstgericht werken valt zowel onder formatieve programma-evaluatie als onder formatief assessment. Assessment for learning valt onder formatief assessment. Diagnostisch toetsen valt onder formatief assessment, maar gaat dieper in op onderwijsbehoeften van leerlingen dan assessment for learning. De niveaus waarop deze drie benaderingen zijn gericht, zijn weergegeven in onderstaande figuur. Het schoolniveau in het figuur op de volgende pagina verwijst naar een school of een schoolbestuur.



Niveaus van assessment en evaluatie in opbrengstgericht werken, assessment for learning en diagnostisch toetsen.

2.

Opbrengstgericht werken

Een manier om de opbrengsten van het onderwijs te verhogen, is door intensiever gebruik te maken van toetsresultaten. Dit houdt in dat het onderwijs zou moeten worden gestuurd door de opbrengsten van het leren zoals gedemonstreerd in toetsen en dagelijkse leeractiviteiten. Dit wordt ook wel opbrengstgericht werken genoemd. Dit hoofdstuk gaat in op de betekenis van het begrip opbrengstgericht werken, wat zal worden afgekort tot OGW.

Definities van OGW

OGW houdt in dat gegevens over opbrengsten worden geanalyseerd en vervolgens benut om het onderwijs af te stemmen op de behoeften van de leerlingen in relatie tot het behalen van de beoogde leeropbrengsten. De term OGW vindt zijn oorsprong in het Engelse data-driven decision making (bijvoorbeeld beschreven door Fullan, 2007). Dit begrip is relatief nieuw, maar de gedachte er achter is dat niet. Er wordt bijvoorbeeld al lang gebruik gemaakt van data over leerlingprestaties, ook wel datafeedback genoemd, om te komen tot beslissingen over de vervolgstappen in het instructieproces (Morrison, Ross, & Kemp, 2007).

In deze Special wordt een combinatie van de volgende definities van OGW gehanteerd:

“Het systematisch, doelgericht werken aan het maximaliseren van prestaties” (Inspectie van het Onderwijs, 2010).

OGW wordt ook wel meetgestuurd onderwijs genoemd. Qua betekenis komt deze term grotendeels overeen met OGW en data-driven decision making. Meetgestuurd onderwijs kan worden gedefinieerd als:

- *“Onderwijs waarbij betrokkenen (scholen, leerkrachten, leerlingen) zich in hun taakuitvoering laten sturen door de uitkomsten van metingen” (Ledoux, Blok, Boogaard, & Krüger, 2009, p. 9).*

Ledoux et al. (2009) geven de voorkeur aan de term meetgestuurd onderwijs boven OGW omdat in de term OGW geen verwijzing zit naar het systematische gebruik van feedback. Met feedback wordt in deze context de uitkomst van metingen van leerlingprestaties bedoeld. In de definitie van de Inspectie van het Onderwijs komt echter wel terug dat OGW een systematisch proces is. Daarentegen wordt er niet gespecificeerd wie de betrokkenen zijn.

Wat is OGW?

OGW omvat een systematisch en cyclisch proces, waarin steeds dezelfde stappen worden uitgevoerd: plannen, uitvoeren, evalueren en handelen. Het evalueren van het onderwijs wordt bij voorkeur gebaseerd op betrouwbare en objectieve gegevens. Deze evaluatie noemt men het toepassen van datafeedback. Er is sprake van OGW wanneer deze datafeedback wordt gebruikt om beslissingen te nemen over de inrichting van het onderwijs (Ledoux et al., 2009).

Zoals blijkt uit de voorgaande paragraaf zijn er verschillende definities van OGW in omloop. Er zijn tussen de verschillende definities echter wel overeenkomsten te vinden. In alle gevallen gaat het namelijk om:

- een systematisch en cyclisch proces,
- met als doel de leeropbrengsten van leerlingen te verhogen.

Wat opvalt aan de definitie van de Inspectie van het Onderwijs is dat wat onder prestaties verstaan wordt zeer breed te interpreteren is. Verder zijn de gegevens over opbrengsten die zouden moeten worden gebruikt om deze prestaties te bereiken niet gespecificeerd.

Aan de ene kant bestaat de visie dat binnen OGW enkel gebruik gemaakt moet worden van data over leerlingprestaties op gestandaardiseerde en genormeerde toetsen, om de kwaliteit van de genomen beslissing te waarborgen.

Aan de andere kant bestaat de visie dat binnen OGW diverse vormen van assessment gebruikt kunnen worden, zoals observaties, methodegebonden toetsen, beoordelingen van schrijfproducten, etc.

In definities over OGW zijn bovendien verschillende meningen terug te vinden over de mate waarin metingen objectief en formeel moeten zijn. De mate waarin de gebruikte datafeedback afkomstig moet zijn uit objectieve en kwalitatief hoogwaardige bronnen hangt af van de consequenties van de te nemen beslissing. Hiermee verschilt de frequentie waarmee de evaluatieve cyclus doorlopen kan worden.

Punten waarop de definities verschillen zijn de niveaus waarop deze gericht zijn, bijvoorbeeld op de individuele leerling, het klasniveau of op het schoolniveau.

Op leerling- en klasniveau is OGW voornamelijk gericht op het aanpassen van de instructie en didactiek aan de behoeften van leerlingen. Door het instructie-aanbod in didactisch opzicht aan te passen op de behoeften van de leerlingen kan voor elke leerling een maximale leerwinst worden bereikt. Hierbij gaat speciale aandacht uit naar de behoefte van opvallende leerlingen. Opvallende leerlingen zijn niet alleen leerlingen die onvoldoende voortgang maken, maar ook leerlingen die juist veel beter presteren dan het gemiddelde niveau dat van hen verwacht mag worden. Hiervoor worden bijvoorbeeld gegevens uit een leerlingvolgsysteem gebruikt, aangevuld met informatie verkregen met methodegebonden toetsen. Deze gegevens worden voornamelijk verzameld met toetsen met een formatief doel, aangezien deze toetsen de leerkracht en leerlingen informatie verschaffen over de mate waarin de leerstof beheerst wordt. Dit valt onder *formatief assessment*.

Op schoolniveau heeft OGW betrekking op kwaliteitszorg in brede zin, bijvoorbeeld schoolzelfevaluaties. De datafeedback die op schoolniveau kan worden verzameld kan afkomstig zijn uit een breed spectrum aan bronnen, zoals leerlingprestaties op een centrale eindtoets, schoolzelfevaluatie-rapportages, of vragenlijsten. Op schoolniveau kunnen dus ook toetsen met een summatief doel waardevolle informatie verschaffen. Wanneer datafeedback op schoolniveau wordt gebruikt om beslissingen te nemen over de inrichting van het onderwijs, spreken we van *formatieve evaluatie*. Daarentegen is er bijvoorbeeld sprake van summatieve evaluatie wanneer deze data worden gebruikt voor kwaliteitsbeoordelingen door de Inspectie van het Onderwijs.

Verder verschilt de aard van de datafeedback die verzameld kan worden op verschillende niveaus. Leerkrachten zullen bijvoorbeeld andere datafeedback verzamelen dan de schooldirectie. Uiteraard vindt er interactie plaats tussen de verschillende niveaus binnen de school en zullen de beslissingen op deze niveaus elkaar beïnvloeden (Mandinach, Honey, & Light, 2006).

Uit het onderzoek van Ledoux et al. (2009) blijkt dat OGW een positief effect kan hebben op de leeropbrengsten van leerlingen. De redenen hiervoor zijn dat OGW het mogelijk maakt om nauwkeuriger zwakke plekken te signaleren in het leerproces van de leerling, de resultaten te vergelijken met die van andere leerlingen (of andere klassen en scholen), of met een referentieniveau. Doordat hierdoor verbeterd inzicht in de zwakke plekken ontstaat, kunnen meer gericht verbeteringsmaatregelen genomen worden. Het gaat er dus om dat leerkrachten weten wat een leerling wel of niet kan om vervolgens het aanbod van de instructie daarop af te stemmen. Dit vergt naast een goede interpretatie van waar nog onderwijsbehoeften liggen ook de competentie om het onderwijsaanbod op basis van verschillen in onderwijsbehoeften te differentiëren.

Data-driven decision making en data-driven teaching

OGW vindt zijn oorsprong in data-driven decision making (datagestuurde beslissingen nemen) en data-driven teaching (datagestuurd lesgeven). Deze twee begrippen zijn nauw gerelateerd, maar datadriven teaching is voornamelijk gericht op het klasniveau, terwijl data-driven decision making alle niveaus in de

school behelst. Om meer duidelijkheid te verschaffen over de kenmerken van OGW gaan we kort in op de begrippen data-driven decision making en data-driven teaching en de totstandkoming hiervan. Traditioneel wordt het onderwijs gezien als een veld waarin leerkrachten op basis van hun intuïtie en instinct beslissingen nemen (Slavin, 2002). De No Child Left Behind Act (NCLB) in de Verenigde Staten heeft er echter voor gezorgd dat het gebruik van data om de schoolpraktijk te informeren aan populariteit heeft gewonnen. De gedachte achter NCLB is dat door standaarden te bepalen en meetbare doelen te stellen de leeropbrengsten van leerlingen verhoogd kunnen worden. Onderzoek heeft aangetoond dat het stellen van nauwkeurige leerdoelen kan leiden tot prestatieverbetering.

In Nederland hebben de beleidsveranderingen uit de Verenigde Staten geleid tot een nieuwe onderwijsvisie: leerlingen moeten, ongeacht eventuele speciale onderwijsbehoeften, de kans krijgen naar dezelfde school te gaan als leerlingen zonder speciale onderwijsbehoeften. Deze visie is onder andere terug te vinden in onderwijsbeleid zoals Weer Samen Naar School (Ministerie van OCW, 1990) en Passend Onderwijs (Ministerie van OCW, 2009). Door deze veranderingen wordt in het reguliere onderwijs belang gehecht aan het aansluiten op de onderwijsbehoeften van alle leerlingen in plaats van alleen zorgleerlingen (leerlingen met speciale onderwijsbehoeften).

Schildkamp en Kuiper (2010) definiëren data-driven decision making als "het systematisch analyseren van bestaande databronnen in de school, uitkomsten van analyses toepassen om instructie, curricula en schoolprestaties te innoveren en implementeren (bijvoorbeeld verbeteringsacties uitvoeren) en deze innovaties te evalueren" (p. 482). Deze databronnen zijn niet alleen toetsresultaten van leerlingen. Bruikbare data zijn bijvoorbeeld schoolzelfevaluatie-data, data over de kenmerken van leerlingen in de school, leerresultaten op een centrale eindtoets, leerresultaten op methodegebonden toetsen en de resultaten uit vragenlijsten afgenomen aan ouders of leerlingen (Schildkamp & Kuiper, 2010). De term OGW zoals deze in de Nederlandse context gebruikt wordt ligt ergens tussen datadriven decision making en data-driven teaching, waarin wordt nagestreefd dat in de school als geheel een meer onderzoekende werkwijze wordt gehanteerd. In Nederland ligt vooral nadruk op leerlingprestaties als bron van datafeedback, terwijl in data-driven decision making ook gebruik gemaakt wordt van andere bronnen van datafeedback, zoals resultaten uit een vragenlijstonderzoek.

Het uitgangspunt van zowel OGW als data-driven decision making en data-driven teaching is dat datafeedback kan worden gebruikt om goed geïnformeerde beslissingen te nemen over het onderwijs. De volgende paragraaf zal hier op ingaan.

Van datafeedback naar beslissingen over het onderwijs

Veel scholen hebben beschikking over objectieve gegevens over de leerprestaties en de leervorderingen van leerlingen, bijvoorbeeld uit een leerlingvolgsysteem. Ook beschikken leerkrachten over datafeedback die in de dagelijkse praktijk met verschillende assessments wordt verzameld. Zoals al eerder aangegeven is er sprake van OGW wanneer datafeedback daadwerkelijk wordt gebruikt om beslissingen te nemen over de inrichting van het onderwijs. In de Nederlandse scholen wordt datafeedback echter nog onvoldoende benut (Ledoux et al., 2009). De beschikbare gegevens worden dus nog in onvoldoende mate gebruikt om gericht veranderingen door te voeren in het onderwijsprogramma om aan de wensen en noden van de leerlingen te voldoen. Wanneer datafeedback goed benut wordt zal dit leiden tot onderwijs dat beter is afgestemd op de behoeften van leerlingen en uiteindelijk tot betere leerresultaten.

Bij OGW speelt de *evaluatieve cyclus* in het onderwijsproces een grote rol. Of in de woorden van Ledoux et al. (2009): "Het gebruik van vorderingengegevens is onlosmakelijk verbonden met de evaluatieve cyclus" (p. 21). Zij geven aan dat in de praktijk de evaluatieve cyclus vaak niet in zijn geheel wordt doorlopen. Hierbij ontbreekt volgens hen het stellen van doelen en vastleggen van standaarden, het interpreteren van data en het nemen van beslissingen. Verder stellen zij dat de kwaliteit van de evaluatieve cyclus rechtstreeks afhangt van de kwaliteit van de datafeedback die is verzameld. Dit betekent dan ook dat onbetrouwbare datafeedback ertoe kan leiden dat onjuiste beslissingen worden genomen. Maar ook de onjuiste interpretatie van betrouwbare data kan leiden tot onterechte beslissingen (Visscher & Ehren, 2011). Het juist kunnen interpreteren van de datafeedback is dus een voorwaarde voor een succesvolle invoering van OGW.

Verder is een risico van het gebruik van een vaste evaluatieve cyclus dat men in een mechanistische aanpak beland zonder aandacht voor de eigen context.

Gebruik maken van datafeedback houdt in dat scholen allerlei bronnen van informatie benutten om zelf te onderzoeken wat wel en niet werkt. Light, Wexler en Heinze (2004) en Madinach et al. (2006) beschrijven het proces van datafeedback naar beslissingen over het onderwijs: Data zijn ruw materiaal, bijvoorbeeld scores van leerlingen op een toets. Of data worden omgezet naar informatie is in de handen van de persoon die de data bestudeert. Nadat de data zijn verzameld moeten deze op een of andere manier worden samengevoegd en gepresenteerd. Dit kan bijvoorbeeld betekenen een rapportage maken om de betekenis van de data eruit te halen, oftewel, data omzetten in informatie. Een leerlingvolgsysteem kan hierbij nuttige ondersteuning bieden.

Vervolgens worden patronen in de data geanalyseerd en de informatie die hieruit wordt gehaald samengevat. In deze fase vindt interpretatie van de data plaats en worden wanneer nodig diepteanalyses uitgevoerd. Ook kan het in deze fase nodig zijn om meer data te verzamelen, bijvoorbeeld met behulp van een diagnostisch toetsinstrument. De verkregen informatie wordt omgezet in kennis.

Om dit te doen moet de verzamelde informatie worden gesynthetiseerd, waarna wordt bepaald welke acties nodig zijn. Dit leidt tot het nemen van beslissingen, die worden uitgevoerd en waarvan vervolgens de opbrengsten worden gemeten. Met andere woorden, er wordt weer teruggegaan naar de eerste stap in het proces. Op deze manier vormt zich een feedbackloop. Binnen de school kunnen veel verschillende feedbackloops ontstaan. De frequentie waarmee een feedbackloop wordt doorlopen is onder andere afhankelijk van het gebruikte type datafeedback. Datafeedback uit formele en objectieve toetsen, bijvoorbeeld uit een leerlingvolgsysteem, zijn minder frequent beschikbaar dan datafeedback uit informele assessment situaties, bijvoorbeeld uit een methodegebonden toets.

Strategieën voor OGW

Zoals al eerder aangegeven speelt de evaluatieve cyclus een grote rol in de systematische benutting van gegevens over leerprestaties van leerlingen. In deze paragraaf wordt verder ingegaan op de evaluatieve cyclus. Door Ledoux et al. (2009) wordt deze cyclus beschreven in vijf stappen:

1. Vastleggen van doelen en standaarden
2. Informatie (data) verzamelen
3. Verzamelde gegevens registreren
4. Resultaten interpreteren
5. Beslissingen nemen

Om een praktisch voorbeeld te nemen zullen wij de cyclus zoals beschreven door Ledoux et al. (2009) doorlopen in de rol van de leerkracht met als doel beslissingen te nemen naar aanleiding van de resultaten op een toets.

Bij **stap één** worden door de school zelf doelen en standaarden vastgelegd of worden extern opgestelde doelen en standaarden gehanteerd.

In **stap twee** wordt informatie verzameld over de leervoortgang van leerlingen, dit is dus de afname van de toets. Zoals is gebleken uit de voorgaande paragraaf kunnen we in deze fase eigenlijk nog niet spreken van informatie. Het zou beter zijn om te spreken van het verzamelen van data, immers, informatie ontstaat pas wanneer er betekenis wordt gegeven aan de data (dit vindt plaats in stap vier van de cyclus). Data zijn in dit voorbeeld toetsscores van leerlingen.

Vervolgens worden in **stap drie** de data (toetsscores) geregistreerd. Bijvoorbeeld het invoeren van toetsresultaten in een leerlingvolgsysteem.

De **vierde stap** is de interpretatie van de verzamelde toetsscores. Bij het interpreteren van de toetsscores gaat het erom de geobserveerde scores te vergelijken met een standaard. In deze stap wordt er kritisch gekeken naar de eigen onderwijsopbrengsten. Samenwerking met collega's kan hierbij nuttig zijn. De vierde

fase resulteert in inzicht in de betekenis van de toetscores voor de eigen onderwijscontext (informatie) en patronen die hierin zichtbaar zijn.

In de **vijfde stap** wordt tenslotte de informatie omgezet in kennis en worden beslissingen genomen over de inrichting van het onderwijs. Een mogelijke actie van een leerkracht na het doorlopen van de vijf stappen in de evaluatieve cyclus is extra aandacht besteden aan een bepaald onderwerp of bepaalde leerlingen extra instructie te geven. De effectiviteit van de genomen beslissing wordt vervolgens geëvalueerd door de cyclus opnieuw te doorlopen, wanneer dit niet gedaan wordt is de feedbackloop niet compleet. De kwaliteit van de cyclus hangt af van de kwaliteit van de data die zijn verzameld. Dit wil zeggen dat onbetrouwbare data kunnen leiden tot het nemen van verkeerde beslissingen. De kwaliteit van de verzamelde data (in stap twee) is dus noodzakelijk voor een goed verloop van de rest van het proces. Afhankelijk van hoe belangrijk de beslissing is die genomen moet worden, kunnen meer of minder strenge eisen worden gesteld aan de kwaliteit van de data. Ook is het belangrijk om elke stap in de evaluatieve cyclus volledig te doorlopen.

Wanneer er bijvoorbeeld geen duidelijke doelen en standaarden zijn zal dit de interpretatie van de toetsresultaten bemoeilijken. Uit onderzoek is gebleken dat scholen vooral moeite hebben met het uitvoeren van stap één, vier en vijf van de cyclus (Ledoux et al., 2009).

Praktische voorwaarden bij OGW

Visscher en Ehren (2011) hebben vier voorwaarden geformuleerd waaraan voldaan moet worden om bovenstaande kerncomponenten succesvol in te voeren:

1. Er is een instrument beschikbaar waarmee de onderwijsresultaten (leeropbrengsten van leerlingen) op valide en betrouwbare wijze kunnen worden gemeten. Aangezien deze data de basis vormen voor de te maken instructiebeslissingen is de kwaliteit van de gebruikte instrumenten van groot belang.
2. Er wordt een vaste cyclus gevolgd. De effectiviteit van de gekozen instructiebenadering wordt geëvalueerd in een volgende meting. Tijdens de uitvoering van de cyclus is het monitoren van de voortgang van belang. Hierbij kunnen de prestaties van leerlingen op methodegebonden toetsen nuttige inzichten opleveren.
3. Er is tijd beschikbaar om de vernieuwingen te ontwikkelen en in te voeren. Ook is er tijd beschikbaar voor het uitvoeren van nieuwe werkzaamheden.
4. Er is een goede afstemming tussen instructie en toetsmaterialen, wat de koppeling tussen datafeedback en instructiebeslissingen vergemakkelijkt.
5. Wanneer scholen opbrengstgericht gaan werken zal dit van de medewerkers binnen de school het volgende eisen (Visscher & Ehren, 2011):
 - Men zal zich moeten verdiepen in nieuwe werkwijze.
 - Men zal nieuwe kennis moeten opdoen en deze ook leren toepassen binnen de werkwijze van OGW.
 - Binnen het schoolteam zullen precieze en realistische doelen moeten worden gesteld en de behaalde resultaten moeten worden geëvalueerd aan de hand van de gestelde doelen.
 - Men zal moeten leren om kritisch naar de eigen resultaten te kijken en een relatie te leggen tussen het gegeven onderwijs en de behaalde resultaten.

3. Assessment for learning

In de introductie is de term ‘assessment for learning’ genoemd als een manier om sturing te geven aan het leerproces, met andere woorden, voor formatieve doeleinden. Assessment for learning betekent letterlijk vertaald ‘assessment voor leren’. In de Nederlandstalige literatuur wordt assessment for learning ook wel ‘Evalueren om te leren’ genoemd. In deze special wordt de originele term ‘assessment for learning’ gehanteerd, die afgekort zal worden als AfL. Dit hoofdstuk gaat in op de theorie achter AfL en de belangrijke aspecten die hierbij een rol spelen.

Definities van AfL

Net zoals voor formatieve toetsing zijn er voor AfL verschillende definities in omloop. In deze Special geven we twee veelgebruikte definities van AfL:

1. AfL is “het proces van zoeken naar en het interpreteren van bewijs om gebruikt te worden door lerenden en hun leerkrachten om te beslissen waar lerenden zijn in hun leerproces, waar zij heen moeten en hoe ze daar kunnen komen” (Broadfoot et al., 2002, p. 2-3).
2. “Assessment for learning is onderdeel van de alledaagse praktijk met leerlingen, leerkrachten en medeleerlingen waarin wordt gezocht naar, gereflecteerd op en gereageerd op informatie uit dialogen, demonstraties en observaties op manieren die continu leren bevorderen” (Klenowski, 2009, p. 264).

In de volgende paragrafen zullen deze definities verder worden toegelicht. Vervolgens geven we op basis van deze verschillende definities en strategieën over AfL praktische handvatten.

Wat is AfL?

AfL kan het best worden gezien als een benadering naar toetsing in de klas, het is geen strikt geformuleerde theorie. In AfL is assessment geïntegreerd in de dagelijkse lespraktijk. Assessments hebben als doel bij te dragen aan het leerproces. De resultaten van assessments worden gebruikt om de positie van de leerling in het leerproces te bepalen ten opzichte van de te bereiken leerdoelen. Hierdoor kan het onderwijs beter worden afgestemd op de behoefte van de leerling(en). De nadruk ligt op wat er is geleerd en de kwaliteit van interacties en relaties in de klas.

Het begrip assessment wordt heel breed geïnterpreteerd. Het bewijs over de leerprocessen van leerlingen dat wordt verzameld in AfL kan zowel *kwalitatief* als *kwantitatief* van aard zijn. De resultaten op een toets leveren bijvoorbeeld kwantitatieve informatie op (een score op de toets) maar ook kwalitatieve informatie, bijvoorbeeld welke soort vragen een leerling juist of onjuist heeft beantwoord. Andere bronnen uit de alledaagse praktijk leveren alleen kwalitatieve informatie op, zoals dialogen en observaties. AfL maakt het mogelijk om te anticiperen op zwakkere punten in het huidige leerproces en vervolgstappen te identificeren om de beoogde doelen te bereiken (Assessment Reform Group, 1999). Een cruciaal aspect van AfL is feedback, wat wordt gebruikt om leerlingen te sturen in hun toekomstig leergedrag (Stobart, 2008).

Meerdere onderzoekers beweren dat AfL de leeropbrengsten van leerlingen kan verhogen. De redenen hiervoor zijn dat:

1. leerlingen weten wat ze moeten leren;
2. assessments worden gebruikt om voortgang te bepalen;
3. leerlingen feedback ontvangen.

In AfL ligt er dan ook veel nadruk op het leerproces en niet alleen op het eindproduct (de leeropbrengsten).

De AfL-benadering omvat meer dan alleen een manier om assessment en resultaten op assessments te gebruiken. De Assessment Reform Group (1999) omschrijft AfL aan de hand van vijf kenmerken (p. 145):

- leerlingen zijn actief betrokken bij hun eigen leerproces;
- effectieve feedback wordt gegeven aan leerlingen;
- instructieactiviteiten worden aangepast naar aanleiding van assessment resultaten;
- leerlingen zijn in staat om self-assessment uit te voeren;
- de invloed van assessment op de motivatie en het zelfvertrouwen van leerlingen wordt erkend.

Stobart (2008) benadrukt dat AfL een sociale activiteit is die zowel de identiteit van de leerling en het soort leren dat plaatsvindt beïnvloed. De nadruk ligt op het begrijpen en het verbeteren van het leren. De kern van deze visie is dat leerlingen actief deelnemen aan hun leerproces en dit zelf reguleren. Ook zijn zij in staat om hun eigen werk en dat van anderen te beoordelen. De basis voor het leren vindt plaats in samenwerking, door gedeelde leerintenties en succescriteria. Peer- en self-assessment (waarbij leerlingen hun medeleerlingen en zichzelf beoordelen) spelen een belangrijke rol in AfL, evenals feedback.

Strategieën voor AfL

Uit de omschrijving van AfL door de Assessment Reform Group wordt duidelijk dat leerkrachten effectieve feedback moeten verschaffen aan leerlingen en dat zij hun instructieactiviteiten moeten afstemmen op de behoeften van leerlingen. Aan de andere kant wordt van leerlingen verwacht dat zij actief deelnemen aan hun eigen leerproces en gebruik maken van de feedback die zij ontvangen. Ook moeten zij in staat zijn om hun eigen leren te beoordelen. Black en Wiliam (2009) stellen op basis van de vijf kernfactoren van de Assessment Reform Group (1999) strategieën voor die gebruikt kunnen worden om AfL in de praktijk toe te passen:

1. Maak duidelijk wat de leerdoelen zijn en bediscussieer succescriteria

De eerste strategie is bedoeld om leerlingen inzicht te geven in de te behalen leerdoelen. Op deze manier hebben zij een idee waar zij heengaan in hun leerproces en wat hun huidige positie is met betrekking tot de beoogde doelen (Harlen & Winter, 2004; Stobart 2008). Het gevaar is echter dat deze leeropbrengsten meer aangekondigd dan onderhandeld worden, dit is in het bijzonder het geval wanneer er voor elke les zeer specifieke en vaste doelen worden gehanteerd. Clarke (2001) gebruikt daarom de term 'leerintentie', deze term is zowel flexibel als breed. In AfL is het van belang dat wanneer dit nodig blijkt, kan worden afgeweken van een lesplan om de bredere doelen te kunnen bereiken. Door over leerintenties te onderhandelen en te discussiëren kan de autonomie van de leerling worden bevorderd.

2. Ontwerp effectieve discussies in de klas en andere leertaken waarmee informatie verzameld kan worden over de leervoortgang van leerlingen

De tweede strategie maakt duidelijk dat niet alleen toetsen een middel zijn om informatie te verzamelen over leervoortgang, maar dat ook activiteiten in de klas zoals discussies een belangrijke informatiebron kunnen vormen (Stobart, 2008). De assessmentvormen die worden gebruikt binnen AfL zijn flexibel en interactief en maken het mogelijk om direct in te grijpen wanneer dit nodig is, door bijvoorbeeld feedback en begeleiding.

3. Geef feedback die leerlingen vooruit helpt

De derde strategie houdt in dat er effectieve feedback moet worden gegeven aan leerlingen. Feedback wordt gezien als een kernmechanisme in AfL, omdat het mogelijk maakt om het gat te vullen tussen waar de leerling momenteel is in het leerproces en waar hij of zij naartoe moet (Stobart, 2008). Zie voor een classificatie van verschillende manieren om feedback te geven en de effectiviteit hiervan de ToetsSpecial Feedback in computergestuurde toetsen van Toetswijzer.

4. Laat leerlingen functioneren als een instructiebron voor elkaar

De vierde strategie verwijst naar peer-assessment, dit wil zeggen dat leerlingen elkaar en zichzelf beoordelen in plaats van dat de leerkracht dit doet. Dit vereist zowel inzicht in hoe een goede prestatie eruitziet

(succescriteria) als inzicht in de eigen positie in het leerproces. Deze vaardigheden vormen de basis voor zelfregulatie, ook wel metacognitie genoemd. Wanneer leerlingen in staat zijn hun eigen leren te reguleren zullen zij op een effectieve manier leren.

5. Maak leerlingen de eigenaar van hun leerproces

De vijfde strategie houdt in dat leerlingen actief moeten deelnemen aan een eigen leerproces, ook hier speelt zelfregulatie een belangrijke rol (Black & William, p. 8). Leerlingen leren niet alleen bepaalde leerstof, zij leren ook hoe ze op een effectieve manier kunnen leren.

Benaderingen van AfL

De hierboven genoemde strategieën zijn een voorbeeld van hoe de vijf kernfactoren van de Assessment Reform Group in de praktijk gebruikt kunnen worden. Er zijn echter verschillende benaderingen met betrekking tot AfL. Op basis van een vragenlijst die is afgenomen bij leerkrachten en schoolhoofden in het Verenigd Koninkrijk heeft Hargreaves (2005) onderscheid gemaakt tussen een *meetgerichte* en een *exploratieve* benadering van AfL. Bij de meetgerichte benadering is het doel om inzicht te krijgen in het huidige niveau van de leerling met betrekking tot vooraf opgestelde leerdoelen door gebruik te maken van toetsen. Hierbij horen begrippen zoals monitoring en het aantonen van een niveaubehersing. Binnen deze benadering worden zowel kwantitatieve als kwalitatieve gegevens verzameld en gebruikt. De exploratieve benadering is meer gericht op het proces dan op de uitkomsten van het proces (leeropbrengsten). In deze benadering heeft een leerkracht een reflectiegeoriënteerde en onderzoekende leerbenadering. De nadruk ligt hier hoofdzakelijk op kwalitatieve gegevens. In beide benaderingen wordt feedback, zoals hierboven beschreven, gebruikt voor sturing van het leerproces en/of de leeropbrengsten van leerlingen. In de praktijk ligt de uitvoering van AfL tussen een meetgerichte en exploratieve benadering in.

Praktische voorwaarden bij AfL

Ondanks dat de AfL-beweging al sinds ongeveer 1998 bestaat, is uit verschillende onderzoeken gebleken dat de praktische invoering van AfL-strategieën in de klas niet altijd realiseerbaar blijkt te zijn. Een van de beperkende factoren is bijvoorbeeld de grootte van de klassen. Klassen van 25 tot 30 leerlingen maken het voor leerkrachten moeilijk om het onderwijs vorm te geven op basis van resultaten uit formatieve evaluaties. Visscher en Ehren (2011) geven aan dat de implementatie van AfL jarenlange samenwerking van experts en praktici blijkt te vereisen.

4.

Diagnostisch toetsen

Diagnostisch toetsen is een systematische en cyclische benadering waarmee het leerproces en de leeropbrengsten van individuele leerlingen in kaart gebracht kunnen worden. Aangezien diagnostisch toetsen in verschillende vakgebieden verschillende definities kent en voor verschillende doeleinden wordt gebruikt, wordt in dit hoofdstuk toegelicht hoe diagnostisch toetsen, toegepast binnen het onderwijs, kan bijdragen aan het optimaliseren van het leerproces van lerenden.

Definities van diagnostisch toetsen

We noemen twee definities van diagnostisch toetsen (DT):

1. In algemene zin is het stellen van een diagnose met DT het kritisch analyseren en verklaren van de leeropbrengsten van leerlingen (Rupp, Templin, & Henson, 2010).
2. Keeley en Tobey (2010) stellen dat het bij DT gaat om het meten van voorkennis en verschillende manieren van redeneren.

De voorkennis en manieren van redeneren van een leerling kunnen geïnterpreteerd worden als voorbeelden van verklaringen van de leerprestaties en daarmee informatie geven over de oorsprong van deze leerprestaties. Naast deze leerlingkenmerken kunnen ook taakkenmerken verklaren waarom een leerling iets wel of niet kan. In deze Special hebben we de specifieke definitie van Keeley en Tobey en de globale definitie van Rupp et al. samengevoegd tot de onderstaande definitie: "Het verkrijgen van een gedetailleerde beschrijving van de prestaties van een leerling en de oorsprong van deze prestaties met als doel een advies te geven over hoe deze prestaties gemaximaliseerd kunnen worden."

Wat is DT?

Diagnostiek in het onderwijs was oorspronkelijk bedoeld als middel om leerlingen te identificeren die niet mee konden komen in het reguliere onderwijs en dus behoefte hadden aan speciaal onderwijs (Stobart, 2008). Ook nu wordt diagnostiek in het onderwijs gebruikt voor het identificeren van leerlingen met speciale onderwijsbehoeften als gevolg van leer- en/of gedragsproblemen. Mede als gevolg van veranderingen in het onderwijsbeleid ligt echter niet langer de nadruk op het aansluiten op onderwijsbehoeften van leerlingen met leer- en/of gedragsproblemen, maar op alle leerlingen. Het gaat, met andere woorden, om het maximaliseren van leeropbrengsten in brede zin. Het gebruik van diagnostische toetsen binnen het onderwijs is waardevol voor alle leerlingen omdat naast data over wat leerlingen wel en niet kennen en kunnen ook informatie verzameld wordt over *waarom* een leerling bepaalde taken wel of niet kan. Zo kan een leerkracht bijvoorbeeld achterhalen waarom een leerling bepaalde vragen fout beantwoord heeft. Vervolgens kan de leerkracht deze informatie gebruiken om te zorgen dat de leerling dat type fout niet meer zal maken. Door op een andere manier te toetsen en daarmee gedetailleerdere informatie te verzamelen over de oorsprong van de leeropbrengsten van leerlingen wordt met DT meer diepgang bereikt dan bij andere benaderingen.

Uit onderzoek blijkt dat leerkrachten van allerlei bronnen van informatie vinden dat deze diagnostische waarde hebben (Rupp et al., 2010). Echter, leerkrachten beschouwen niet alle bronnen van informatie als even nuttig voor het inrichten van hun onderwijs. Uit het onderzoek is gebleken dat leerkrachten behoefte hebben aan informatie waarmee ze hun strategieën voor remediëring kunnen bepalen. Dit betekent dat leerkrachten behoefte hebben aan gedetailleerde informatie over het leerproces en de ontwikkeling van leerlingen. In de volgende paragrafen is beschreven hoe je met diagnostische toetsen deze informatie kan verzamelen.

Diagnostische cyclus en functies van diagnoses

Het stellen van diagnoses binnen de hulpverlening aan jongeren en volwassenen is een cyclisch proces waarbij elke stap in een diagnose resulteert. Binnen het onderwijs geeft het volgen van de diagnostische cyclus informatie over het leerproces van leerlingen.

Fasen in de diagnostische cyclus (De Bruyn, Ruijsenaars, Pameijer, van Aarle, 2007):

1. verhelderen
2. onderkennen (verkennen)
3. verklaren
4. indiceren (adviseren)

De fasen in de diagnostische cyclus verschillen in de mate waarin ze gedetailleerde informatie geven over het leerproces. Verondersteld wordt dat fases 3 en 4 resulteren in diagnoses met meer praktische waarde omdat deze informatie leerkrachten beter kan ondersteunen bij het nemen van beslissingen over het inrichten van de leeromgeving dan de diagnoses van fases 1 en 2.

1. Verhelderen: Waar zit de leerling ten opzichte van het hoofddoel of de hoofddoelen?

Binnen het onderwijs wordt doorgaans in doelen beschreven wat leerlingen aan het eind van een bepaalde onderwijsperiode moeten kennen en kunnen. Wanneer de doelen een groot deel van één of meerdere schooljaren beslaan, kan gesproken worden van hoofddoelen. In deze fase is DT gericht op het vaststellen in hoeverre de leerling hoofddoelen heeft bereikt en of bijsturing van het leerproces nodig is. Vanuit deze diagnose kan worden besloten of het nodig is om meer informatie over de oorsprong van de prestaties van de leerling te verzamelen. Omdat de informatie die deze fase oplevert geen informatie geeft over de onderliggende factoren die een rol spelen in de prestatie van de leerling, heeft deze informatie nog niet veel diagnostische waarde voor de leerling en de leerkracht. De resultaten van een dergelijke diagnose zijn te vergelijken met gebruikelijke gestandaardiseerde toetsen die meestal voor summatieve doeleinden gebruikt worden.

2. Verkennen: Hoe presteert de leerling op subdoelen?

In deze fase wordt informatie gegeven over de prestaties van de leerling op subdoelen. De subdoelen worden geformuleerd aan de hand van de hoofddoelen en zijn voorwaardelijke deelvaardigheden en kennisaspecten voor het hoofddoel. Dit principe is gebaseerd op de Piagetiaanse ontwikkelingspsychologie waarin wordt verondersteld dat leerlingen een aantal *ontwikkelingsdrempels* moeten nemen alvorens ze het hoofddoel kunnen bereiken (Verhofstadt-Denève, Van Geert, & Vyt, 2003). Een ontwikkelingsdrempel is een drempel die leerlingen moeten nemen om taken met een grotere complexiteit te kunnen uitvoeren. Het zijn met andere woorden essentiële kennis en/of vaardigheden om complexere taken te kunnen uitvoeren. Bijvoorbeeld bij het leren rekenen zal een leerling bij opgaven tot 10 voldoende hebben aan het één voor één tellen op de vingers, maar naar mate de taken complexer worden (grotere getallen) zullen leerlingen met grotere sprongen moeten gaan tellen.

Deze fase resulteert in een diagnose waarin per subdoel een *sterkte-/zwakteprofiel* van de leerling wordt beschreven. In deze fase kan ook in kaart gebracht worden hoe de leerling op taken met bepaalde kenmerken presteert, op die manier wordt informatie gegeven over de leerdrempels die de leerling wel en niet genomen heeft. In het sterkte-/zwakteprofiel wordt dan gespecificeerd in welke taaksituaties de leerling een bepaalde taak wel of niet kan uitvoeren.

3. Verklaren: Hoe komt het dat de leerling sterk of zwak is binnen dit (sub)domein?

Het doel van deze fase is om met de antwoorden van leerlingen te toetsen welke verklaringen voor de leerprestaties het meest waarschijnlijk zijn voor deze leerling. Er zijn twee hoofdbenaderingen van het verklaren van de leerprestaties van leerlingen. Op de eerste plaats kan gekeken worden naar wat de leerling met hulp, zonder hulp en nog niet (ongeacht de aangeboden hulp) kan. Wanneer met een diagnostische toets in kaart wordt gebracht wat een leerling nog niet zelfstandig, maar wel met hulp kan, kan de leerkracht hierop inspelen door de leeromgeving zo in te richten dat de leerling steeds meer zelfstandigheid verwerft. De tweede benadering is gericht op het diagnosticeren van strategieën en systematische fouten

► **Onderwijsbehoefte en zone van naaste ontwikkeling**

Vygotsky vond dat om een leerling vooruit te kunnen helpen in zijn ontwikkeling, het vooral van belang is om vast te stellen wat een kind kan leren in plaats van te meten wat een kind kan (Verhofstadt-Denève et al., 2003). Er moet met andere woorden vastgesteld worden wat de Zone van Naaste Ontwikkeling (ZNO) van de leerling is. Met de ZNO bedoelt Vygotsky datgene dat de leerling wel kan met hulp van volwassenen of leeftijdsgenoten die verder zijn in hun ontwikkeling, maar nog niet zelfstandig. Door het uitvoeren van taken met hulp van anderen krijgen leerlingen de gelegenheid om kennis en vaardigheden te ontwikkelen waardoor ze op den duur in staat zijn de taken zelfstandig uit te voeren. Naast hulp van anderen kunnen leerlingen sommige taken wel met een hulpmiddel uitvoeren, bijvoorbeeld een getallenlijn bij het maken van rekenopgaven.

► **Oplossingsstrategieën en juistheid van antwoorden**

Bij DT ligt de focus niet alleen op de juistheid van de antwoorden die leerlingen geven maar ook op de oplossingsstrategieën die leerlingen gebruiken om taken te maken (Cito, 2011; Kraemer, 2011). Deze oplossingsstrategieën en het type fouten kunnen gedeeltelijk de prestaties van de leerling verklaren. Leerlingen gebruiken kennis bij het bepalen van de strategie om de taak te maken. Zo gebruiken ze kennis over de begrippen en de relaties tussen de begrippen bij het maken van een taak (Leighton & Gierl, 2007). Bij bijvoorbeeld het beantwoorden van vragen bij een tekst kan het behulpzaam zijn als een leerling kennis heeft van de verschillende doelen die teksten kunnen hebben, zoals informeren of amuseren. Door data te verzamelen over hoe leerlingen taken maken, krijgt de leerkracht inzicht in de kennis die de leerling gebruikt bij het oplossen van een taak. Met deze inzichten kan de leerkracht het onderwijs zo inrichten dat leerlingen de kans krijgen efficiëntere en effectievere oplossingsstrategieën te ontwikkelen.

Wanneer diagnosticeren bestaat uit het vaststellen waardoor een leerling onjuiste antwoorden geeft, is het essentieel om voorafgaand aan de afname van de diagnostische toets te bepalen welke informatie over fouten leerkrachten waardevol vinden voor de inrichting van hun lespraktijk. In literatuur over assessment wordt onderscheid gemaakt tussen fouten en misstappen; leerlingen die een misstap zetten hebben zich vergist in de strategie en/of kennis die ze hebben gebruikt bij het maken van de taak (Bennett, 2011; Kraemer, 2011). Mistappen ontstaan voornamelijk doordat leerlingen minder geconcentreerd werken doordat ze ongemotiveerd zijn of worden afgeleid. Systematische fouten daarentegen zijn meestal het gevolg van onjuiste kennis (misconcepties) of van het ontbreken van kennis en/of vaardigheden die essentieel zijn voor het maken van de betreffende taak. Het diagnosticeren van fouten gaat dus voornamelijk om het vaststellen welke typen systematische fouten resulteren in de onjuiste antwoorden die de leerling heeft gegeven. Deze informatie heeft diagnostische waarde omdat leerkrachten het kunnen gebruiken om de leerlingen feedback te geven over de fouten die ze maken.

4. Adviseren: Op welke doelen moet de interventie zich richten?

Op basis van het integratieve beeld wordt een advies gegeven over de aanpassingen aan de leeromgeving met als doel het leerproces te optimaliseren. Om dit mogelijk te maken moet de diagnose zich richten op het identificeren van sterke en zwakke punten in de prestaties in leerlingen die een zo direct mogelijke relatie hebben met de inrichting van de leeromgeving. Ten eerste wordt, voor elke leerling apart, advies gegeven over de doelen van de interventie en de prioriteit van de doelen.

Het tweede deel van de indicatie bestaat uit advies over de duur, frequentie en de vorm (individueel of in een groep) van de interventie.

Strategieën voor DT

Er zijn twee strategieën om met DT in te spelen op de onderwijsbehoeften van leerlingen (Stobart, 2008):

1. **remediërend**: verhelpen van een gesignaleerd probleem;
2. **preventief**: voorkomen van een probleem.

De eerste strategie kan worden ingezet wanneer tijdens het leerproces problemen worden gesignaleerd of wanneer aan het einde van een leerproces wordt geconcludeerd dat leerdoelen niet zijn behaald. Het is zelf-evident dat de tweede strategie het meest effect is wanneer deze aan het begin van het leerproces wordt gebruikt om de voorkennis van de leerling te bepalen en problemen te voorkomen. Zo kan bijvoorbeeld uit de resultaten van een diagnostische toets blijken dat een leerling bepaalde fouten, die door de meeste leerlingen wel gemaakt worden, niet maakt. In dat geval wordt vastgesteld dat de leerling over eerder verworven kennis (voorkennis) beschikt. De leerkracht zou dan kunnen kiezen voor het verkorten, versnellen en/of verdiepen van het leerproces van de leerling.

Praktische voorwaarden bij DT

Een van de voordelen van gestandaardiseerde en computergestuurde diagnostische toetsen is dat deze betrouwbaardere en meer valide resultaten opleveren. Een ander voordeel is dat computergestuurde toetsen een lagere arbeidsintensiteit hebben dan niet-computergestuurde toetsen. Dergelijke toetsen zijn echter in Nederland nog niet veel ontwikkeld. Maar er zijn ook andere diagnostische assessmentmethoden mogelijk. Eén van de mogelijkheden is het diagnostische interview. In een diagnostisch interview worden verschillende technieken gezamenlijk gebruikt om een diagnose te stellen (Moyer & Milewicz, 2002). Deze technieken zijn:

1. interviewtechnieken (vragen stellen en luisteren);
2. hard op laten uitvoeren van een taak;
3. observatietechnieken.

Uit onderzoek blijkt dat het bij diagnostisch onderzoek belangrijk is om deze technieken op de juiste momenten in te zetten. Zo is het bijvoorbeeld van belang dat de interviewer op het juiste moment doorvraagt, maar moet de leerling ook voldoende tijd krijgen om te kunnen antwoorden (Moyer & Milewicz, 2002). Even (2005) beschreef dat het zeer moeilijk is objectief naar de antwoorden van leerlingen te luisteren, omdat de interviewer geneigd is te interpreteren wat de leerling zegt op basis van eigen ervaringen en opvattingen. Het afnemen van een diagnostisch interview vergt, met andere woorden, behoorlijk wat vaardigheden en kennis. Om die reden is het handig om samen een diagnostisch onderzoek goed voor te bereiden door concreet te maken welke informatie je van de leerlingen wilt weten en door vooraf vragen op te stellen. Een gevaar van te veel structuur aanbrengen is echter dat de interviewer vergeet te luisteren.

Een andere mogelijkheid is gebruik maken van een door een externe partij ontworpen diagnostisch interview. Een voorbeeld van een dergelijk gestructureerd diagnostisch interview is Diagnosticeren en Plannen in de Onderbouw (Cito, 2011). 'Diagnosticeren en Plannen in de Onderbouw' richt zich op het diagnosticeren van strategieën en fouten van leerlingen in het primair onderwijs binnen het vak rekenen-wiskunde. De inhoud (welke vragen) en de vorm (volgorde van de vragen) van het interview zijn volledig vastgelegd. In de uitgave 'Diagnosticeren en Plannen in de Onderbouw' zitten werkbladen met de opgaven voor de leerling en observatieformulieren. Op de observatieformulieren staan de opgaven weergegeven inclusief de veel voorkomende oplosstrategieën die leerlingen bij die opgaven gebruiken. De leerkracht die het interview afneemt codeert welke strategieën de leerling gebruikt en welke fouten hij maakt. Hierdoor krijgt de leerkracht inzicht in de mogelijke verklaringen (zie fase 3 van de diagnostische cyclus) voor de reken-wiskunde prestaties van de leerling.

Gegeven de hoge arbeidsintensiteit van niet geautomatiseerde diagnostische instrumenten is het belangrijk dat geïnvesteerd wordt in het ontwikkelen van geautomatiseerde diagnostische toetsen. Op die manier kan de brug die wordt geslagen tussen de theoretische principes van diagnostisch toetsen en in de implementatie van deze principes in de praktijk worden versterkt.

5.

Samenvatting van de belangrijkste kenmerken van OGW, AfL en DT

In dit hoofdstuk geven we een samenvatting van de belangrijkste kenmerken van opbrengstgericht werken, assessment for learning en diagnostisch toetsen.

Opbrengstgericht werken (OGW)

OGW volgt een vaste cyclus van doelen bepalen, data verzamelen, resultaten registreren, resultaten interpreteren en het onderwijs aanpassen. Deze onderwijsaanpassingen kunnen gericht zijn op leerling-, klas- of schoolniveau. Voor de dataverzameling worden voornamelijk objectieve, betrouwbare en gestandaardiseerde instrumenten gebruikt, met name wanneer de consequenties van de beslissing grote gevolgen hebben voor degene over wie de beslissing wordt genomen. Beslissingen die op schoolniveau worden genomen hebben vaker grotere gevolgen dan bijvoorbeeld beslissingen die leerkrachten dagelijks in de klas maken. Ook zak-/slaagbeslissingen hebben verstrekkende gevolgen voor leerlingen en worden daarom ondersteund met objectieve en betrouwbare datafeedback. OGW is hoofdzakelijk gericht op het verhogen van de uitkomsten van het leerproces en wordt minder gebruikt voor het verbeteren van het leerproces zelf.

Assessment for learning (AfL)

Bij AfL gaat het om het continu verzamelen van bewijs over leervoortgang van leerlingen in de alledaagse klassituaties en het inspelen hierop. In AfL is assessment geïntegreerd in de dagelijkse lespraktijk. Het assessmentproces is een sociale activiteit waarbij de nadruk ligt op het begrijpen en het verbeteren van het leren binnen de klas. Leerlingen nemen actief deel aan hun leerproces en kunnen hun eigen leerproces reguleren. AfL heeft dus betrekking op het leerling- en klasniveau. Er zijn vele manieren waarop dit gedaan kan worden, hier worden dus niet noodzakelijk schriftelijke toetsen bij gebruikt. AfL vindt niet noodzakelijk plaats binnen een bepaalde activiteitscyclus, het is voornamelijk een flexibel en interactief proces. De nadruk ligt vooral op het veelvuldig verzamelen van informatie op allerlei manieren en hier direct op inspelen met aangepaste instructie en feedback. Omdat het in AfL gaat om het creëren van een optimaal leerproces speelt feedback die aansluit op de leerbehoefte van leerlingen een essentiële rol.

Diagnostisch toetsen (DT)

DT is niet alleen gericht op het meten van het leerproces en de leerprestaties van leerlingen, maar ook in de mogelijke verklaringen van deze leeropbrengsten. Deze verklaringen kunnen worden gezocht in de voorkennis die leerlingen hebben of de manier waarop leerlingen redeneren. De verklaringen zijn over het algemeen leerling- en taakkenmerken die beïnvloeden of een leerling bepaalde taken nog niet (zelfstandig) kan uitvoeren. Daarnaast kan een deel van de leerprestaties worden verklaard aan de hand van de oplossingsstrategieën en systematische fouten die leerlingen maken. DT is een cyclisch proces waarbij het leerproces en de leerprestaties van de leerling op een steeds gedetailleerder niveau beschreven worden. Hierbij gaat het om individuele leerlingen; de beslissingen die worden genomen over de inrichting van de leeromgeving gaan over individuele leerlingen. Wanneer gebruik gemaakt wordt van diagnostische toetsen is veelal sprake van objectieve en gestandaardiseerde toetsen die worden afgenomen op formele toetsmomenten. Er zijn in Nederland echter nog niet veel diagnostische toetsen ontwikkeld die geautomatiseerd het leerproces en de leerprestaties van leerlingen in kaart brengen. Leerkrachten kunnen echter ook gebruikmaken van andere assessmentmethoden zoals diagnostische interviews.

6.

Overeenkomsten en verschillen

De overeenkomsten en verschillen tussen de drie benaderingen - opbrengstgericht werken (OGW), assessment for learning (AfL) en diagnostisch toetsen (DT) - worden in dit hoofdstuk behandeld. Hierbij focussen wij ons op het niveau, de doelen, het tijdspad, de frequentie van beslissingen, de assessmentmethoden en het assessmentproces.

Niveau

OGW heeft betrekking op alle niveaus, wat betekent dat het zowel over assessment als over evaluatie gaat. AfL heeft betrekking op het klas- en leerlingniveau. DT heeft betrekking op het leerlingniveau.

Doelen

Het doel van OGW is om te toetsen of te evalueren of bepaalde doelen zijn behaald. Daarmee kan worden besloten of er veranderingen moeten worden doorgevoerd in de leeromgeving. Het doel van AfL is om de kwaliteit van het leerproces te verbeteren, hierbij ligt minder nadruk op de uitkomsten van het leerproces. DT heeft als doel het voorkomen of remediëren van problemen die zich kunnen voordoen tijdens het leerproces.

Tijdspad

Het tijdspad waarin beslissingen worden genomen verschilt ook. In OGW kan het variëren van onmiddellijk tot een aantal jaren na het verzamelen van de data. Feedback komt dan ook vaak niet direct. Dit is met name het geval wanneer feedback van een gestandaardiseerde toets wordt gegeven door een externe partij. In AfL worden beslissingen bijna altijd onmiddellijk tijdens het leerproces gemaakt. Feedback wordt continue gegeven op basis van behoeften van leerlingen. Het tijdspad van DT wordt bepaald door de behoefte aan bijsturing van het leerproces van de leerling, maar het is aan te raden om de tijd tussen de toets en de beslissing beperkt te houden. Feedback is het meest effectief wanneer deze zo snel mogelijk na de toets wordt gegeven.

Frequentie van beslissingen

Daarnaast verschillen de benaderingen in de frequentie waarmee beslissingen worden genomen. In OGW verschilt de frequentie van dagelijks tot eens in een paar jaar. De frequentie van beslissingen hangt samen met de zwaarte van de beslissing; zwaardere beslissingen worden minder vaak genomen. In AfL maken leerkrachten continue beslissingen op basis van de informatie die zij op dat moment beschikbaar hebben. Bij DT hangt de frequentie van de beslissing af van de behoefte aan bijsturing van het leerproces.

Assessmentmethoden

De assessmentmethoden die worden gebruikt in OGW zijn voornamelijk gestandaardiseerde toetsen, die resulteren in kwantitatieve datafeedback. Deze datafeedback moet worden geïnterpreteerd door de gebruiker. Assessmentmethoden in AfL zijn meestal niet gestandaardiseerd, maar juist interactief en flexibel. Deze resulteren voornamelijk in een kwalitatieve beschrijving van wat leerlingen kennen en kunnen. Instrumenten voor DT zijn meestal gestandaardiseerd, deze kunnen leiden tot kwantitatieve of kwalitatieve resultaten. Voor alle benaderingen geldt dat de kwaliteitseisen die worden gesteld aan de instrumenten afhangen van de consequenties van de te nemen beslissingen.

Assessmentproces

In OGW en DT wordt een vaste cyclus gevolgd. Kenmerkend voor deze cyclus is een systematische manier van toetsen op formele meetmomenten. In AfL wordt geen vaste cyclus gebruikt en er is geen sprake van een systematisch proces. Er zijn geen aparte meetmomenten; assessment is geïntegreerd in de instructie. Verder is in AfL de interactie tussen leerlingen belangrijk, bijvoorbeeld self assessment en peer assessment.

7.

Een aanzet tot een programma voor formatieve toetsing

In dit hoofdstuk noemen we vijf elementen uit de drie benaderingen die gezamenlijk een programma voor formatieve toetsing kunnen vormen. Door de hier beschreven elementen van de drie benaderingen te combineren kan de formatieve functie van toetsen en evaluaties optimaal worden benut. Op deze manier kan worden bijgedragen aan de kwaliteit van het leerproces en het verhogen van de leeropbrengsten.

1. Gericht op alle aggregatieniveaus

Door elementen van de drie benaderingen te combineren is de formatieve functie van toetsen en evaluaties gericht op alle niveaus binnen de school (zie figuur in hoofdstuk 1).

2. Gericht op optimalisatie van het leerproces en de leeropbrengsten

Het voordeel van een programma voor formatieve toetsing is dat zowel het leerproces en de leeropbrengsten geoptimaliseerd kunnen worden. In een dergelijk programma is ruimte voor zowel remediëring als preventie van leerproblemen. Effectieve feedback geven tijdens het leerproces is hierbij essentieel.

3. Gericht op lange- en kortetermijnbeslissingen

Door zowel korte als lange feedbackloops te doorlopen, kan zowel op korte termijn als op lange termijn de effectiviteit van gekozen strategieën geëvalueerd worden. Hierdoor kan continue afstemming op de onderwijsbehoeften van leerlingen worden bereikt.

4. Gericht op de benutting van verschillende databronnen

Door het gebruik van verschillende assessmentbenaderingen kan een rijker beeld verkregen worden van het leren van leerlingen. Dit betekent dat kwalitatief hoogwaardige en objectieve data worden gecombineerd met kwalitatieve data die in alledaagse klassituaties worden verzameld. Door aanvullend gebruik te maken van gedetailleerde databronnen over de oorsprong van de leeropbrengsten kan het leerproces gericht gestuurd worden.

5. Gericht op het systematisch en flexibel data verzamelen

Door gebruik te maken van systematische manieren van data verzamelen kan worden gewaarborgd dat alle stappen van een evaluatieve cyclus worden doorlopen. Door daarnaast ook flexibel data te verzamelen kan een rijker beeld van het leerproces en de leeropbrengsten worden gevormd. Hierdoor kan afstemming op de onderwijsbehoeften van leerlingen worden gewaarborgd.

8.

Literatuur

Assessment Reform Group (1999). *Assessment for learning: Beyond the black box*. Cambridge University.

Bennett, R.E. (2011). Formative assessment: a critical review. *Assessment in Education: Principles, Policy & Practice*, 18(1), 5-25. doi:10.1080/0969594X.2010.513678.

Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment Evaluation and Accountability*, 21(1), 5-31. doi:10.1007/s11092-008-9068-5.

Broadfoot, P.M., Daugherty, R., Gardner, J., Harlen, W., James, M., Stobart, G. (2002). *Assessment for learning: 10 principles*. Cambridge, UK: University of Cambridge School of Education.

Cito (2011). *Diagnostieren en plannen in de bovenbouw*. Arnhem: Cito. Zie: [productfolder](#).

Clarke, S. (2001). *Unlocking formative assessment*. Londen: Hodder and Stoughton.

De Bruyn, E.E.J., Ruijsenaars, A.J.J.M., Pameijer, N.K., & Van Aarle, E.J.M. (2003). *De diagnostische cyclus. Een praktijkleer*. Leuven, België: Acco.

Even, R. (2005). Using assessment to inform instructional decisions: How hard can it be? *Mathematics Education Research Journal*, 17(3), 45-61.

Fullan, M. (2007). *The new meaning of educational change*. New York and London: Teacher College Press.

Hargreaves, E. (2005). Assessment for learning? Thinking outside the (black) box. *Cambridge Journal of Education*, 35(2), 213-224. ISSN Print: 0305-764X

Harlen, W., & Winter, J. (2004). The development of assessment for learning: learning from the case of science and mathematics. *Language testing*, 21(3), 390-408. doi:10.1191/0265532204lt289oa.

Inspectie van het Onderwijs (2010). *De staat van het onderwijs. Onderwijsverslag 2008/2009*. Utrecht: Inspectie van het Onderwijs.

Keasley, P., & Tobey, C.R. (2010). *Mathematics formative assessment*. Thousand Oaks, CA: Corwin.

Klenowski, V. (2009). Assessment for learning revisited: An Asia-Pacific perspective. *Assessment in Education: Principles, Policy & Practice*, 16(3), editorial.

Kraemer, J.M. (2011). *Oplossingsmethoden voor aftrekken tot 100*. Eindhoven: Technische Universiteit Eindhoven; Arnhem: Cito.

Ledoux, G., Blok, H., Boogaard, M., & Krüger, M. (2009). *Opbrengstgericht werken: over de waarde van meetgestuurd onderwijs*. SCO-Rapport 812. Amsterdam: SCO-Kohnstamm Instituut.

Leighton, J.P., & Gierl, M.J. (Eds.) (2007). *Cognitive diagnostic assessment for education. Theory and applications*. New York: Cambridge University Press.

Light, D., Wexler, D., & Heinze, J. (2004). *How practitioners interpret and link data to instruction: Research findings on New York City schools' implementation of the grow network*. Paper presented at the annual meeting of the American Educational Research Association, San Diego, CA.

Mandinach, E.B., Honey, M., & Light, D. (April, 2006). *A theoretical framework for data-driven decision making*. Paper presented at the meeting of the American Educational Research Association, San Francisco.

Ministerie van OCW (2009). *Derde voortgangsrapportage Passend Onderwijs*. Den Haag: Ministerie van OCW.

Ministerie van Onderwijs en Wetenschappen (1990). *Weer samen naar school*. Zoetermeer: Ministerie van Onderwijs en Wetenschappen.

Morrison, G.R., Ross, S.M., & Kemp, J.E. (2007). *Designing effective instruction*. New York, NY: John Wiley & Sons.

Moyer, P.S., & Milewicz, E. (2002). *Learning to question: Categories of questioning used by preservice teachers during diagnostic mathematics interviews*. *Journal of Mathematics Teacher Education*, 5, 293-315.

Rupp, A.A., Templin, J., & Henson, R.A. (2010). *Diagnostic measurement. Theory, methods, and applications*. New York: The Guilford Press.

Sanders, P. (2011). *Het doel van toetsen*. In: P. Sanders (Ed.), *Toetsen op school* (pp. 9-20). Arnhem: Cito.

Schildkamp, K., & Kuiper, W. (2010). Data-informed curriculum reform: Which data, what purposes, and promoting and hindering factors. *Teaching and Teacher Education*, 26, 482-496.

Scriven, M. (1967). *The Methodology of Evaluation*. Washington, DC: American Educational Research Association.

Shepard, L.A. (2005). *Formative assessment: Caveat emptor*. Paper presented at the ETS Invitational Conference, The Future of Assessment: Shaping Teaching and Learning, New York.

Slavin, R.E. (2002). Evidence-based education policies: Transforming educational practice and research. *Educational Researcher*, 21(7), 15-21.

Stobart, G. (2008). *Testing times: The uses and abuses of assessment*. London: Routledge.

Van der Kleij, F.M., Vermeulen, J.A., Schildkamp, K., Eggen, T.J.H.M. (submitted). *Data-based decision making, assessment for learning, and diagnostic testing in formative assessment*. Manuscript submitted for publication.

Verhofstadt-Denève, L., Van Geert, P., & Vyt, A. (2003). *Handboek ontwikkelingspsychologie. Grondslagen en theorieën*. Houten: Bohn Stafleu Van Loghum.

Visscher, A. & Erhen, M. (2011). *De eenvoud en complexiteit van opbrengstgericht werken. Analyse in opdracht van het Ministerie van Onderwijs, Cultuur en Wetenschap*. Enschede: University of Twente.

Colofon

Copyright © Stichting Cito Instituut voor Toetsontwikkeling

Het auteursrecht op de ToetsSpecial 'Leren van toetsen; een cyclisch proces' berust bij Cito. Overname van de inhoud van de Special is uitsluitend toegestaan na toestemming van de redactie van ToetsWijzer. Verzoeken tot overname van de inhoud dienen te worden gericht aan de redactie van ToetsWijzer, onder vermelding van het medium, de oplage en de doelgroep.

Contact

Meer weten? Reageren? Neem contact op met de redactie van ToetsWijzer.

Websites

Cito | www.cito.nl

RCEC | www.rcec.nl

Universiteit Twente | www.utwente.nl

ToetsWijzer | www.toetswijzer.nl



Over deze ToetsSpecial

In de afgelopen jaren zijn er veel begrippen ontstaan die samenhangen en deels samenvallen met het begrip *formatieve toetsing*, zoals opbrengstgericht werken, ‘assessment for learning’ en diagnostisch toetsen. Het doel van deze special is om de lezer wegwijs te maken in de betekenis van deze begrippen, de overeenkomsten en verschillen tussen deze begrippen te verduidelijken en een beeld te schetsen van de betekenis van deze begrippen voor de onderwijspraktijk in de klas.