

## Taaltoetsing in vmbo en mbo

URIËL SCHUURS & LUDO VERHOEVEN

In dit artikel komen bevindingen uit een literatuurreview en een praktijkonderzoek naar taaltoetsing in vmbo en mbo aan de orde. De auteurs schetsen een beeld van verschillende typen toetsen, en van de eisen die men volgens de vakliteratuur aan toetsen met verschillende functies zou kunnen stellen. Daarna worden de uiteenlopende eisen besproken die blijken een kleinschalig praktijkonderzoek door taaldocenten en door toetsdeskundigen worden gesteld aan formatieve taaltoetsen (om het onderwijs bij te kunnen sturen) en aan summatieve taaltoetsen (ter afsluiting van een onderwijsperiode). Aan de literatuurreview en het praktijkonderzoek verbinden de auteurs enkele conclusies en aanbevelingen.

Leonie en Anne zetten hun schoolbankjes alvast tegen elkaar, terwijl de leraar een tekstbegriptoets uitdeelt. Ze hebben inmiddels ook allebei het boek voor Nederlands bij het betreffende hoofdstuk opengeslagen. In de war gebracht door de twee leerlingen staat de leraar te twijfelen: moet hij - al dan niet gespeeld - boos worden? Het verbod op spieken en samenwerken herhalen? Zijn twijfel duurt net iets te lang en de hele klas zit gniffelend te wachten op wat komen gaat... Wat zou u eigenlijk doen in die situatie?

Anne en Leonie vonden dat ze best hun boek mochten gebruiken bij de toets: je moet er toch wat van kunnen leren? Dat de toets vervolgens geen betrouwbaar beeld meer geeft van wat de leerlingen er op eigen kracht van kunnen, dat is voor hen vers twee. Voor de meeste docenten zijn doel en functie van een toets gelukkig helderder dan voor de hierboven opgevoerde leerlingen. Daarmee is niet gezegd dat de inzet van toetsen in het taalonderwijs altijd op een geslaagde manier plaatsvindt. Uit onderzoek dat we in schooljaar 2009-2010 hebben verricht<sup>1</sup>, blijkt dat docenten in vmbo en mbo vaak wel notie hebben van de eisen die je aan een toets zou kunnen stellen, maar dat daar in de praktijk niet steeds de juiste consequenties aan worden verbonden. In dit artikel schetsen we eerst verschillende typen toetsen en de uiteenlopende eisen die aan toetsen kunnen worden gesteld; daarna geven we een klein inkijkje in opvattingen binnen het onderwijsveld en tot slot doen we enkele aanbevelingen.<sup>2</sup>

### Ander onderwijs, andere toetsen

Met name in het mbo is het onderwijs de laatste decennia sterk veranderd: meer compe-

tentiegericht onderwijs en meer zelfstandigheid voor deelnemers. In het kielzog daarvan heeft men vooral in het mbo gezocht naar andersoortige toetsvormen; in het vmbo was deze verschuiving afwezig of in elk geval minder duidelijk zichtbaar. Voor het beoordelen van de taalvaardigheid had men behoefte aan een vorm van geïntegreerde toetsing, waarbij taalcompetenties niet meer los van de beroepscompetenties getoetst worden, maar geïntegreerd in het professionele handelen (vgl. Van Kleef, 2010)<sup>3</sup>. Dit heeft geleid tot een breed palet van mogelijk inzetbare toetsvormen, variërend van een traditionele kennisstoets tot een *self-assessment*. Tegelijkertijd is er een tendens ontstaan om toetsing niet alleen achteraf te laten plaatsvinden (*assessment OF learning* ofwel summatieve toetsing), maar ook tijdens, of zelfs geïntegreerd binnen het onderwijs (*assessment FOR learning* ofwel formatieve toetsing). Deze verandering vond vooral plaats in het mbo, maar is in lichtere mate ook in het vmbo aanwijsbaar.

Een duidelijk voorbeeld van formatieve toetsing is de diagnostische toets. Deze is bedoeld om na te gaan welke onderdelen door een leerling nog onvoldoende worden beheerst zodat daar het onderwijs op kan worden gericht. Met name de invloedrijke metastudie van Black & Wiliam, (1998) laat zien dat nauwkeurig gedoseerde formatieve toetsing in veel gevallen leidt tot een sterke vooruitgang in leerrendement. Dit geldt vooral wanneer leerlingen bewust gemaakt worden van de gestelde onderwijsdoelen, hun duidelijk gemaakt wordt hoe ze in dat licht hun werk kunnen optimaliseren en hun de mogelijkheid wordt geboden om deelgenoot te zijn in het evaluatieproces (bijvoorbeeld door *self- and peer assessment*, vgl. Pellegrino, 2008).

Het is een interessante vraag welke toetsvormen geschikt zijn voor de twee onderscheiden functies van *assessment*, of in meer algemene zin: welke eisen te stellen zijn aan

toetsen, afhankelijk van de beoogde functie. Die vraag hebben we op twee manieren geprobeerd te beantwoorden:

- aan de hand van onderzoeksgegevens uit een literatuurreview;
- aan de hand van bevraging van leerkrachten en experts.

Hier vatten we enkele opmerkelijke resultaten van de literatuurreview kort samen.<sup>4</sup> In deze literatuurreview hebben we specifiek gekeken naar gegevens over relatief nieuwe toetsvormen zoals *self-assessment*, *peer-assessment* en portfolio-beoordeling.

Er is in het onderwijs een kennelijke behoefte om gebruik te maken van *self-assessment* (Booth, 2003). Scores op deze assessmentvorm blijken echter ronduit slecht te correleren met andere, aantoonbaar betrouwbare indicaties van taalvaardigheid (vgl. o.a. Alderson & Huhta, 2005; Matsuno, 2009). De betrouwbaarheid van *peer assessment* is iets hoger dan van *self-assessment* (vgl. Matsuno, 2009; Topping, 2003) en bovendien lijkt *peer assessment* een beter aantoonbaar positief effect op het onderwijs te hebben. Als toetsvorm – en zeker als summatieve toetsvorm – zijn beide instrumenten echter ongeschikt.

Er is in het onderwijs ook een grote behoefte om gebruik te maken van portfolio's. Aan het portfolio als tastbaar 'bewijs' van het leren worden drie hoofdfuncties toegekend (vgl. Wolf et al., 1997):

- het ontwikkelingsgerichte portfolio kan het leerproces zichtbaar maken, zowel voor de omgeving als voor het individu zelf;
- het assessmentportfolio (ook wel 'beoordelingsportfolio') bevat materiaal op basis waarvan uitspraken worden gedaan over bekwaamheden van de leerling;
- het presentatieportfolio (ook wel *showcase*) heeft vooral een toonfunctie, als hulp bij bijvoorbeeld sollicitaties. Dit type portfolio

moet uitspraken aannemelijk maken die de samensteller doet over zijn of haar eigen competenties / capaciteiten.

Vooraf bij universiteiten, hbo- en mbo-instellingen heeft het portfolio een hoge vlucht genomen (vgl. Inspectie van het Onderwijs, 2003). Het gebruik van portfolio's in het onderwijs heeft aantoonbaar nut (Davies & LeMahieu, 2003), maar de beoordeling van portfolio's is problematisch: leerkrachten weten niet goed wat ze beoordelen (vgl. Smith & Tillemans, 2007). Al kan training van de beoordelaars enige verbetering brengen (Sluijsmans et al., 2008), toch is vooral de betrouwbaarheid van de beoordeling van portfolio's problematisch. Deze betrouwbaarheid zou vergroot kunnen worden door eenduidige criteria tussen beoordelaars af te spreken en door ervoor te zorgen dat de performance indicators representatief zijn voor de competenties die beheerst moeten worden. In Nederland heeft vooral Straetmans, (2006) er werk van gemaakt om de portfolio-scoring te standaardiseren met behulp van protocollaire beoordelingsprocedures (vgl. ook Sluijsmans et al., 2008). In hoeverre de ontwikkelde methoden daadwerkelijk in praktijk worden gebracht, en met welk effect, hebben we niet kunnen achterhalen.

Een en ander betekent dat veelbesproken en voor het mbo gepropageerde toetsvormen zoals *peer assessment*, *self-assessment* en *portfolio* niet geschikt zijn om belangrijke beslissingen op te baseren. Wel zijn ze bruikbaar als onderwijsmiddel om de leerling duidelijk te maken aan welke standaarden hij aan het eind van het onderwijsleerproces moet voldoen.

### Aan welke eisen moet een toets voldoen?

Vanouds worden er twee eisen genoemd waaraan een toets zou moeten voldoen: de

betrouwbaarheid en de validiteit.

De betrouwbaarheid zegt iets over de mate waarin gemeten waarden overeenkomen met de feitelijke waarden. Om betrouwbaar te mogen heten, moet een assessment nauwkeurig meten en bovendien bij herhaaldelijk meten tot hetzelfde resultaat leiden. Naarmate een toets betrouwbaarder is, zijn kleinere verschillen in scores nog op een zinvolle manier te interpreteren. De betrouwbaarheid is de belangrijkste eis die aan een assessment te stellen is. Als een assessment niet voldoende betrouwbaar is, heeft het weinig zin om de validiteit aan de orde te stellen.

De validiteit van een assessment is de mate waarin het instrument meet wat gemeten moet worden. Waar betrouwbaarheid een objectief gegeven is, moet de validiteit van een toets worden beredeneerd: aangetoond moet worden dat het beoogde construct wordt gemeten. In een retrospectief overzicht merkt Spolsky (2008, 448) op dat de beschikbaarheid van statistische technieken heeft bevorderd dat individuele metingen werden vervangen door lange reeksen van eenvoudig scoorbare opgaven, zodat de betrouwbaarheid in beeld kon worden gebracht, maar dat overeenstemming over de validiteit van een meetinstrument aanzienlijk lastiger te bereiken is. Het lijkt erop dat veel vormen van alternatieve assessment (bijvoorbeeld *self-assessment*, *peer-assessment* en *portfolio*) ontwikkeld zijn vanuit onvrede met de validiteit van de toetsen die op dat moment in gebruik waren. Het beeld dringt zich op als zouden gestandaardiseerde toetsen met een acceptabele betrouwbaarheid niet goed passen bij het gegeven onderwijs, en om die reden zijn vervuld voor toetsen die naar het oordeel van leerkrachten wel valide zijn, maar waarvan de betrouwbaarheid ontoereikend of onbekend is.

Sinds de opkomst van de zogenaamde *alternative assessment* worden er naast betrouw-

baarheid en validiteit een groot aantal andere eisen genoemd die men zou kunnen kan stellen aan toetsen. Zo benoemen Segers et al. (2003) zes dimensies waarlangs de veranderingen in assessment in de laatste twee decennia zich bewegen: onder andere de authenticiteit van de taakstelling en de soort aspecten die beoordeeld worden (naast cognitieve aspecten kunnen ook metacognitieve, sociale en psychomotorische aspecten beoordeeld worden). Bachman & Palmer (1996) introduceerden het begrip *usefulness* ofwel de bruikbaarheid van een test. Deze bruikbaarheid hangt in hun optiek af van de volgende testkarakteristieken:

- de betrouwbaarheid;
- de constructvaliditeit;
- de authenticiteit: de taken moeten lijken op wat de leerling later in de praktijk moet kunnen;
- interactiviteit: de leerling moet de taakstelling als realistisch kunnen ervaren, omdat resultaten in een onechte taakomgeving nietszeggend zouden zijn;
- de impact van de resultaten op het vervolgonderwijs waarbij positieve impact als een belangrijk kenmerk van *authentic assessment* wordt gezien (vgl. Wiggins, 1989; Sluijsmans et al., 2008);
- praktische aspecten: zo moet een *assessment*-afname inpasbaar zijn in de dagelijkse gang van zaken en de kosten en de baten moeten in verhouding zijn.

Ook door anderen (vgl. Booth et al., 2003; Straetmans, 2006) zijn er kwaliteitscriteria benoemd waaraan een assessment zou moeten voldoen. Aan de hand van deze en soortgelijke publicaties hebben we een lijst met 20 mogelijk relevante kenmerken van assessment-instrumenten opgesteld, op basis van frequentie waarin ze genoemd zijn in de literatuur:

1. taken in de toets komen inhoudelijk overeen met taken in de toekomstige beroepspraktijk;

2. taken in de toets zijn competentiegericht;
3. taken in de toets hebben dezelfde moeilijkheidsgraad als die in de praktijk;
4. de toets geeft een goed beeld van het prestatieniveau in de praktijk;
5. bij de toets horen van tevoren vastgestelde kwaliteitseisen;
6. de leerlingen weten wat er van ze verwacht wordt;
7. de validiteit van de toets is in orde;
8. er is sprake van beoordelaarsafhankelijkheid;
9. de betrouwbaarheid van de toets is goed toetsdoelen en onderwijsdoelen komen overeen;
11. de toets wordt gemaakt als de leerling eraan toe is;
12. de toets is leerzaam voor de leerling;
13. de toets is leerwegaafhankelijk;
14. de kosten van de toets staan in verhouding tot de opbrengst;
15. de toets draagt bij aan het leerproces door de feedback die de leerling krijgt;
16. het resultaat van de toets heeft invloed op de inhoud van het onderwijsleerproces;
17. het maken van de toets bevordert het vermogen tot zelfbeoordeling;
18. de toetsscores voorspellen de prestaties van de leerling in het onderwijs;
19. de toetsscores zijn generaliseerbaar over taken;
20. de toetsscore geeft een goede voorspelling van de prestaties in de beroepspraktijk.

Aan deze criteria wordt uiteenlopend belang gehecht in de vakliteratuur. De uitgevoerde literatuurreview laat zien dat er wel – naast de eisen van betrouwbaarheid en validiteit – vrijwel volledige eenstemmigheid heerst over vier kenmerken waaraan een assessment zou moeten voldoen:

1. de taken in de taaltoets komen overeen met taken in de praktijk;
2. de taken in de taaltoets en taken in de

- praktijk zijn vergelijkbaar moeilijk;
- 3. de toetsscore geeft een goed beeld van het prestatieniveau in de praktijk;
- 4. de toetsscores hebben predictieve validiteit voor de beroepspraktijk.

Opvallend is dat bij het bespreken van wenselijke of noodzakelijke kenmerken van assessments in de literatuur doorgaans niet wordt gespecificeerd welke functie aan de betref-

fende assessment wordt toegekend. Toch is wel duidelijk dat bij summatieve toetsing - dus bij afsluitende en certificerende toetsen - de betrouwbaarheid van het toetsinstrument van groter belang moet worden geacht dan bij *assessment for learning*. Hetzelfde geldt voor bijvoorbeeld eisen die betrekking hebben op de generaliseerbaarheid van de resultaten naar toekomstige beroepstaken: bij een toets die bedoeld is om leerlingen meer inzicht te

	VMBO	MBO
	REGELMATIG TOT VAAK	REGELMATIG TOT VAAK
Kennistoets	50,0%	42,9%
Essaytoets	25,0%	0,0%
Vaardigheidstoets	62,5%	71,4%
Casustoets	0,0%	57,1%
Voortgangstoets	37,5%	28,6%
Peer assessment	12,5%	14,3%
Self-assessment	0,0%	0,0%
Stage-opdracht	25,0%	57,1%
Projectopdracht	12,5%	42,9%
Gedragassessment	12,5%	42,9%
Afstudeeropdracht	37,5%	28,6%
Portfolio assessment	0,0%	42,9%
Criteriumgericht interview	0,0%	28,6%
Reflectie-opdracht	0,0%	42,9%

Tabel 1: Mate van gebruik van verschillende assessmentvormen

geven in wat ze nog niet zo goed beheersen, is het minder belangrijk of de taken wel precies even moeilijk zijn als taken die ze later in hun beroep moeten uitoefenen.

### Een inkijkje in de opvattingen binnen het onderwijsveld

In aanvulling op het eigenlijke onderzoek - de literatuurreview - leek het ons zinvol om in een kleinschalig aanvullend onderzoekje te polsen of de in de literatuur gevonden inzichten overeen komen met opvattingen binnen het onderwijsveld. Met het oog daarop hebben we een klein aantal leerkrachten in vmbo en mbo schriftelijk bevestigd. Deze kleinschalige bevestigingsronde ging als volgt: via een open call in oktober 2009 op de mailinglist <list-nederlands@digischool.nl> is aan leerkrachten gevraagd om een vragenlijst in te vullen die was geplaatst op een besloten website. Het uiteindelijke aantal respondenten is klein: veel leerkrachten die aanvankelijk hadden toegezegd, zijn blijkbaar terugheschrokken door de forse omvang van de vragenlijst. De enquête is tussen 20 november en 20 december 2009 beantwoord door 15 leerkrachten (7 leerkrachten uit het mbo en 8 leerkrachten Nederlands uit het vmbo). Daarnaast is een achtal selectief gekozen experts 5) gevraagd om hun visie. We presenteren de gegevens in het besef dat voor verregaande conclusies een herhaling met meer respondenten noodzakelijk is.

Allereerst is aan de leerkrachten gevraagd om van een aantal soorten toetsen op een vierpuntsschaal aan te geven hoe vaak ze er gebruik van maken. De in de vragenlijst genoemde toetsen zijn afkomstig uit de literatuurstudie en werden in de vragenlijst voor de respondenten van een korte omschrijving voorzien. Antwoordmogelijkheden waren: (bijna)

nooit / af en toe / regelmatig / (erg) vaak. Om de contrasten tussen vmbo en mbo goed zichtbaar te maken is in tabel 1 weergegeven hoeveel procent van de onderzochten regelmatig tot vaak gebruik maakt van de genoemde toetsvormen:

Deze kleinschalige veldbevestiging onderstreept het in paragraaf 1 al gesignaleerde verschil tussen vmbo en mbo. In het vmbo zijn de kennistoets en de vaardigheidstoets verreweg het meest gebruikt. Daarnaast wordt er met enige regelmaat gebruik gemaakt van voortgangstoetsen, essaytoetsen, stage-opdrachten en afstudeeropdrachten.

In het mbo zien we een ander beeld: daar zijn de vaardigheidstoets, de casustoets en de stageopdracht de meest gebruikte vormen van assessment. Verder wordt in het mbo met enige regelmaat gebruik gemaakt van twee vormen van assessment die in het vmbo helemaal niet genoemd zijn, namelijk de portfolio-assessment en de reflectie-opdracht. Opvallend is dat in beide schooltypes weinig gebruik gemaakt wordt van de in de literatuur veelbesproken assessmentvormen, zoals het criteriumgericht interview, de peer assessment en de self-assessment.

Kern van de schriftelijke vragenlijst voor leerkrachten betreft de eerder genoemde lijst met 20 eisen die aan assessmentinstrumenten gesteld kunnen worden. Bij elk van deze eisen is gevraagd hoe belangrijk de leerkracht deze eis vindt; daarbij is een onderscheid gemaakt naar instrumenten met een formatieve functie en met een summatieve functie. Er is bijvoorbeeld gevraagd hoe belangrijk het is dat bij een formatieve taaltoets de toetsscores een goede voorspeller zijn van prestaties in de beroepspraktijk. Als antwoordmogelijkheden kon worden gekozen uit Erg onbelangrijk / Onbelangrijk / Belangrijk / Erg belangrijk / Weet niet of Geen mening.

	(verkorte) omschrijving eis	FORMATIEVE TOETSEN			SUMMATIEVE TOETSEN		
		leerkracht vmbo	leerkracht mbo	expert	leerkracht vmbo	leerkracht mbo	expert
1	taken komen overeen met taken praktijk	87,5%	85,7%	58,0%	87,5%	100,0%	71,0%
2	taken zijn competentiegericht	62,5%	57,1%	43,0%	100,0%	71,4%	71,0%
3	moeilijkheidsgraad als in praktijk	75,0%	85,7%	58,0%	100,0%	100,0%	71,0%
4	geeft goed beeld prestatieniveau praktijk	75,0%	100,0%	58,0%	87,5%	100,0%	57,0%
5	tevoren vastgestelde kwaliteitseisen	100,0%	100,0%	71,0%	100,0%	100,0%	86,0%
6	leerlingen weten wat van ze verwacht wordt	87,5%	85,7%	72,0%	100,0%	100,0%	57,0%
7	validiteit	100,0%	85,7%	100,0%	100,0%	100,0%	100,0%
8	beoordelaarsafhankelijkheid	62,5%	57,1%	43,0%	75,0%	85,7%	86,0%
9	betrouwbaarheid	100,0%	85,7%	72,0%	100,0%	100,0%	86,0%
10	toets- en onderwijsdoelen komen overeen	100,0%	100,0%	87,0%	100,0%	100,0%	100,0%
11	toets maken als leerling eraan toe is	87,5%	28,6%	86,0%	87,5%	42,9%	43,0%
12	toets is leerzaam voor de leerling	87,5%	85,7%	87,0%	75,0%	57,1%	14,0%
13	toets is leerwegaafhankelijk	62,5%	71,4%	57,0%	75,0%	71,4%	73,0%
14	kosten staan in verhouding tot opbrengst	75,0%	100,0%	87,0%	75,0%	100,0%	73,0%
15	draagt bij aan leerproces door feedback	87,5%	85,7%	100,0%	87,5%	100,0%	29,0%
16	invloed op inhoud onderwijsleerproces	100,0%	85,7%	100,0%	100,0%	71,4%	29,0%
17	bevordert vermogen tot zelfbeoordeling	87,5%	71,4%	100,0%	75,0%	71,4%	43,0%
18	voorspellen prestaties leerling in onderwijs	50,0%	71,4%	43,0%	100,0%	100,0%	100,0%
19	scores zijn generaliseerbaar over taken	50,0%	42,9%	57,0%	62,5%	85,7%	100,0%
20	voorspellen prestaties in beroepspraktijk	37,5%	85,7%	43,0%	87,5%	85,7%	73,0%
	<b>GEMIDDELD</b>	<b>78,8%</b>	<b>78,6%</b>	<b>70,4%</b>	<b>88,8%</b>	<b>87,1%</b>	<b>68,1%</b>

Tabel 2. Het belang van verschillende assessment-kenmerken volgens leerkrachten vmbo, leerkrachten mbo en experts

Ook hier zijn bij de gegevensverwerking de scores op de eerste twee en de laatste twee antwoordcategorieën gesommeerd. In tabel 2 is per onderscheiden soorten toets en per groep respondenten aangegeven hoeveel procent heeft gekozen voor de antwoorden Belangrijk of Erg belangrijk.

De resultaten maken duidelijk dat zowel leerkrachten als experts vinden dat toetsdoelen en onderwijsdoelen overeen moeten komen (10) en daarnaast validiteit (7) zeer belangrijk vinden; dit geldt voor zowel de formatieve als de summatieve toetsen. Waar het de betrouwbaarheid (9) betreft, ligt dit anders: leerkrachten vinden de betrouwbaarheid vrijwel unaniem van groot belang bij zowel formatieve als summatieve toetsen. De experts vinden betrouwbaarheid in mindere mate van belang bij summatieve toetsen - wellicht in de wetenschap dat geen enkel instrument volledig betrouwbaar kan zijn - en in nog mindere mate bij de formatieve toetsen. Ter toelichting daarbij werd opgemerkt dat het voornaamste doel van een formatieve toets niet is om een betrouwbare meting te doen, maar om indicaties te verkrijgen rondom wat beheerst wordt en wat niet. In dezelfde lijn ligt de bevinding van experts dat een formatieve toets niet per se beoordelaarsafhankelijk (8) scorebaar hoeft te zijn: de kosten en moeite wegen in dat geval niet op tegen de opbrengsten.

Bij de formatieve toetsen (*assessment for learning*) zijn de experts minder veeleisend dan de leerkrachten: dit is vooral het geval waar het gaat om de predictieve validiteit, zoals die tot uiting komt in de mate waarin toetstaken de taken in de beroepspraktijk weerspiegelen (4), de mate waarin de prestaties op de toets die in de praktijk voorspellen (20) en de moeilijkheidsgraad van de assessment die overeen zou moeten komen met de praktijk (3). Het zijn vooral de mbo-leerkrachten die benadrukken dat de assessment-taken even moeilijk moeten zijn als die in de

praktijk en dat de assessment de prestaties in de beroepspraktijk zouden moeten kunnen voorspellen.

Bij de summatieve toetsen (*assessment of learning*) benadrukken de experts - naast de al genoemde validiteit (7) en de parallellie tussen toets- en onderwijsdoelen (10) - de voorspellende waarde van assessments voor het vervolgonderwijs (18) en de generieke werking van assessments: wat getoetst wordt, moet generaliserende uitspraken opleveren voor andersoortige taken (19). Opvallend is verder het verschil van mening tussen de experts en de leerkrachten over de functie van de assessments voor het onderwijs zelf: de leerkrachten vinden in meerderheid dat summatieve toetsing leerzaam moet zijn voor de leerlingen (12), door feedback moet bijdragen aan het leerproces (15), het onderwijsleerproces inhoudelijk moet beïnvloeden (16) en het vermogen tot zelfbeoordeling moet bevorderen (17). De experts benadrukken dat dit alles juist niet de functie van summatieve toetsing is.

Meer in algemene zin lijken leerkrachten alle mogelijke kenmerken van toetsen van groot belang te vinden; experts lijken eerst te kijken naar het doel van de toetsing om op grond daarvan te definiëren aan welke kenmerken een toets zou moeten voldoen. Het meest extreem komt deze opvatting tot uiting bij de vraag of toetsen moeten bijdragen aan het leerproces door middel van feedback (15): het merendeel van de leerkrachten vindt dit nodig bij zowel formatieve als summatieve toetsen. De experts achten dit unaniem nodig bij de formatieve toetsen, maar slechts een kleine minderheid van de experts vindt dit noodzakelijk bij de summatieve toetsen.

### Een terugblik

Wat heeft de literatuurreview en de daaraan gekoppelde veldbevraging nu opgeleverd?

In de eerste plaats heeft de veldbevestiging duidelijk gemaakt dat leerkrachten niet steeds goed zicht hebben op de verschillende functies van assessment-instrumenten. Het lijkt daarom de moeite waard om in een professionaliseringstraject expliciet aandacht te besteden aan de vraag welke eisen aan toetsen kunnen worden gesteld, afhankelijk van de functie waarmee de instrumenten worden ingezet.

Ook hebben we beter zicht gekregen op een belangrijk verschil tussen vmbo en mbo. Tabel 1 laat zien dat met name in het mbo veel gebruik wordt gemaakt van formatieve toetsen waarvan de kwaliteit niet duidelijk is, zoals het gedragsassessment en het portfolio. De kans is reëel dat deze toetsen een lage betrouwbaarheid hebben en subjectief beoordeeld worden; bovendien is een acceptabele mate van validiteit niet zonder meer gegarandeerd. Zolang dit het geval is, verdient het aanbeveling om waar mogelijk naast deze - vaak zelfgemaakte - formatieve toetsen ook gestandaardiseerde formatieve toetsen in te zetten.

In het vmbo wordt relatief gezien weinig gebruik gemaakt van instrumenten voor *assessment-for-learning*. Hoewel we willen blijven benadrukken dat *peer learning*, *peer evaluation* en het portfolio onvoldoende betrouwbaar zijn om als summatieve toets te kunnen worden ingezet, zouden deze instrumenten wel vaker in het vmbo kunnen worden ingezet. Mits op de juiste manier als onderwijsmiddel gebruikt, kan dit leiden tot vergroting van het onderwijsrendement. Bovendien kan het helpen de grote cultuurverschillen tussen vmbo en mbo te verkleinen.

Het mbo lijkt ermee gediend dat er weer gebruik gemaakt wordt van summatieve toetsen waarvan de kwaliteit gecontroleerd is. Om gestandaardiseerd te meten is het noodzakelijk om te beschikken over gestandaardiseerde meetinstrumenten die op een meer dan lokaal niveau meten en voor sum-

matieve doeleinden kunnen worden ingezet. Inmiddels is een traject gestart dat ervoor zorgt dat in 2014 een vorm van centrale examinering voor Nederlands en rekenen/wiskunde wordt ingevoerd, gebaseerd op de referentieniveaus van Meijerink (Expertgroep Doorlopende Leerlijnen Taal en Rekenen, 2009). Gelet op de diversiteit aan opleidingen in het mbo zal het lastig zijn om op korte termijn inhoudelijk voldoende draagvlak te vinden voor generieke toetsen voor alle mbo-richtingen. Het verdient daarom aanbeveling om leerkrachten ervan te overtuigen dat gestandaardiseerde, generieke taaltoetsen een aanwinst kunnen zijn voor alle beroepsopleidingen.

#### NOTEN

1. Dit artikel is gebaseerd op onderzoek dat werd gefinancierd door NWO op basis van de aanvraag 'Meten van leerprestaties in het onderwijs Nederlands op het (v)mbo' (PROO Review Studies 2008).
2. Eerstgenoemde auteur is werkzaam bij Cito. Om die reden vermelden we dat opvattingen in dit artikel niet noodzakelijk overeenkomen met algemene instituu- opvattingen van Cito.
3. Deze 'geïntegreerde' vorm van toetsing maakt het lastig om nog te spreken van een taaltoets: bij geïntegreerde toetsing wordt aan de beoordeling van een toets op het vlak van de beroepsmatige kennis, bijvoorbeeld een portfolio, ook een oordeel over de taalvaardigheid van de student gekoppeld. In zo'n geval is er strikt genomen geen sprake van een taaltoets, terwijl wel taalvaardigheid wordt getoetst.
4. Een uitgebreid overzicht van de resultaten is te vinden in Schuurs & Verhoeven, 2010.
5. De experts waren allen verbonden aan universiteiten en/of onderwijsverzorgende instellingen en zijn werkzaam op het gebied van toetsing.

#### LITERATUUR

- Alderson, J. C., & Huhta, A. (2005). The Development of a Suite of Computer-Based Diagnostic Tests Based on the Common European Framework. *Language Testing*, 22(3), 20.
- Bachman, L. F., & Palmer, A. S. (1996). *Language assessment in practice: Developing language tests and justifying their use in the real world*. Oxford: Oxford University Press.
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education*, 5, 1-74.
- Booth, R., Clayton, B., Hartcher, R., Hungar, S., Hyde, P., & Wilson, P. (2003). *The Development of Quality Online Assessment in Vocational Education and Training. Volume 1 and Volume 2*. Leabrook, Australia: National Centre for Vocational Education Research.
- Davies, A., & LeMahieu, P. (2003). Assessment for learning: Reconsidering portfolios and research evidence. In M. Segers, F. Dochy & E. Cascallar (Eds.), *Optimising new modes of assessments: In search of qualities and standards* (pp. 141-169). Dordrecht: Kluwer Academic Publishers.
- Expertgroep Doorlopende Leerlijnen Taal en Rekenen (2009). *Over de drempels met taal en rekenen*. Enschede.
- Inspectie van het Onderwijs (2003). *Zicht op toetsen / Toetsing en examinering in het hoger onderwijs*. Den Haag.
- Kleef, A. van. (2010) Beroepsvaardigheden? Het gaat altijd over taal. *Levende Talen Magazine*, 97(8), 24-27.
- Matsuno, S. (2009). Self-, peer-, and teacher-assessments in Japanese university EFL writing classrooms. *Language Testing*, 26(1), 75-100.
- Pellegrino, J. (2008). Assessment for learning: Using assessment formatively in classroom instruction. *International Journal of Psychology*, 43(3-4), 381-381.
- Schuurs, U., Verhoeven, L. (2010). *Meten van leerprestaties in het (v)mbo: assessment for learning en assessment of learning*. Nijmegen: Expertisecentrum Nederlands.
- Segers, M., Dochy, F., & Cascallar, E. (2003). The era of assessment engineering: Changing perspectives on teaching and learning and the role of new modes of assessment. In M. Segers, F. Dochy & E. Cascallar (Eds.), *Optimising new modes of assessments: In search of qualities and standards* (pp. 1-12). Dordrecht: Kluwer Academic Publishers.
- Sluismans, D. M. A., Straetmans, G. J. J. M., & van Merriënboer, J. J. G. (2008). Integrating Authentic Assessment with Competence-Based Learning in Vocational Education: The Protocol Portfolio Scoring. *Journal of Vocational Education and Training*, 60(2), 159-172.
- Smith, K., & Tillema, H. (2008). The challenge of assessing portfolios: In search of criteria. In A. Havnes & L. McDowell (Eds.), *Balancing dilemmas in assessment and learning in contemporary education* (pp. 183-195). New York: Routledge.
- Spolsky, B. (2008). Language assessment in historical and future perspective. In E. Shohamy & N. H. Hornberger (Eds.), *Encyclopedia of language and education*, 2nd edition, Volume 7: Language Testing and Assessment (pp. 445-454). New York: Springer Science+Business Media LLC.
- Straetmans, G. J. J. M. (2006). *Bekwaam beoordelen en beslissen. Beoordelen in competentiegerichte beroepsopleidingen* (lectorale rede). Enschede: Saxion Hogescholen.
- Topping, K. (2003). Self and peer assessment in school and university: Reliability, validity and utility. In M. Segers, F. Dochy & E. Cascallar (Eds.), *Optimising new modes of assessments: In search of qualities and standards* (pp. 55-87). Dordrecht: Kluwer Academic Publishers.

Wiggins, G. (1989). A true test: toward more authentic and equitable assessment. *Phi Delta Kappan*, 70(9), 703–713.

Wolf, K., Lichtenstein, G., & Stevenson, C. (1997). Portfolios in teacher education. In J. H. E. Strange (Ed.), *Evaluating teaching: a guide to current thinking and best practice* (pp. 193–214). Thousand Oaks, California: Corwin Press Inc.

URIËL SCHUURS publiceert met regelmaat over taalonderwijs en -toetsing. Centrale thema's in zijn werk zijn de optimalisering van taalonderwijs, zinnig gebruik van ict-toepassingen in onderwijs en toetsing en effectmeting. Daarnaast verzorgt hij schrijfvaardigheidstrainingen. Ten tijde van het onderzoek dat in dit artikel beschreven wordt, werkte Uriël Schuurs als senior onderzoeker bij Expertisecentrum Nederlands. Vanaf maart 2011 is hij werkzaam als examenmaker bij het Cito in Arnhem. E-mail: <uriel.schuurs@cito.nl>.

LUDO VERHOEVEN is hoogleraar Pedagogische Wetenschappen aan de Radboud Universiteit Nijmegen en coördinator van het onderzoeksprogramma *Learning & Plasticity* binnen het Behavioural Science Institute. Zijn primaire onderzoeksbelangstelling gaat uit naar processen van gesproken en geschreven taalontwikkeling in relatie met leerproblemen. Vanaf 1996 is hij tevens wetenschappelijk directeur van het landelijk Expertisecentrum Nederlands. Van daaruit is het concept van interactief taalonderwijs ontwikkeld waarbij voor mondelinge communicatie en geletterdheid tussendoelen en onderwijsprototypen zijn ontwikkeld.

## VERSCHENEN PROEFSCHRIFT



Coosje van der Pol (2010). *Prentenboeken lezen als literatuur*, een structuralistische benadering van het concept 'literaire competentie' voor kleuters. Delft: Eburon (Stichting Lezen Reeks deel 16). Isbn 978 90 597 2423 5, 296 blz.

Draagt het voorlezen van prentenboeken bij aan de literaire competentie van kleuters? Bij het beantwoorden van deze vraag hanteert Coosje van der Pol in haar proefschrift een structuralistische benadering van het lezen van literatuur, gebaseerd op de publicatie *Structuralist poetics: Structuralism, linguistics and the study of literature* van J. Culler (2002, 1975).

Om een tekst te kunnen lezen als literatuur moet de lezer kennis hebben van hoe die tekst is gestructureerd en met welke literaire middelen en conventies hij tot stand is gebracht. Het lezen van verhalen als literatuur is een vaardigheid die geleerd kan worden. Dat leerproces kan al starten in het kleuteronderwijs door tijdens het voorlezen op een impliciete manier aandacht te besteden aan die literaire middelen en conventies. Bijvoorbeeld dat een personage altijd is verzonnen en dat je dus niet bij hem op bezoek kunt gaan om hem iets te vragen. Bekendheid met dit literaire systeem achter verhalen kan resulteren in een rijkere leeservaring.

Deze structuralistische benadering van het prentenboek staat op nogal gespannen voet met de praktijk van het kleuteronderwijs. Daarin worden weliswaar heel veel prentenboeken voorgelezen, maar het ontsleutelen van de literaire betekenis van die verhalen speelt nauwelijks een rol. Een eerste gangbare reden om prentenboeken voor te lezen is taalontwikkeling, c.q. het uitbreiden van de woordenschat. In ruimere zin gaat het bij die taalontwikkeling ook om boekpromotie en leesbevordering. Een tweede reden waarom leerkrachten verhalen voorlezen is het stimuleren van de sociaal-emotionele ontwikkeling van kinderen. Als derde reden geldt dat prentenboeken voorgelezen worden ter introductie, ondersteuning en uitdieping van een (wereldoriënterend) thema. Terwijl bij het literaire lezen de aandacht in eerste instantie op het verhaal is gericht, fungeert in de huidige voorleespraktijk het verhaal vooral als kapstok voor andere leerprocessen, waarbij