

HET VERANTWOORD GEBRUIK VAN DIDACTISCHE LEEFTIJDSEQUIVALENTEN

De kritiek weerlegd

Gerard Melis, Paul Oosterveld en Johan Schokker

Het regelmatig toetsen van schoolvaardigheden is een niet meer weg te denken praktijk in het hedendaagse basisonderwijs. Een toetsscore van een leerling op een toets zegt op zichzelf echter niets, deze moet geïnterpreteerd worden. Daarvoor maakt men gebruik van normscores. Een van de manieren om een toetsscore te normeren is de Didactische LeeftijdsEquivalenten (DLE's).

Hoewel de toepassing van DLE's de afgelopen decennia wijdverbreid is in het basisonderwijs, en in bijna alle leerlingvolgsystemen is geïmplementeerd, bestaat er stevige kritiek op de methodologische steekhoudendheid van deze maat. In dit artikel worden de voornaamste kritiekpunten op het gebruik van DLE's besproken. Veel van deze kritiek blijkt bij nadere beschouwing onterecht, want gebaseerd op een verkeerde interpretatie van wat een DLE precies is.

1 Het nut van normeren

Het toetsen van leerlingen in het basisonderwijs heeft vooral ten doel om te bekijken waar een leerling staat ten opzichte van zijn of haar groepsgenoten en ten opzichte van de onderwijsdoelen, die aan het einde van het basisonderwijs in principe gehaald moeten worden (de zogenoemde 'referentieniveaus').

Een leerkracht kan daarbij gebruik maken van methodegebonden toetsen. De resultaten van dergelijke toetsen kunnen heel inzichtelijk zijn, maar hebben het nadeel dat ze afhankelijk zijn van de gevolgde onderwijsmethode en dat de vergelijking van de toetsresultaten dus enigszins beperkt is. In de beoordeling van de scholen eist de Onderwijsinspectie daarom het gebruik van niet-methodegebonden toetsen. Tegenwoordig zijn dat vooral de toetsen van het Cito LOVS.

Het is duidelijk dat een ruwe score op zo'n toets (bijvoorbeeld het aantal goed gemaakte opgaven) weinig inzichtelijk is. Niet alleen verschilt het aantal opgaven per toets, maar ook de moeilijkheid van de opgaven. Men kan over een leerling die de meeste opgaven van bijvoorbeeld een Begrijpend Lezen toets correct maakt, gerust beweren dat hij de stof waarop de toets in kwestie een beroep doet, voldoende beheerst. Om echter te weten hoe uitzonderlijk een dergelijke prestatie is, of hoe alledaags, dient men gebruik te maken van normen. Uiteraard is het van belang dat de normen zijn gebaseerd op een representatieve steekproef van Nederlandse

basisschoolleerlingen, zodat de scores van de normgroep een betrouwbare referentie zijn. Daarnaast is voor het toetsen van schoolvaardigheden ook het normeringsmoment van belang, want basisschoolleerlingen ontwikkelen zich in hun schoolvaardigheden gedurende het schooljaar, en leerlingen presteren – een enkele uitzondering daargelaten – aan het eind van het schooljaar beter dan aan het begin. Het Cito heeft haar toetsen in de regel in een specifieke maand genormeerd. Daarom dient in dezelfde periode waarin de toets is genormeerd ook de toets te worden afgenomen. Wanneer men een schoolvaardigheidstoets buiten de toetskalender om afneemt, raken de normen vertekend.

Er bestaan verschillende manieren om normscores te representeren. Wanneer de normen zijn gebaseerd op rangorde, zoals percentielscores, kan men aan de hand van de normscore bepalen waar een leerling in de normgroep staat. Een veel gebruikte indeling is de I t/m V-niveaus van het Cito. Een leerling in niveau I valt onder de 20% hoogst scorende leerlingen, terwijl een leerling in niveau V onder de 20% laagst scorende leerlingen valt. Daarnaast wordt de oudere Cito-indeling A- t/m E-niveau nog veel gebruikt. Een leerling op het E-niveau valt in het domein van de 10% laagst scorende leerlingen in de normgroep (zie tabel 1).

Tabel 1. Het Cito-normsysteem

I 20%	20% hoogst scorende leerlingen	A 25%	25% hoogst scorende leerlingen
II 20%	20% boven het landelijk gemiddelde	B 25%	25% ruim boven tot net boven het landelijk gemiddelde
III 20%	20% landelijk gemiddelde	C 25%	25% net tot ruim onder het landelijk gemiddelde
IV 20%	20% onder het landelijk gemiddelde	D 15%	15% ruim onder het landelijk gemiddelde
V 20%	20% laagst scorende leerlingen	E 10%	10% laagst scorende leerlingen

2 Het Didactische LeeftijdsEquivalent

Een alternatief voor de Cito-indeling vormt het Didactische LeeftijdsEquivalent (DLE). DLE is niet gebaseerd op rangorde, maar op het beheersingsniveau van de schoolvaardigheid en de onderwijsmaand dat gemiddeld genomen overeenkomt met dat beheer-

HET VERANTWOORD GEBRUIK VAN DIDACTISCHE LEEFTIJDSEQUIVALENTEN

singsniveau. De onderwijsmaand wordt uitgedrukt als Didactische Leeftijd (DL).

Aangezien er in het basisonderwijs elk leerjaar tien onderwijsmaanden zijn, en het basisonderwijs vanaf groep 3 uit zes leerjaren bestaat, heeft een basisschoolleerling aan het einde van zijn basisschoolloopbaan 60 maanden onderwijs gevolgd (afgezien van doublures of het eventueel overslaan van een leerjaar). Strikt genomen wordt de DL begrepen als het aantal onderwijsmaanden dat een leerling gemiddeld heeft genoten. Zo heeft een leerling in groep 4 eind januari 15 maanden onderwijs gekregen. Met andere woorden, de DL van een leerling is in groep 4 eind januari gelijk aan 15. Doordat een leerling een klas heeft overgeslagen of is blijven zitten, kan het totale aantal onderwijsmaanden in individuele gevallen afwijken. DLE's zijn echter een normsysteem, waardoor het niet nodig is dergelijke individuele verschillen te verdisconteren. Met andere woorden: de DL is geen persoonskenmerk, maar een klassekenmerk. In tabel 2 wordt een overzicht gegeven van de onderwijsmaand en de corresponderende DL.

Stel dat een toets Rekenen-Wiskunde in januari in groep 3 is genormeerd en er gemiddeld in de normgroep een score van 45 is behaald. Leerlingen die lager scores, scores dan ondergemiddeld en leerlingen die hoger scores bovengemiddeld. De toetscore 45 representeert de prestatie die men van een willekeurige leerling zou verwachten met een DL van 15. We spreken daarom van een Didactisch LeeftijdsEquivalent (DLE). Feitelijk is dit een andere uitdrukkingwijze; in plaats van de toetscore uit te drukken in percentielen, wordt een te verwachte score gegeven

bij een bepaalde onderwijsmaand. Sommigen leerlingen zullen daaraan voldoen, anderen zullen daarvan afwijken. Die afwijkingen worden bij de toepassing van DLE's niet ten opzichte van de groep geduid, maar ten opzichte van de onderwijsmaand.

Omdat een kind 60 maanden onderwijs krijgt, en er dus 60 didactische leermaanden zijn, loopt ook een DLE-schaal in principe van 0 tot 60; elke DLE staat voor een bepaald kennis- of beheersingsniveau, die wordt uitgedrukt in een toetsscore op een bepaalde toets. Zo meet de Schoolvaardigheidstoets Technisch Lezen (Teije de Vos, 2007) hoe vlot een leerling kan lezen. Daarvoor moet een leerling gedurende één minuut hardop een tekst lezen; het aantal woorden dat hij in die tijd correct kan lezen, is de toetsscore. Wanneer leerlingen in de normgroep in april in groep 5, dat is DL 28, gemiddeld 120 woordjes kan lezen, correspondeert een toetsscore van 120 met een DLE van 28. Wanneer een leerling in januari in groep 5 120 woordjes leest, dat is DL 25, krijgt hij een DLE van 28 en doet hij het dus beter dan de gemiddelde leerling.

Met behulp van DLE's kan men een leerling aan de hand van een toetsscore plaatsen in de onderwijsmaand van zijn niveau, en zo zijn relatieve leerachterstand of leervoorsprong bepalen. Daarbij zijn DLE's vooral nuttig om leervorderingen in beeld te krijgen. De niveauideling van het Cito, die relatief grof is, leent zich daar minder voor. Wanneer een leerling bijvoorbeeld op Rekenen-Wiskunde M5 een DLE scoort van eind groep 4, maakt dat inzichtelijk waar een leerling qua beheersingsniveau staat. Men kan dan zijn leerachterstand uitdrukken in het aantal onderwijsmaanden.

Tabel 2. Onderwijsmaand en DL

Maand	Groep 3	Groep 4	Groep 5	Groep 6	Groep 7	Groep 8
September	1	11	21	31	41	51
Oktober	2	12	22	32	42	52
November	3	13	23	33	43	53
December	4	14	24	34	44	54
Januari	5	15	25	35	45	55
Februari	6	16	26	36	46	56
Maart	7	17	27	37	47	57
April	8	18	28	38	48	58
Mei	9	19	29	39	49	59
Juni	10	20	30	40	50	60

HET VERANTWOORD GEBRUIK VAN DIDACTISCHE LEEFTIJDSEQUIVALENTEN

DLE's zijn met name ontworpen voor toetsen in het basisonderwijs. Er is immers in het basisonderwijs elk leerjaar een vast onderwijsprogramma van tien maanden, en het aanleren van schoolse vaardigheden is een activiteit die zich gedurende de gehele schoolloopbaan in het basisonderwijs afspeelt. Dit geldt voor bijna alle schoolse vaardigheden, hoewel er ook wel enige verschillen zijn. In de eerste leerjaren ligt het accent bijvoorbeeld op Technisch Lezen, in de latere leerjaren verschuift het accent naar Begrijpend Lezen. Maar voor veel schoolvaardigheden geldt een groeiend beheersingsniveau; leerlingen gaan gedurende hun schoolloopbaan in het basisonderwijs, – als het goed is – steeds beter rekenen en lezen.

3 Valkuilen bij het interpreteren van een DLE

Bij de interpretatie van DLE's dient men wel enige slagen om de arm te houden.

Ten eerste, stel dat een leerling de Cito schoolvaardigheidstoets Rekenen-Wiskunde M5 heel goed maakt, en zijn DLE situeert hem op een niveau begin groep 6. Dan kan men niet zeggen, om voor de hand liggende redenen, dat deze leerling de rekenstof van begin groep 6 beheerst; die stof is immers nog niet in het lesprogramma behandeld. Het is beter om te zeggen: de DLE-score wijst erop dat deze leerling de rekenstof van M5 beheerst als de gemiddelde leerling aan het begin van groep 6.

In dit verband is het interessant kennis te nemen van hetgeen het Cito opmerkt met betrekking tot het door hen gehanteerde begrip 'functioneringsniveau', waarbij dezelfde redenering wordt gegeven:

“Dit niveau geeft aan met welke gemiddelde leerling in het regulier basisonderwijs de vaardigheid van een getoetste leerling te vergelijken is. Een leerling heeft op de begrijpend leestoets M5 een vaardigheidsscore behaald van 8. Deze vaardigheidsscore komt overeen met een vaardigheidsscore van een gemiddelde leerling op toetsmoment M4. Het functioneringsniveau van de leerling is dus M4 [= DLE 15, red.]. Door de vaardigheidsscore om te zetten naar functioneringsniveaus, is in een oogopslag duidelijk welke lesstof bij de leerling past. Een functioneringsniveau M5 geeft aan dat deze leerling instructie

tot midden groep 5 vlot kan doorlopen”

(Cito Alternatief leerlingrapport, maart 2012)

Ten tweede, wanneer een leerling op de schoolvaardigheidstoets Rekenen-Wiskunde M5 een DLE van 20 behaalt, kan men concluderen dat hij op het niveau rekt van een gemiddelde leerling uit eind groep 4 en een leerachterstand heeft van vijf maanden. Uiteraard zegt dat op zichzelf niet of de leerling in kwestie in staat zal zijn deze achterstand in te halen en op welke termijn. Dat hangt helemaal af van de situatie van de leerling, de aard van zijn problematiek, hoe hij die leerachterstand heeft opgelopen, enzovoorts.

De methodiek van DLE's maakt het dus mogelijk een leerling op een niveau te plaatsen dat boven of onder zijn eigenlijke onderwijsmaand valt. Dat heeft echter wel zijn grenzen; het is nogal onzinnig om een leerling die uitstekend een schoolvaardigheidstoets Rekenen-Wiskunde M5 heeft gemaakt, een DLE toe te schrijven die gesitueerd is in groep 7. Wanneer een leerling duidelijk buiten het vaardigheidsniveau valt waar de toets een beroep op doet – hij is of veel té goed of veel té slecht ten opzichte van het vaardigheidsniveau dat door de toets wordt gemeten –, moet men een dergelijke DLE eerder beschouwen als een indicatie dat niet de juiste toets is afgenomen.

Om deze reden wordt ook in de DLE-tabellen, die worden opgesteld voor de indicatiestelling voor LWOO en PrO (Melis, 2012) bij een leerling die meer dan een jaar boven of onder het DLE van het eigenlijke normeringsmoment scoort (dus een leerjaar boven of onder waarvoor de toets is geconstrueerd), aanbevolen 'terug te toetsen' of 'door te toetsen'. Dat wil zeggen dat een leerling een makkelijkere of een moeilijker toets moet krijgen. Bij een tweede toetsafname kan men dan preciezer het niveau van de leerling bepalen zodat hij niet teveel opgaven voorgelegd krijgt die te moeilijk of te makkelijk voor hem of haar zijn.

Het door- of terugtoetsen is mogelijk omdat in een leerlingvolgsysteem voor elk leerjaar aparte toetsen worden ontwikkeld op het geëigende niveau. Zo beschikt het Cito LOVS over een reeks van toetsen Rekenen-Wiskunde, vanaf groep 3 tot en met groep 8.

In zekere zin zou je kunnen zeggen dat het steeds om dezelfde vaardigheid gaat, met toenemende complexiteit. Wanneer deze aanname geldt, is het moge-

HET VERANTWOORD GEBRUIK VAN DIDACTISCHE LEEFTIJDSEQUIVALENTEN

DLE
BOEK

lijk de vaardigheid op één schaal te zetten. Met behulp van de itemrespons-theorie en een procedure die bekend staat als *test-equating* worden door het Cito de vaardigheidsscores van alle toetsen van een bepaalde vaardigheid op een en dezelfde schaal gezet. Hoe hoger de vaardigheidsscore, des te groter de vaardigheid van een leerling. Elke toetsscore die op een bepaalde toets wordt behaald, correspondeert met een vaardigheidsscore. Zo kan men in principe het niveau van een leerling door alle leerjaren van het basisonderwijs vergelijken. Wat dat betreft volgen vaardigheidsscores en DLE's hetzelfde principe, namelijk dat de te toetsen vaardigheid op één schaal kan worden geplaatst, van geen tot weinig beheersing naar een hoge beheersing, voor alle leerjaren.

Van een schaal is duidelijk sprake wanneer men steeds dezelfde toets hanteert om een vaardigheid te meten in alle leerjaren van het basisonderwijs. De Schoolvaardigheidstoets Technisch Lezen is bijvoorbeeld zo'n toets; in alle leerjaren krijgen leerlingen dezelfde tekst voorgelegd en wordt gekeken hoeveel woordjes de leerlingen in één minuut correct kunnen lezen. Deze toets is in meerdere leerjaren genormeerd en kan daardoor op verschillende momenten in de basisschoolloopbaan worden afgenomen.

Maar bij de meeste schoolvaardigheidstoetsen is dit niet het geval. Zo kent het Cito LOVS voor de meeste schoolvaardigheden aparte toetsen voor elk leerjaar. Wanneer men met DLE's werkt, is het dus van belang goed in de gaten te houden voor welk leerjaar de toets is bedoeld. De DLE-schaal werkt optimaal voor het leerjaar waarvoor de toets ook is genormeerd, het gebruik van DLE's wordt hachelijker als de DLE duidt op een DL die te ver van dat leerjaar is verwijderd. Dat kan men ook zien doordat het domein van scores op een toets, de DLE-schaal niet langer volledig dekt, en er grote gaten vallen in de schaal.

4 Bezwaren tegen DLE's

Hoewel er in het basisonderwijs veel wordt gewerkt met DLE's, is er ook kritiek op het gebruik ervan. Maatgevend in de kritische perceptie van DLE's is een artikel van Arne Evers en Wilma Resing (Evers & Resing, 2007). De titel van hun artikel liegt er niet om: 'Het drijfzand van didactische leeftijdsequivalenten'. Evers en Resing brengen tien bezwaren tegen het gebruik van DLE's naar voren. We geven steeds kort hun bezwaar weer en gaan daar vervolgens op in.

1. *De veronderstelling dat scores op een toets in de tijd lineair zullen toenemen en voor alle leerlingen eenzelfde verloop zullen hebben, blijkt vaak niet houdbaar. Kinderen leren niet volgens een regelmatige stabiele curve en zeker niet volgens een strikt lineair patroon.*

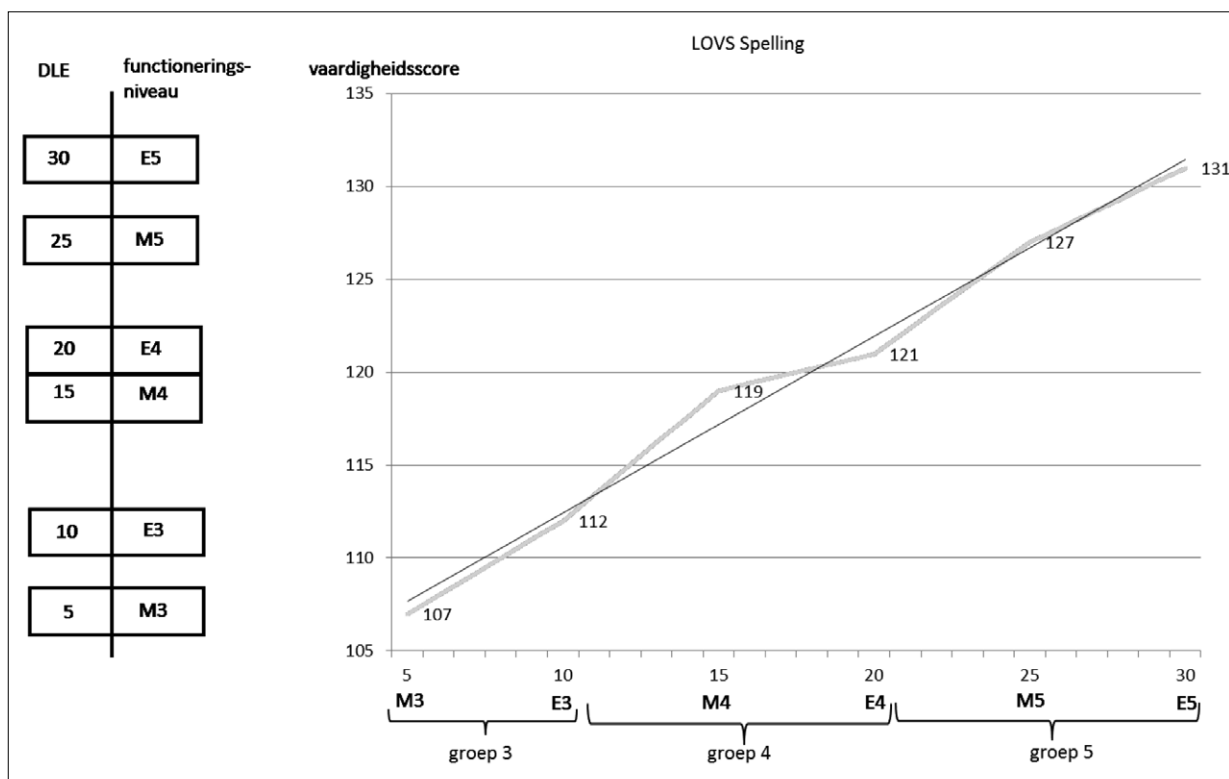
De gedachte dat de schoolvaardigheden bij leerlingen op lineaire wijze toenemen, is uiteraard van elke theoretische rechtvaardiging of praktische aanwijzing ontbloeit; leerlingen kunnen onverwachte leersprongen doormaken, en ook tijdelijke (en in een enkel geval permanente) terugval vertonen. Schoolprestaties zijn afhankelijk van een gevarieerde reeks factoren, zoals de aanleg, de motivatie en het welbevinden van het kind, de sociaaleconomische status van het gezin, de kwaliteit van de leerkracht en het leermateriaal, enzovoorts. Over het belang van al deze factoren bestaat veel wetenschappelijke literatuur, maar over de wijze waarop al deze factoren met elkaar interacteren bestaan slechts vermoedens; de werkelijkheid van het schoolgaande kind is gewoonweg te complex. Er lijkt in elk geval weinig evidentie te bestaan voor de veronderstelling van een 'regelmatige stabiele curve'.

Dit is echter ook geen veronderstelling van de toepassing van DLE's. DLE's zijn immers normscores. Ze pretenderen dan ook een beschrijving te geven van de norm en niet van het leertraject van de individuele leerling. Dat leerlingen op individueel niveau allerlei onregelmatige leercurves laten zien, sluit niet uit dat een lineaire ontwikkeling een goede beschrijving geeft van de gemiddelde leerling. De 'gemiddelde leerling' is een *fictief* construct, maar dat neemt niet weg dat het vergelijken van de prestaties op een toets van een individuele leerling met een dergelijk construct nuttig kan zijn.

Ook is het niet helemaal correct te beweren dat een DLE-schaal uitgaat van een strikt lineaire ontwikkeling, zoals Evers en Resing doen. Vaak is een toets namelijk op meerdere momenten in een schooljaar genormeerd. De interpolatie tussen de nomeringsmomenten gaat inderdaad uit van een lineaire toename, maar wanneer er ook halverwege het schooljaar normeringsgegevens zijn verzameld, hoeft de vaardigheidsbeheersing, en daarmee ook de onder de opgestelde DLE-schaal gepositioneerde normscores, niet perse lineair toe te nemen. In een DL/DLE-grafiek wordt de tijd als meeteenheid genomen – na een bepaald aantal maanden

HET VERANTWOORD GEBRUIK VAN DIDACTISCHE LEEFTIJDSEQUIVALENTEN

Grafiek 1. DLE-schaal over verschillende normeringsmomenten



onderwijs verwacht men dat de gemiddelde leerling een daarbij behorende normscore behaalt – terwijl in de schaalgrafieken van de Cito LOVS-toetsen de vaardigheidsscore de meeteenheid is.

Nemen we het voorbeeld van LOVS Spelling. Er is een DLE-schaal opgesteld voor de didactische leermaanden van 5 tot 30. Wanneer we alleen over de normeringsmaanden M3 en E5 zouden beschikken, zouden we op basis van de getrokken stippelijntje tussen deze normeringsmomenten een DLE-schaal opstellen. We beschikken echter daarnaast over gegevens van de normeringsmaanden E3, M4, E4 en M5. Daaruit blijkt dat in termen van vaardigheidsscores, de groei tussen E4 (DL/DLE 20) en M5 (DL/DLE 25) groter is dan tussen M4 (DL/DLE 15) en E4 (DL/DLE 20). Omdat we over deze gegevens beschikken, kunnen we hiermee rekening houden in de opstelling van de DLE-schaal. Van werkelijke lineariteit is dus geen sprake. Uit dit voorbeeld blijkt ook dat de kwaliteit van een DLE-schaal afhankelijk is van het aantal normeringsmomenten. In bovenstaande voorbeeld beschikken we om de vijf maanden over normeringsgegevens. Interpolatie tussen de normeringsmomenten vindt

wel plaats, maar als de perioden tussen de verschillende normeringsmomenten niet te groot zijn, is dit een aanvaardbare procedure. Zeker wanneer men bedenkt dat men bij het gebruik van percentiëlnormen ook dikwijls gedwongen is normen toe te passen buiten de eigenlijke normeringsperiode. Hierdoor kunnen vertekeningen optreden, maar die zullen gering zijn wanneer men niet al te zeer afwijkt van de normeringsperiode. Zo geldt ook voor een DLE-schaal dat interpolatie aanvaardbaar is, wanneer de afstand die is gelegen tussen de verschillende normeringsmomenten, niet te groot is.

2. Er wordt geen rekening gehouden met bodem- en/of plafondeffecten.

Het is waar dat toetsen met een bodem- en/of plafondeffect niet kunnen differentiëren tussen respectievelijk zwakke of goede leerlingen. Maar het is de vraag of dit als een bezwaar kan gelden in het gebruik van DLE's. Het optreden van bodem- en/of plafondeffecten duidt over het algemeen op een onjuiste toepassing van toetsen (bijvoorbeeld wanneer men een Rekenen-Wiskunde toets afneemt die veel te moeilijk is voor de leerlingen) of een slechte toets-

HET VERANTWOORD GEBRUIK VAN DIDACTISCHE LEEFTIJDSEQUIVALENTEN

constructie. Een toets met een duidelijk bodem-effect is dan ook niet geschikt om zwakke leerlingen te differentiëren. Dat staat echter los van de DLE als normscore. Ook bij andere normscores, zoals de Cito-niveaus A tot en met E, treden problemen op in geval van bodem- en plafondeffecten.

3. *Kleine scoreverschillen kunnen leiden tot aanzienlijke verschillen in DLE's.*

Dat is een gelijksoortig bezwaar als bezwaar 2. Wanneer men bijvoorbeeld een schoolvaardigheidstoets die bedoeld is voor groep 4, afneemt bij een groep in de bovenbouw, kan men verwachten dat de meeste leerlingen een maximale score zullen behalen (plafondeffect). Het aantal fouten dat dan gemaakt wordt, is gewoonweg te klein om de leerlingen op betrouwbare wijze te kunnen differentiëren. Dit vertaalt zich ook in de DLE-schaal, waar in de bovenste regionen van de schaal 'gaten' vallen.

Overigens kan hetzelfde bezwaar gelden voor de wijze waarop het Cito toetsscores normeert: als de scores in de normgroep weinig verschillen, worden toch sommige leerlingen als A en anderen als E geclassificeerd, hoewel die classificatie gebaseerd is op kleine verschillen.

Wanneer kleine scoreverschillen leiden tot aanzienlijke verschillen in DLE's, is dat dus een aanwijzing dat de toets in dat scoredomein niet voldoende differentieert. Meestal betekent dat, dat de toets te moeilijk of te makkelijk is voor de betreffende leerlingen.

4. *Bij het opstellen van de DLE-schaal worden DLE's bepaald buiten de normeringsmomenten door middel van extrapolatie.*

In sommige situaties vindt men het wenselijk normscores door te trekken naar een lager of een hoger leerjaar door middel van extrapolatie. Dit is echter een hachelijke praktijk; er zijn ter ondersteuning namelijk geen toetsgegevens beschikbaar. Men neemt in deze situaties aan dat de lineaire toename van de vaardigheidsbeheersing zich voortzet, zonder dat daarvoor voldoende empirische evidentie beschikbaar is. In de tabellen zoals die in het DLE Boek worden gepresenteerd, komt deze praktijk dan ook niet voor. Dit kan dus niet als een algemeen geldend argument tegen DLE's worden opgevoerd.

5. *Bij het opstellen van de DLE-schaal worden 'onmogelijke' DLE's meegenomen.*

De DLE-schaal is ontworpen voor het basisonderwijs. Het voortzetten van de schaal na de overgang naar het voortgezet onderwijs is vooral twijfelachtig wanneer de schaal is opgesteld door middel van extrapolatie. In de praktijk scoren leerlingen van het eerste leerjaar vmbo op bijvoorbeeld de Tempo Toets Rekenen van Teije de Vos op een DLE rond de 40 en 50, dus op het niveau van leerlingen uit groep 6 en 7 van het basisonderwijs. Hoewel de toets niet specifiek voor deze groep is genormeerd, kan men niet ontkennen dat een dergelijke normscore niet inzichtelijk is; een lwoo-leerling blijkt dan het niveau te hebben van een leerling uit groep 5. In het ideale geval beschikt men ook over percentielscores, zodat naast de leerachterstand van de leerling ten opzichte van de gemiddelde basisschoolleerling, ook kan worden bepaald hoe erg dat is in vergelijking met leerlingen van het eigen onderwijsniveau.

In het DLE Boek worden geen DLE-scores gerapporteerd boven de 60, tenzij er bij een toets sprake is van een of meer genormeerde afnamemomenten in het voortgezet onderwijs, maar dit heeft dan ook een duidelijke rationale.

Daarnaast is ook besloten om in het DLE Boek negatieve DLE-scores op te nemen. Op het eerste gezicht komt dit wat vreemd over. DLE's waren aanvankelijk bedoeld voor het basisonderwijs, zonder dat daarbij het kleuteronderwijs is inbegrepen. Het Cito heeft echter toetsen voor kleuters uitgebracht, zoals Taal voor Kleuters en Rekenen voor Kleuters. Om de toetscores op ook deze toetsen uit te kunnen drukken in DLE's, is de schaal onder het nulpunt doorgetrokken. Dat maakt op het eerst gezicht weliswaar een wat vreemde indruk, maar veronderstelt geen principiële bezwaar tegen het gebruik van DLE's in deze leerjaren.

6. *Onjuiste veronderstelling van gelijke spreiding van de scores.*

DLE's zijn gebaseerd op gemiddelde en/of mediane scores (de mediaan geniet de voorkeur want deze is minder sterk afhankelijk van extreme waarden) op de verschillende normeringsmomenten; over de spreiding wordt geen vooronderstelling gedaan. Er is ook geen directe relatie tussen percentielscores en DLE-scores. In dit verband halen Evers en Resing

HET VERANTWOORD GEBRUIK VAN DIDACTISCHE LEEFTIJDSEQUIVALENTEN

Tabel 3. Toetsscores en normscores PI-dictee van twee leerlingen

DL	Score	Percentiel	DLE	Leer-achterstand	Leerr-ndement totaal	Leerrndement tussen toetsmomenten
Leerling A						
5	6	5	1	4	20%	
15	42	15	11	4	73%	100%
25	70	21	21	4	84%	100%
35	92	28	31	4	89%	100%
Leerling B						
5	6	5	1	4	20%	
15	31	5	8	7	53%	70%
25	59	5	17	8	68%	90%
35	75	5	23	12	66%	60%

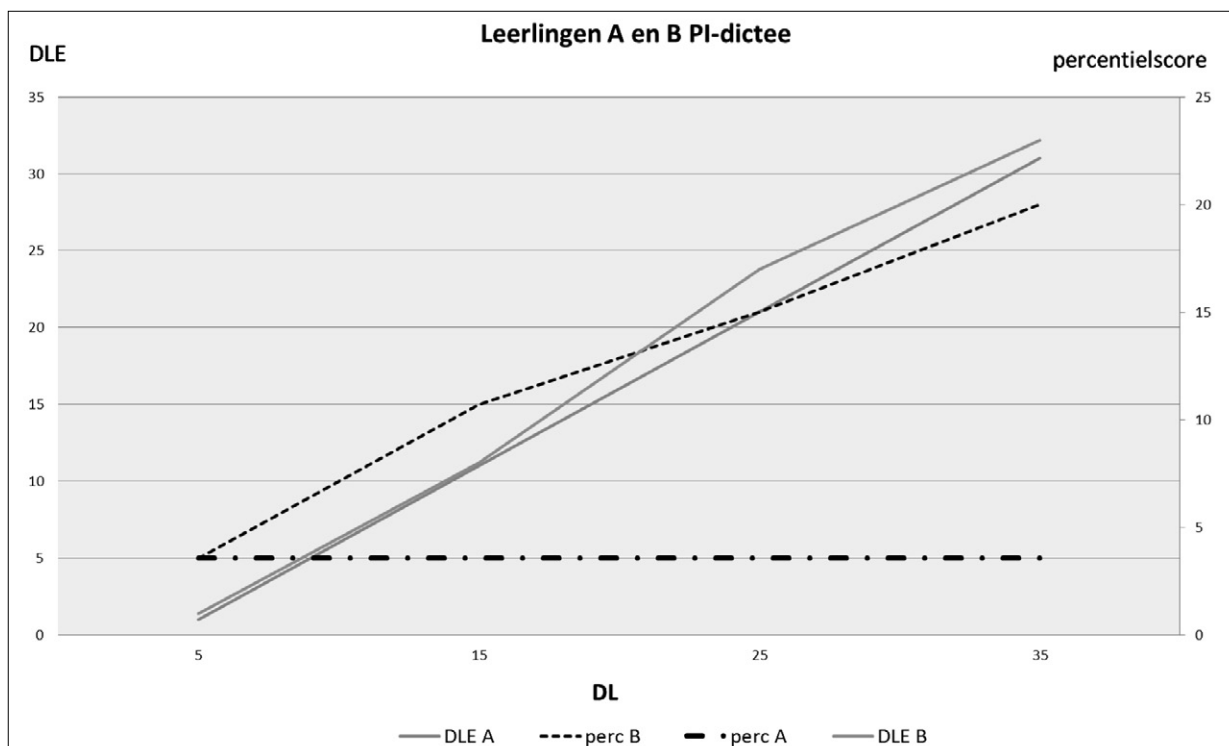
een voorbeeld van Geelhoed en Reitsma aan over het PI-dictee. Het PI-dictee is een spellingstoets voor het basisonderwijs, die kan worden afgenomen vanaf groep 3. In groep 3 kunnen de leerlingen nog niet zoveel woorden schrijven, dus valt de ruwe score relatief laag uit. We veronderstellen nu dat we het PI-dictee in januari bij leerlingen uit groep 3 afnemen en in de daarop volgende drie leerjaren nemen we nogmaals deze zelfde toets af.

Dit levert voor leerling A en leerling B de volgende scores op (zie tabel 3, Geelhoed en Reitsma, p. 30-31).

Leerling A heeft dusdanig gescoord dat zijn leerachterstand steeds vier maanden is gebleven, de scores van leerling B zijn dusdanig dat hij steeds een percentielscore van 5 heeft behaald, dat wil zeggen dat hij zich bevindt in het gebied van de 5% zwakst scorende leerlingen van de normgroep.

Het kritiekpunt is nu dat wanneer men uitsluitend oog heeft voor de leerachterstand, bij leerling A de indruk ontstaat dat er geen verbetering van de leerprestaties heeft plaatsgevonden. Zijn leerachterstand blijft immers dezelfde.

Grafiek 2. DLE en percentielscores van leerling A en leerling B



HET VERANTWOORD GEBRUIK VAN DIDACTISCHE LEEFTIJDSEQUIVALENTEN

Dit is echter een merkwaardige interpretatie van de gegevens. Een stabiele achterstand is namelijk niet een stilstand in de ontwikkeling. Leerling A haalt elk schooljaar een DLE-score die 10 punten hoger ligt dan het jaar daarvoor (per 10 maanden onderwijstijd steeds 10 maanden vooruitgang; een leerrendement van 100%). Alleen is de leerling niet in staat gebleken de aanvankelijke leerachterstand van vier maanden bij de eerste meting in te halen.

Wanneer men bij de interpretatie van de normscores ook de percentielscores betreft, is dat zeker inzichtelijk. Daaruit blijkt dat leerling A van een relatief slechte leerling zich op het PI-dictee heeft verbeterd tot een bijna gemiddelde leerling. Deze relatieve positieverbetering ten opzichte van de medeleerlingen is niet direct af te leiden uit de DLE-score.

Daar staat tegenover dat wanneer men uitsluitend gebruik maakt van percentielscores, dit ook zijn beperkingen kent. Dit blijkt uit het voorbeeld van leerling B, dat eveneens door Geelhoed en Reitsma wordt gegeven. Leerling B behaalt steeds een percentielscore van 5, en daaruit is moeilijk af te leiden of deze leerling ondanks zijn constante percentielscore zich in de spelling heeft verbeterd. Leerling B is een zwakke leerling, maar door te kijken naar de DLE's wordt duidelijk in hoeverre hij zich in het spellen heeft verbeterd en krijgt men inzicht in zijn leertempo. Wanneer men uitsluitend naar de percentielscores kijkt, zou men gemakkelijk het idee kunnen krijgen dat deze leerling stil is blijven staan. De DLE's van leerling B maken duidelijk dat hij wel degelijk vooruit gaat, hoewel zijn leerachterstand in de loop der jaren is toegenomen.

7. *Misverstand dat leerachterstand ook gelijk staat aan benodigde leertijd.*

Het idee van Evers en Resing dat een leerachterstand van een leerling in termen van DLE's ook aangeeft wat de benodigde leertijd van de leerling zou zijn om de betreffende achterstand in te halen, berust op een misvatting. De DLE's doen onder meer een uitspraak over de *gemiddelde* leerachterstand van een leerling; het gebruik ervan is vooral nuttig om te bezien, aan de hand van de toetsresultaten van een leerlingvolgsysteem, of de eventuele leerachterstand toe- of afneemt. Zoals eerder aangegeven, kan een leerachterstand uitgedrukt in DLE's niet direct worden vertaald in het aantal maanden onderwijs

dat een specifieke leerling nodig zou hebben om die achterstand in te halen; dat hangt geheel van de betreffende leerling af en zijn of haar omstandigheden.

8. *Onzuiverheid van de normgroep; gedoubleerde leerlingen.*

Volgens Evers en Resing wordt het gebruik van DLE's vertroebeld door onzuiverheid van de normgroep, waarin immers ook gedoubleerde leerlingen zijn opgenomen. Zoals hiervoor aangegeven, dient de DL te worden begrepen als het aantal onderwijsmaanden dat de gemiddelde leerling op dat specifieke moment heeft genoten. Dat moment correspondeert met een gemiddeld onderwijsniveau in de normgroep. Gedoubleerde leerlingen maken altijd deel uit van de populatie basisschoolleerlingen in een bepaald leerjaar en moeten daarom ook niet buiten de normgroep worden gehouden.

Toetsontwikkelaars zoals het Cito sluiten voor de bepaling van het niveau A t/m E gedoubleerde leerlingen evenmin uit; zowel daar als bij de daaraan ontleende DLE-schaal is sprake van 'onzuiverheid' van de normgroep.

Overigens, wanneer men te maken heeft met een gedoubleerde leerlingen wordt meestal aangeraden de DL van zo'n leerling aan het begin van het leerjaar tien maanden terug te zetten. De leerling presteert immers op het onderwijsniveau waarin hij wordt onderwezen, en ook het toetsen vindt plaats op dat niveau. Men is geïnteresseerd in hoeverre een leerling de onderwezen leerstof beheerst, en het zou niet terecht zijn voor een gedoubleerde leerling een maatstaf te hanteren van een klas hoger; hij is immers niet voor niets gedoubleerd.

9. *Toetsen hebben te weinig opgaven om een volledige DLE-schaal op te stellen.*

Om een volledige DLE-schaal op te stellen, dient een toets inderdaad uit ten minste 60 opgaven te bestaan. Dit is in de regel niet het geval, omdat een toets zelden is bedoeld voor alle leerlingen van het basisonderwijs, en in de praktijk de afnameperiode van een toets beperkt is. De DLE-schaal wordt dan ook in de praktijk meestal opgesteld voor een beperkt domein, wat betekent dat men de toetsscore maar met een beperkt aantal DLE's kan verbinden. Voor het overgrote deel van de gevallen voldoet een dergelijke schaal ook.

HET VERANTWOORD GEBRUIK VAN DIDACTISCHE LEEFTIJDSEQUIVALENTEN

Bovendien speelt het probleem niet wanneer, zoals bij het Cito LOVS, de toetsen voor de verschillende leerjaren op één vaardigheidsschaal worden gezet door middel van itemrespons-technieken.

10. De leerstof is niet vergelijkbaar over de verschillende leerjaren, en scholen verschillen in het moment waarop ze de leerstof aanbieden.

Zoals al eerder aangegeven, dient men precies te zijn in de interpretatie van een DLE. Het is duidelijk dat men geen uitspraak kan doen over een toetsinhoud die ook niet getoetst wordt. Dat scholen verschillen in de tijd waarop ze bepaalde lesstof aanbieden, is voor gebruik van DLE's niet relevant, maar betreft het verschil tussen methodegebonden toetsen en methodeonafhankelijke toetsen.

5 Besluit

Geen van de tien bezwaren die Evers en Resing hebben aangedragen tegen het gebruik van DLE's is zwaarwegend. Vaak blijkt het te gaan om beperkingen van de toets, zoals bodem- en plafondeffecten, die in de DLE-schaal tot uiting komt, of om een onjuist opgestelde DLE-schaal. Natuurlijk bestaan er DLE-schalen met duidelijke mankementen, maar dat kan niet gelden als een principieel bezwaar tegen het gebruik van DLE's.

Sinds kort heeft het Cito in het al langer bestaande 'alternatief leerlingrapport'- bedoeld voor leerlingen die zich anders ontwikkelen dan gemiddeld (zoals de Cito-brochures "Het alternatief leerlingrapport" en "Toetsscore, vaardigheidsscore... en dan?" het formuleren) – nadrukkelijker en duidelijker dan voorheen een koppeling aangebracht tussen vaardigheidsscore en onderwijstijd:

"Leerlingen met een achterstand behalen meestal een vaardigheidsniveau E of V. Deze niveauaanduiding zegt niets over de groei die zij doormaken. Daarom staat in het alternatief leerlingrapport de ontwikkeling van de vaardigheidsscore in de tijd centraal."

Met dat doel brengt het Cito een correspondentie aan tussen vaardigheidsscore en zogenoemde 'functioneringsniveaus' in termen van de genormeerde toetsmomenten M3, E3, M4, M5, E5, enzovoorts.

Het aanbrengen van deze koppeling is voor het Cito van belang met het oog op het opstellen van een ontwikkelingsperspectief. Dit functioneringsniveau doet sterk denken aan de DLE, alleen wordt een toetsscore van een leerling niet uitgedrukt in onderwijsmaand, maar in de iets bredere term van onderwijsperiodes als M3, E3, M4, enzovoorts. Het concept van de 'functioneringsniveaus' komt dus overeen met dat van de DLE, met dit verschil dat een DLE-score specifiek wordt uitgedrukt en daarmee ook eenduidiger gebruikt kan worden bij het opstellen van een leerrendementsverwachting als aspect van het ontwikkelingsperspectief. De argumenten die het Cito aandraagt voor het hanteren van een functioneringsniveau zijn ook de argumenten die pleiten voor het gebruik van DLE's; het geeft een duidelijke indicatie van het niveau van de vaardigheid van de leerling en is geschikter om de vorderingen van een individueel leertraject in beeld te brengen. Het blijft wel enigszins merkwaardig dat de argumenten die voor het Cito juist lijken te pleiten voor het gebruik van het functioneringsniveau, niet lijken te pleiten voor het gebruik van DLE's.

Wat de critici van DLE's kan worden verweten, is dat ze blijkbaar zo content zijn met hun manier van kijken dat ze geen of weinig aandacht hebben voor het feit dat de meeste leerkrachten graag (ook) op een andere manier naar de gegevens kijken. Een manier van kijken waarmee ook inzichtelijker aan de ouders over de leerontwikkeling van hun kinderen verslag kan worden gedaan. Het is zeker niet zo dat de ene manier beter is dan de andere: het hangt er vanaf waar je naar wilt kijken. Om sommige zaken te zien moet je op een bepaalde manier kijken, om andere zaken te zien moet je op een andere manier kijken. Dat blijkt bijvoorbeeld uit het besproken voorbeeld van de twee leerlingen van het PI-dictee, dat Geelhoed en Reitsma juist aandragen als een argument tegen het gebruik van DLE's. Feit is dat leerkrachten het gebruik van DLE's in het algemeen nuttig vinden, dus graag op ook die manier naar de gegevens kijken.

Het is bovendien verstandig om bij de interpretatie van de DLE-score, en dan met name de ontwikkeling ervan bij individuele leerlingen, ook de rangordenormen zoals percentielscores te betrekken. Het komt ons voor dat juist de combinatie van beide het meest volledige beeld kan opleveren. Vooral ook omdat de leerlingvolgsystemen in het basisonderwijs zich vaak beperken tot de relatief grove Cito-niveaus.

HET VERANTWOORD GEBRUIK VAN DIDACTISCHE LEEFTIJDSEQUIVALENTEN

DLE
BOEK

Literatuur

- Arne Evers en Wilma Resing. (2007). Het drijfzand van didactische leeftijdsequivalenten. *De Psycholoog*, 466-472.
- Cito. (2012). *Alternatiefleerlingrapport*.
- Cito. (2012). *Ontwikkelingsperspectief vaststellen op basis van vaardigheidsscores*.
- Cito. (2012). *Toetsscore, vaardigheidsscore... en dan?*
- Gerard Melis. (1995 en later). *DLE Boek*. Amsterdam: Boom test uitgevers.
- Gerard Melis. (2012). *DLE Schalen indicatiestelling LWOO en Pro voor instroom in schooljaar 2013-2014*. Amsterdam: Boom test uitgevers.
- Henriëtte de Groot, Marjan Huizing en Anneke Noteboom. (1999). *Leerlingvolgsysteem en speciale leerlingen*. Arnhem: Cito.
- Jan Geelhoed en Pieter Reitsma. (2004). *PI-dictee*. Amsterdam: Pearson.
- G.N. Melis en F.R. Sonsma. (1989). *Het SAVU-leerlingvolgsysteem. Functie-Opzet-Invoering*. Hoevelaken: Christelijk Pedagogisch Studiecentrum.
- G.N. Melis. (1991). DLE en leerling volgsysteem. Ja, nee... toch maar; of... juist wel doen?! *School & Begeleiding*, 29, 71-79.
- A.H.J. Moelands, M.J.C. Mommers en J.H.L. Oud. (1990). Leerlingvolgsystemen verklaard en vergeleken. *School & Begeleiding*, 26, 19-28.
- J.H.L. Oud en M.J.C. Mommers. (1990). De valkuil van het didactisch leeftijdsequivalent. *Tijdschrift voor Orthopedagogiek*. 29, 445-459.
- Teije de Vos. (2007). *Schoolvaardigheidstoets Technisch Lezen*. Amsterdam: Boom test uitgevers.
- Teije de Vos. (2012). *De geschiedenis van de DLE's*. <http://testen-en-toetsen.blogspot.com/2012/04/de-geschiedenis-van-de-dles.html>
- Toetsen voor speciale leerlingen. Voor groep 3, 4 en 5*. Arnhem: Cito.

Over de auteurs

Gerard Melis is als orthopedagoog/gz-psycholoog werkzaam in de stafdienst Zorg van een grote brede scholengemeenschap voor voortgezet onderwijs en auteur van onder meer het DLE Boek (Boom test uitgevers, 4^e druk, 2012).

Paul Oosterveld is methodologisch en psychometrisch adviseur. Hij was vijf jaar werkzaam als universitair docent bij de UvA en is auteur van de d2 Aandachts- en concentratietest voor kinderen.

Johan Schokker is werkzaam als methodoloog bij Boom test uitgevers.

Dit artikel verscheen in verkorte vorm in het septembernummer van het *Tijdschrift voor Remedial Teaching* (jaargang 20, 2012, nummer 4, pp. 30-33).