

# Werken met leerlingen met hoge vaardigheidsscores in het Cito Volgsysteem primair en speciaal onderwijs

## SAMENVATTING

Het Cito Volgsysteem primair en speciaal onderwijs kan behulpzaam zijn bij het afstemmen van het onderwijs op leerlingen die het heel goed doen op school en om die reden extra aandacht verdienen. Ook voor deze leerlingen geven de toetsen van het volgsysteem valide en betrouwbare resultaten. Alleen bij leerlingen die alles, of bijna alles, goed hebben is de toetsscore wat minder betrouwbaar. Als een leerling alle opgaven van een toets correct beantwoordt, weten we immers alleen dat de leerling goed is, maar niet precies hoe goed. Hoe betrouwbaar de toetsscore moet zijn, hangt ervan af of een leerkracht een leerling met de toetsscores volgt over de tijd, of dat de leerkracht de toetsscores gebruikt om te bepalen of de leerling toe is aan versnellen, verrijken of verdiepen. In het laatste geval is het voldoende om aan te tonen dat de leerling een forse voorsprong heeft; een minder betrouwbare toetsscore is geen groot bezwaar. Om een leerling optimaal te kunnen volgen over de tijd is een meer betrouwbare toetsscore wenselijk. Dit is te realiseren door een toets van een hoger niveau af te nemen.

In dit artikel gaan we in op de vraag wat LVS-toetsen kunnen betekenen voor het werken met leerlingen die hoge vaardigheidsscores behalen

## 1 Inleiding

Het Cito Volgsysteem primair en speciaal onderwijs, ook wel LVS genoemd, is een geheel van toetsen en rapportages waarmee leerkrachten de leervorderingen van leerlingen in kaart kunnen brengen en deze kunnen volgen over de tijd. Dit gebeurt door de vaardigheidsscores die leerlingen behalen bij verschillende leerstofgebieden te vergelijken met die van leeftijdsgenoten en door ontwikkelingen in het prestatieniveau van de leerlingen te monitoren. Aan de hand van de rapportages kunnen leerkrachten tijdig bijsturen als dit nodig is. Bij het gebruik van LVS-toetsen lijkt in de praktijk vaak de nadruk te liggen op de groep leerlingen die zwak presteren of hiaten in hun leerprestaties laten zien en daarom extra hulp kunnen gebruiken. LVS-toetsen kunnen echter ook behulpzaam zijn bij het afstemmen van het onderwijs op leerlingen die het heel goed doen op school en daarom speciale aandacht nodig hebben.

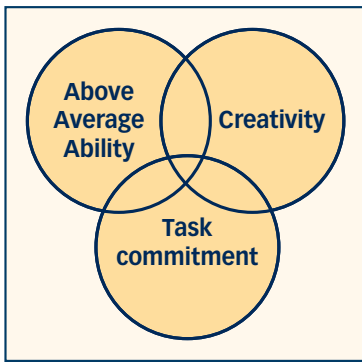
In dit artikel gaan we in op de vraag wat LVS-toetsen kunnen betekenen voor het werken met leerlingen die hoge vaardigheidsscores behalen. Allereerst geven we een omschrijving van de groep leerlingen die we bedoelen. Om welke leerlingen

gaat het en hoe verhouden ze zich tot de leerlingen die in de literatuur worden aangeduid met de term begaafd of hoogbegaafd? Vervolgens gaan we in op de vraag hoe het onderwijs kan inspelen op de behoeften van deze leerlingen. Om die afstemming goed te kunnen realiseren, is het nodig om te weten hoe goed de leerlingen zijn in de verschillende leerstofgebieden. We stuiten daarbij op het probleem dat de LVS-toetsen in de hoogste prestatiegebieden in de regel minder betrouwbaar meten. We gaan dieper in op de oorzaken daarvan, behandelen de begrippen ‘lokale betrouwbaarheid’, ‘betrouwbaarheidsinterval’ en ‘plafondeffect’, en presenteren een oplossing. Ten slotte gaan we in op de veelgehoorde vraag of hoog presterende leerlingen op sommige opgaven ‘anders’ reageren dan hun klasgenoten. Beantwoorden zij sommige relatief eenvoudige opgaven bijvoorbeeld fout doordat ze anders redeneren?

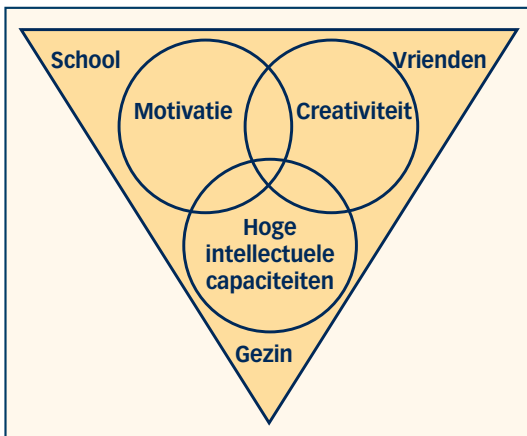
## 2 Leerlingen met hoge vaardigheidsscores

Historisch gezien kent de aandacht voor de hoog presterende leerling in het onderwijs een vrij cyclisch karakter. Het onderwerp mag zich verheugen in een wisselende belangstelling, waarbij uiteenlopende termen vanuit een verschillende invalshoek worden gebruikt. In de jaren tachtig van de vorige eeuw werden bijvoorbeeld de begrippen ‘begaafd’ en ‘hoogbegaafd’ in Nederland geïntroduceerd (zie Mönks & Span, 1985). Tegenwoordig hebben beleidsmakers het dikwijls over ‘excellente leerlingen’ en het stimuleren van ‘toptalent’ (OCW, 2013). Het valt op dat in de gehanteerde terminologie accentverschillen waarneembaar zijn. Soms ligt de nadruk op wat in potentie aanwezig is, soms op wat de leerling daadwerkelijk realiseert. De huidige interesse vanuit OCW lijkt vooral ingegeven door het feit dat de Nederlandse leerling het weliswaar goed doet in de onderste prestatiegebieden (bijna alle leerlingen behalen het lage en gemiddelde niveau in de TIMMS- en PIRLS-onderzoeken), maar dat in vergelijking met toptanden slechts een relatief klein percentage functioneert op het niveau dat wordt omschreven als ‘advanced’. In de kaderbrief van maart 2013 geeft de staatssecretaris van OCW aan dit probleem te willen aanpakken door zich te richten op de in potentie 20% best presterende leerlingen in het funderend onderwijs (OCW, 2013, p. 2). Overigens laat de staatssecretaris in hetzelfde schrijven ook de term ‘hoogbegaafden’ vallen die volgens hem deel uitmaken van deze groep.

We kunnen de omschrijving van onze doelgroep benaderen vanuit een definitie van begaafdheid. Dat is van belang omdat veel literatuur over het onderwijs aan hoog presterende leerlingen uit deze hoek komt. Probleem is dat het begrip begaafdheid verre van eenduidig is omschreven. Hany (1987) vond meer dan honderd definities en categoriseerde ze (zie ook Van Boxtel, Sanders & Venmans, 1990). De zogeheten *capaciteitsmodellen* gaan uit van een bepaald minimaal intelligentiequotiënt ( $> 120$  of zelfs  $> 130$ ) en lijken daarmee een overwaardering van (intelligentie-)testscores als stabiele eigenschappen te impliceren. In de zogeheten *componentenmodellen* worden de cognitieve vermogens van leerlingen uiteengelegd in deelvaardigheden en -processen (zie Sternberg, 1985). Deze modellen zijn voor de onderwijspraktijk (nog) weinig vruchtbaar gebleken. De zogeheten *prestatiegerichte modellen* hebben duidelijk meer weerklank gevonden. In tegenstelling tot de capaciteitsmodellen gaan deze modellen niet uit van potentieel, maar van de ‘begaafde’ of ‘excellente’ prestatie zelf. Er wordt aangege-



Figuur 1 **Renzulli's 'three ring conception of giftedness' model**



Figuur 2 **Mönks' 'triadische model van begaafdheid**

ven welke constellaties van factoren voorwaardelijk zijn om tot dergelijke prestaties te komen. Zo geeft Renzulli (1978) in zijn 'Three ring conception of giftedness' (Figuur 1) aan dat een bovengemiddelde intelligentie, creativiteit en taakgerichte motivatie essentiële ingrediënten zijn voor een excellente prestatie.

Ook *socioculturele modellen* hebben ingang gevonden. Deze modellen zijn in feite een uitbreiding van de eerdergenoemde prestatiegerichte modellen. Er wordt niet alleen verwezen naar kenmerken binnen de persoon, maar ook naar factoren en omstandigheden in de omgeving. Zo refereert Mönks (zie Mönks & Van Boxtel, 1985) in zijn zogeheten 'triadische model' (Figuur 2) aan de sociale context van het gezin, de *peer*-groep en de school. Als we het gezin en de

leeftijdsgenoten buiten beschouwing laten, moet ook het onderwijs aan voorwaarden voldoen, bijvoorbeeld boeiend en uitdagend zijn, wil een begaafde leerling tot excellente prestaties komen. Daarmee zijn we terug bij de kaderbrief van de staatssecretaris. We kunnen ons vinden in zijn voorstel voor een brede aanpak die focust op de 20% best presterende leerlingen. Deze groep omvat namelijk met hoge waarschijnlijkheid óók de leerlingen met een hoge tot zeer hoge intelligentie en doet bovendien recht aan de begaafdheidsdefinities die minder hoge eisen stellen aan het intelligentiequotiënt. Daarnaast is de aanpak heel praktisch. De 20% die de staatssecretaris noemt, komt exact overeen met de omschrijving van niveau I in de indeling in niveaugroepen I tot en met V die gehanteerd wordt in het Cito-LVS.

### 3 Onderwijsarrangementen

Volgens het 'triadische model' van Mönks kan een begaafde leerling alleen tot excellente prestaties komen als het onderwijs aansluit bij zijn behoeften en kenmerken. Er zijn verschillende manieren om het onderwijs af te stemmen op de 20% best presterende leerlingen. We zetten de belangrijkste onderwijsarrangementen hieronder kort uiteen.

#### *Versnelling*

In dit onderwijsarrangement krijgt de leerling de gelegenheid zich in een hoger tempo dan zijn medeleerlingen door de leerstof heen te werken. Wanneer de leerling op zo goed als alle leerstofgebieden zoveel verder is dat het reguliere lesprogramma niet meer van toepassing is, kan de leerling één- of zelfs meermalen een klas overslaan. Ook een vervroegde overgang naar het voortgezet onderwijs, en uiteindelijk zelfs een vroegtijdige instroom in het hoger onderwijs, behoort tot de mogelijkheden. Hoewel versnelling een relatief eenvoudige oplossing is, is het een nadeel dat de leerling zijn eigen klas en klasgenoten moet verlaten. Niet elke leerling zal het daar even gemakkelijk mee hebben. Veel ouders en leerkrachten vrezen sociaal-emotionele problemen. Toch toont internationaal onderzoek aan dat deze problemen zich maar zelden voordoen. Hoogeveen (2008) concludeert in

haar proefschrift bijvoorbeeld dat *versnelde* hoogbegaafde leerlingen niet slechter functioneren dan *niet versnelde* hoogbegaafde leerlingen. Zij constateert echter wel dat leerlingen die op de basisschool een jaar of meer hebben overgeslagen in de eerste leerjaren van het voortgezet onderwijs in mindere mate sociaal geaccepteerd worden. Dat geldt vooral voor jongens, omdat zij gemiddeld gezien minder snel in de puberteit komen.

### ***Verdieping en verrijking***

In dit onderwijsarrangement krijgt de leerling de gelegenheid om zich zelfstandig of in kleine groepen te wijden aan extra activiteiten binnen de reguliere leerstofgebieden (verdieping) of aan activiteiten op geheel andere gebieden die zijn interesse hebben (verrijking). De leerling voert de verdiepings- en/of verrijkingsactiviteiten in principe uit binnen de context van zijn eigen klas. Een bekend verdiepings- en verrijkingsstelsel voor hoog presterende leerlingen is het zogeheten 'Draaideurmodel' (of 'Revolving Door Identification Model') dat in de jaren tachtig door Renzulli, Reis en Smith (1981) in de Verenigde Staten werd geïntroduceerd. In dit systeem wordt niet alleen naar intelligentiescores gekeken, maar ook naar andere factoren en informatiebronnen. Voor elke leerling wordt eerst vastgesteld hoe deze de reguliere leerstof snel, effectief, en zonder zich te hoeven vervelen, tot zich kan nemen. Vervolgens krijgt de leerling verdiepings- of verrijkingsactiviteiten toegewezen. Hij verlaat dus als het ware via een draaideur tijdelijk de reguliere leerstof. Ten slotte brengt de leerling verslag uit voor een publiek, hetzij in de vorm van een paper, hetzij in de vorm van een presentatie, optreden of tentoonstelling.

### ***Speciale scholen en klassen***

In dit onderwijsarrangement worden hoog presterende leerlingen als groep bij elkaar geplaatst. Doorgaans ligt daaraan een selectiebeleid ten grondslag, waarbij de leerling aan bepaalde voorwaarden moet voldoen. Soms worden leerlingen in een zogeheten 'plus-klas' geplaatst. Ook zijn er speciale scholen voor hoog presterende leerlingen zoals Leonardo-scholen. Dit schooltype is in 2007 geïntroduceerd. De selectiecriteria voor dit schooltype lijken samen te vallen met de factoren van Renzulli's 'Three ring conception of giftedness', gecombineerd met een zeer hoog intelligentiequotiënt ( $> 130$ ). Het onderwijs wordt zoveel mogelijk afgestemd op de interesses en mogelijkheden van individuele leerlingen. Er wordt aandacht besteed aan vakken als filosofie, grammatica en wiskunde. Ook wordt er in projecten veelvuldig gebruikgemaakt van nieuwe media.

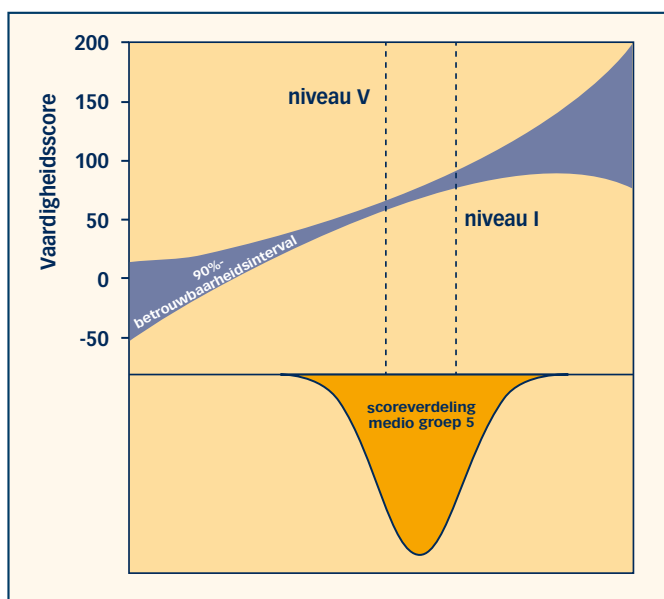
## **4 Aandachtspunten bij het toetsen van hoog presterende leerlingen**

Door het aanbieden van passende onderwijsarrangementen krijgen begaafde leerlingen optimaal de kans om zich te ontwikkelen. Om ervoor te zorgen dat de onderwijsarrangementen aansluiten bij de mogelijkheden van de leerling is het belangrijk om zicht te hebben op de vaardigheden van de leerling. De LVS-toetsen zijn hierbij een hulpmiddel. Uit de LVS-toetsen komt een vaardigheidsscore. Het is belangrijk om bij de interpretatie van vaardigheidsscores rekening te houden met de betrouwbaarheid van het resultaat, zeker als de vaardigheidsscore heel hoog (of laag) is. We lichten dit hieronder toe aan de hand van een beschrijving van

de opbouw van een toets en de begrippen 'lokale betrouwbaarheid', 'betrouwbaarheidsinterval' en 'plafondeffect'. Ook de validiteit vormt een aandachtspunt. Vanuit scholen komen signalen dat leerlingen met hoge vaardigheidsscores bij sommige opgaven soms meer fouten maken dan verwacht, bijvoorbeeld doordat ze de opgave anders interpreteren. We gaan in op deze signalen op basis van een analyse van data uit de meest recente 'Periodieke Peiling van het Onderwijsniveau' (PPON).

### Betrouwbaarheid

Niet elk toetsresultaat is even betrouwbaar. Als een leerling alle opgaven in een toets correct beantwoordt, weten we dat de leerling goed is, maar niet precies hoe goed. Als een leerling een aantal opgaven fout maakt en een aantal goed, weten we precies wat de leerling wel kan en wat niet. In het tweede geval weten we



Figuur 3 **Betrouwbaarheid van toetsscores**

meer over de vaardigheid van de leerling dan in het eerste geval. Het toetsresultaat zal dus ook betrouwbaarder zijn. Aangezien de betrouwbaarheid verschilt per vaardigheidsscore, wordt er ook wel gesproken van 'lokale betrouwbaarheid'. Figuur 3 geeft ter illustratie weer hoe betrouwbaar de LVS-toets Rekenen-Wiskunde voor medio groep 5 meet. Het blauwe vlak geeft voor verschillende vaardigheidsscores het 90%-betrouwbaarheidsinterval weer. Het interval laat zien welke vaardigheidsscores een leerling naar verwachting behaalt als de LVS-toets herhaald afgenomen zou worden en er geen sprake zou zijn van onderwijs- en herinneringseffecten. Het oranje vlak laat zien welke scores leerlingen in de praktijk halverwege groep 5 behalen op deze toets.

We zien dat de vaardigheidsscores die net iets onder het gemiddelde - links van de piek - liggen het betrouwbaarst zijn. Het betrouwbaarheidsinterval is daar immers het kleinst. Dit komt doordat de toets relatief veel opgaven bevat die qua moeilijkheidsgraad aansluiten bij het prestatieniveau van deze leerlingen. In Figuur 3 is ook duidelijk te zien dat bij leerlingen met extreem lage of hoge scores, dus aan de uiteinden van de scoreverdeling, het betrouwbaarheidsinterval het grootst is. Deze scores zijn dus minder betrouwbaar. Dit is vooral het geval bij de echt extreme scores (bijvoorbeeld bij 0 of 1 fout). Er is dan sprake van een 'plafondeffect'. De leerling is wellicht nog vaardiger, maar de LVS-toets Rekenen-Wiskunde voor medio groep 5 bevat geen opgaven meer om dit vast te stellen. Er wordt door Cito bewust voor gekozen om geen extreem moeilijke (of gemakkelijke) opgaven in de LVS-toetsen op te nemen. We willen enerzijds betrouwbaar meten bij zoveel mogelijk leerlingen én er anderzijds voor zorgen dat zoveel mogelijk leerlingen een succeservaring opdoen. Het is belangrijk om hier rekening mee te houden bij de interpretatie van toetsresultaten.

### Validiteit

Hoog scorende leerlingen interpreteren opgaven mogelijk anders dan hun klasgenoten. In dat geval zouden hoog scorende leerlingen beter of juist minder goed kunnen



presteren dan we mogen verwachten op basis van toetsgegevens die in het Nederlandse onderwijs verzameld worden. Cito heeft hier een verkennend onderzoek naar gedaan. Op basis van de informatie uit de PPO-peilingen Rekenen-Wiskunde in groep 5 en 8 zijn de prestaties van de 10% leerlingen met de hoogste vaardigheidsscores per opgave vergeleken met de prestaties van de overige leerlingen. Het onderzoek liet zien dat hoog scorende leerlingen de meeste opgaven (96%) consistent iets beter maken dan de overige leerlingen. Dit is wat we mogen verwachten. Tegelijkertijd liet het onderzoek zien dat de 10% leerlingen met de hoogste vaardigheidsscores enkele opgaven significant beter maakt dan verwacht (3.5%) en enkele andere opgaven significant slechter dan verwacht (0.5%). Dus waar we een consistent hoger prestatieniveau zouden verwachten over alle opgaven bleek dit bij een zeer beperkt aantal opgaven (4%) niet het geval te zijn. Er kon geen patroon ontdekt worden in het type opgaven dat de hoogst scorende leerlingen significant beter of slechter maakte dan verwacht. De analyses geven geen aanleiding om te veronderstellen dat de LVS-toetsen voor Rekenen-Wiskunde bij hoog presterende leerlingen een andere vaardigheid meten dan bij minder hoog presterende leerlingen.

We kunnen de hoog presterende leerlingen een LVS-toets van een hoger niveau aanbieden en het resultaat via de vaardigheidsscore toch vergelijken met de resultaten van de leerlingen die de LVS-toets maken die hoort bij het afnamemoment. Het 'toetsen op maat' heeft hierbij in veel gevallen de voorkeur boven het principe 'doortoetsen'.

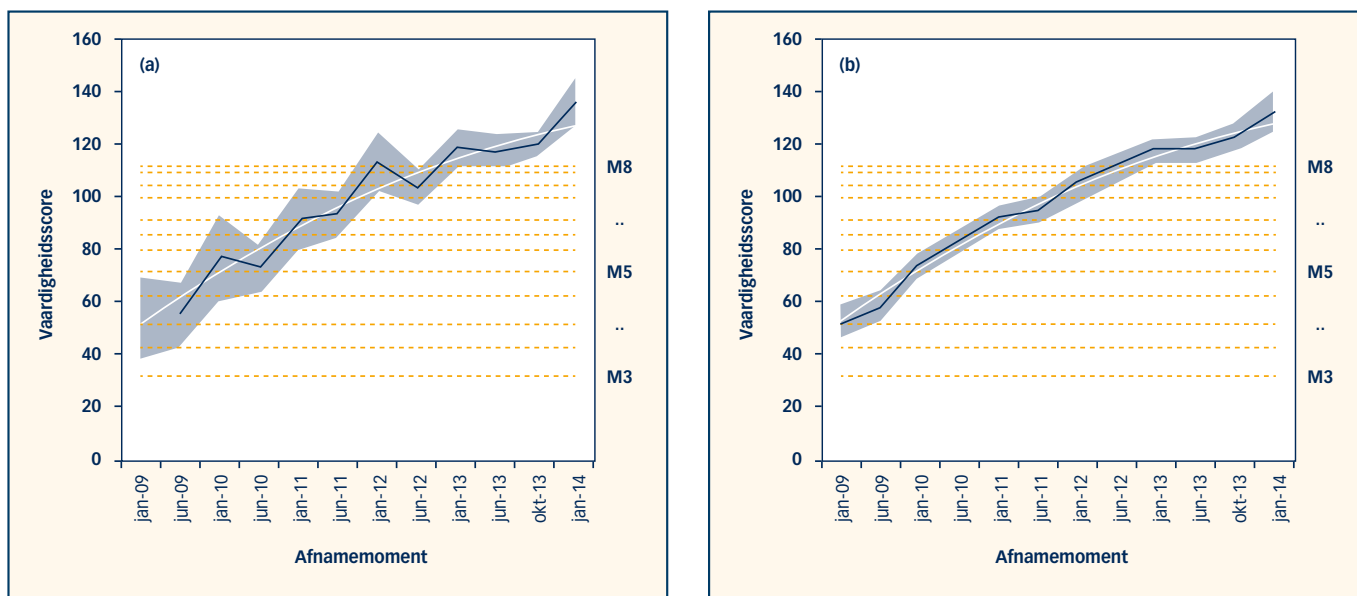
## 5 Betrouwbaarder meten

We hebben gezien dat de scores van leerlingen die bijna alle toetsopgaven goed (of fout) maken wat minder betrouwbaar zijn. Er zijn mogelijkheden om dit probleem met het Cito-LVS op te lossen. Alle LVS-toetsen van een bepaald leerstofgebied liggen namelijk op één onderliggende meetschaal. Deze meetschaal maakt het mogelijk om de prestaties van leerlingen die verschillende versies van toetsen maken, vergelijkbaar te maken. We kunnen, met andere woorden, de hoog presterende leerlingen een LVS-toets van een hoger niveau aanbieden en het resultaat via de vaardigheidsscore toch vergelijken met de resultaten van de leerlingen die de LVS-toets maken die hoort bij het afnamemoment. Het 'toetsen op maat' heeft hierbij in veel gevallen de voorkeur boven het principe 'doortoetsen'.

Bij **doortoetsen** maakt de leerling op één afnamemoment meerdere toetsen. Wanneer blijkt dat de afgenomen toets onvoldoende bij het prestatieniveau van de leerling past, bijvoorbeeld doordat de leerling (bijna) alles goed heeft gemaakt, biedt de leerkracht op hetzelfde moment nog een toets van een ander (hoger) niveau aan. Op deze manier kan de betrouwbaarheid van de meting vergroot worden, omdat het niet aannemelijk is dat de leerling bij een moeilijker toets opnieuw bijna alle opgaven correct beantwoordt. Hoewel de betrouwbaarheid van de meting ermee vergroot kan worden, is doortoetsen in de praktijk in veel gevallen niet echt nodig. Wanneer uit de toetsafname blijkt dat er sprake is van een forse voorsprong heeft de leerkracht in combinatie met de methodegebonden toetsen immers voldoende handvatten om in de klas aan de slag te gaan. De score uit de te gemakkelijke of te moeilijke toets is vanwege de geringere betrouwbaarheid minder geschikt om de vaardigheid van leerlingen over de tijd te volgen.

Bij **toetsen op maat** maakt de leerling pas bij een volgend afnamemoment een toets die aansluit bij zijn prestatieniveau. Vanaf dat moment is het niet alleen mogelijk om vast te stellen welk onderwijs bij de leerling past, maar kunnen ook ontwikkelingen in het prestatieniveau van de leerling (stabiel) over de tijd gevolgd worden. Toepassing van het principe ‘toetsen op maat’ is alleen wenselijk wanneer de leerling een forse voorsprong heeft op zijn leeftijdsgenoten. Voor verreweg de meeste leerlingen, ook voor de leerlingen in niveaugroep I, meet de toets die hoort bij het afnamemoment namelijk betrouwbaar genoeg. Dit blijkt onder andere uit Figuur 3 waarin we kunnen zien dat bij een groot deel van de leerlingen in niveaugroep I het betrouwbaarheidsinterval niet veel groter is dan het betrouwbaarheidsinterval van de leerlingen in de andere niveaugroepen. Over het algemeen is het bij het Cito-LVS zo dat een leerling met een voorsprong van minimaal 1 tot 1½ jaar op zijn leeftijdsgenoten beter getoetst kan worden met een toets van een hoger niveau.

In Figuur 4 is te zien welk effect de principes ‘doortoetsen’ en ‘toetsen op maat’ kunnen hebben op de toetsresultaten die we in de praktijk vinden. In deze figuur staan de toetsresultaten van een leerling met een forse voorsprong op zijn leeftijdsgenoten. Op het eerste afnamemoment in januari 2009 (M3) heeft de leerling een voorsprong van ongeveer één jaar. De toetsscore komt op dat moment immers overeen met de toetsscore die leerlingen halverwege groep 4 (M4) gemiddeld behalen. Deze voorsprong blijft in de schooljaren erna bestaan. We zien dat aan de witte lijn die de veronderstelde werkelijke vaardigheidsgroei van de leerling weergeeft. Deze lijn is normaal gesproken niet bekend. De donkerblauwe lijn geeft de vaardigheidsgroei weer die we in de praktijk kunnen observeren bij een leerling die de veronderstelde ontwikkeling doormaakt. Het lichtblauwe vlak is het bijbehorende 90%-betrouwbaarheidsinterval. Figuur 4a geeft de (fictieve) situatie weer waarbij er niet op maat is getoetst. In Figuur 4b is verondersteld dat er wel op maat is getoetst.



Figuur 4a en 4b **Vaardigheidsgroei van een leerling bij niet op maat (a) en wel op maat (b) toetsen**

Vanwege de voorsprong heeft de leerling in het eerste scenario steeds een te gemakkelijke toets gemaakt. De vaardigheidsscores zullen daardoor wat minder betrouwbaar zijn. Dat is in Figuur 4a zichtbaar door het grote betrouwbaarheids-

interval. Ook observeren we een zeer grillige ontwikkeling over de tijd. Soms valt de geobserveerde score hoger uit dan de werkelijke vaardigheid en wordt de vaardigheid van de leerling overschat. Dit is bijvoorbeeld het geval op afname-moment medio groep 6, in januari 2012. De toets geeft aan dat de leerling boven het niveau van medio groep 8 (M8) presteert, terwijl de werkelijke vaardigheid op het niveau van medio groep 7 (M7) ligt. Bij het volgende afname-moment, in juni 2012, is er daarentegen sprake van een onderschatting van de vaardigheid. De geobserveerde score komt overeen met het gemiddelde niveau van medio groep 7 (M7), terwijl de werkelijke vaardigheid op dat moment dicht bij het niveau van begin groep 8 (B8) ligt. Doordat de toetsen niet aansloten bij het niveau van de leerling werd de vaardigheid het ene moment dus (sterk) overschat en het volgende moment (sterk) onderschat. Daardoor lijkt het voor de leerkracht alsof de leerling in de tweede helft van 2012 flink achteruit is gegaan, terwijl dat in werkelijkheid niet zo is. Dit probleem ontstaat doordat er op basis van de te makkelijke toets slechts een grove (vrij onbetrouwbare) schatting van de vaardigheid gemaakt kon worden.

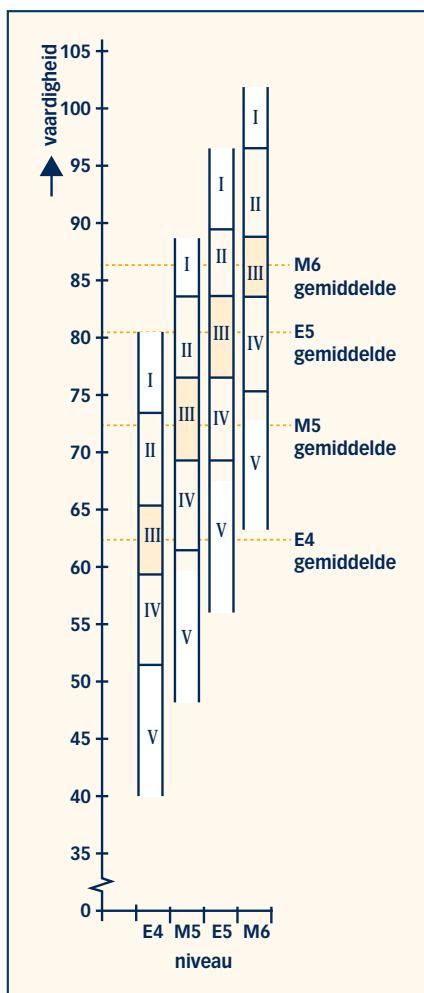
In het tweede scenario (Figuur 4b) is hiervan geen sprake meer. De toetsen sloten in dit scenario aan bij het prestatieniveau van de leerling. We konden betrouwbare schattingen maken van de vaardigheid, en de leerkracht observeert daardoor een tamelijk stabiele en betekenisvolle ontwikkeling over de tijd. Een groot voordeel van het afnemen van toetsen op een passend niveau is dus dat de ontwikkeling van leerlingen nauwkeuriger en betekenisvoller te volgen is. Een ander voordeel is dat de toetsafname ook meer analysemogelijkheden biedt. De toetsafname laat immers niet alleen zien of de leerling alles kan wat hem gevraagd is, maar laat ook zien of er nog specifieke hiaten zijn die om aandacht vragen.

Voor verreweg de meeste leerlingen is het dan ook het beste om de toets af te nemen die specifiek voor een bepaald afname-moment ontwikkeld is. In die gevallen waar de leerling geen enkele, of slechts één, opgave fout maakt, kan het voorleggen van een toets van een hoger niveau wenselijk zijn. Het is dan wel van belang de leerling te vertellen dat hij opgaven kan verwachten met stof waarmee in de les nog niet geoefend is en dat de toets alleen wordt afgenomen om te kijken hoe ver zijn vaardigheid reikt

## 6 Kiezen van een toets op niveau

Het is wenselijk dat hoog presterende leerlingen zoveel mogelijk op een passend niveau getoetst worden. In de praktijk kan het echter lastig zijn om een geschikte toets te kiezen. Wanneer een leerling nauwelijks fouten in een toets heeft gemaakt, zal de toets die er direct op volgt immers ook aan de gemakkelijke kant zijn. Wanneer een toets van een aanzienlijk hoger niveau wordt voorgelegd, zal de inhoud niet goed aansluiten bij het onderwijs dat de leerling heeft gehad. Het zogeheten 'alternatief leerlingrapport' van het Cito-LVS kan behulpzaam zijn bij het maken van een keuze. In Figuur 5 wordt ter illustratie een deel van het alternatief leerlingrapport weergegeven. We zien dat wanneer een leerling bij de toets voor eind groep 4 (E4) een vaardigheidsscore behaalt die hoog in niveaugroep I valt (bijvoorbeeld 82), ook bij de toets voor medio groep 5 (M5) naar verwachting veel opgaven goed zal maken. Ook deze toets had dus zonder problemen voorgelegd kunnen worden. De toets voor einde groep 5 (E5) sluit vermoedelijk nog be-





Figuur 5 **Toetsen op maat op basis van overlap tussen vaardigheidsscores van verschillende toetsen**

ter aan bij het prestatieniveau van de leerling. De vaardigheidsscore van de leerling valt dan namelijk in niveaugroep III. Het is het gemakkelijkst om bij het kiezen van een toets voor hoog presterende leerlingen in het alternatief leerlingrapport te kijken naar de stippelijnen met de gemiddelde vaardigheidsscores op verschillende afnamemomenten. We baseren de toetskeuze dan op het functioneringsniveau van de leerling. De vaardigheidsscore van 82 ligt bijvoorbeeld bijna op het 'E5 gemiddelde'. Dit geeft aan dat de score ongeveer overeenkomt met de score die leerlingen gemiddeld aan het einde van groep 5 behalen en dat de leerling dus op het niveau van eind groep 5 functioneert. Voor deze leerling zou de toets voor eind groep 5 daarom een goede keuze zijn. Voor andere vaardigheidsscores en afnamemomenten kan de keuze op hetzelfde principe gebaseerd worden.

Hoewel toetsen van een hoger niveau het toetsresultaat op dezelfde meetschaal uitdrukken als toetsen van een lager niveau, is de inhoud van de toetsen afgestemd op het onderwijs dat de leerlingen op een bepaald moment in de schoolloopbaan gehad hebben. In de toetsen Rekenen-Wiskunde voor groep 7 komen bijvoorbeeld breuken en procenten voor. In groep 5 hebben leerlingen deze stof nog niet gehad en daarom zit deze stof ook niet in de toetsen voor groep 5. Wanneer een leerling uit groep 5 dus een toets van groep 7 krijgt voorgelegd, kunnen daar onderwerpen in voorkomen die de leerling nog niet heeft gehad. Voor leerlingen kan het onplezierig zijn en soms oneerlijk aanvoelen om opgaven te moeten maken waarvan zij de stof nog niet hebben gehad. Voor verreweg de meeste leerlingen is het dan ook het beste om de toets af te nemen die specifiek voor een bepaald afnamemoment ontwikkeld is. In die gevallen waar de

leerling geen enkele, of slechts één, opgave fout maakt, kan het voorleggen van een toets van een hoger niveau wenselijk zijn. Het is dan wel van belang de leerling te vertellen dat hij opgaven kan verwachten met stof waarmee in de les nog niet geoefend is en dat de toets alleen wordt afgenomen om te kijken hoe ver zijn vaardigheid reikt. Bij het toetsen op een hoger niveau moet er bovendien op gelet worden dat voor de interpretatie van de score de niveaugroepen behorende bij het afnamemoment gebruikt worden. Wanneer een leerling medio groep 4 dus de toets van medio groep 5 maakt, moet voor het vaststellen van het vaardigheidsniveau van de leerling de normering van medio groep 4 gebruikt worden.

## 7 Conclusies

Bij leerlingen die geen of nauwelijks fouten in de toets maken is de toetsscore wat minder betrouwbaar. Ook weet een leerkracht dan niet wat de leerling net wel en niet kan, en wat de leerling dus nog moet leren. Dit is meestal niet zo'n probleem. Beslissingen over versnellen, verrijken en verdiepen zijn ook op basis van toetscores die iets minder betrouwbaar zijn goed te onderbouwen. Om de prestaties van de leerling goed te kunnen volgen over de tijd is het echter aan te bevelen om bij volgende afnamemomenten toetsen te kiezen die beter aansluiten bij het prestatie-

niveau van de leerling. De leerling krijgt op deze manier de kans om te laten zien wat hij allemaal al kan en wat hij nog niet kan. Zo krijgt de leerkracht optimale informatie voor het handelen in de klas, onder andere doordat een toets van een passend niveau meer analysemogelijkheden biedt en doordat de vaardigheidsgroei betrouwbaarder in beeld wordt gebracht. Dergelijke informatie is juist voor leerlingen voor wie het reguliere onderwijsprogramma niet toereikend is, dringend gewenst.

## GERAADPLEEGDE LITERATUUR

Zie hiervoor <http://tvodigitaal.nl> - september - 'Artikelen, Columns, Mededelingen'.

## OVER DE AUTEURS



**Judith Hollenberg** is ontwikkelingspsycholoog en werkzaam als toetsdeskundige bij Cito. Daar houdt zij zich bezig met de ontwikkeling van toetsen voor het primair en speciaal onderwijs. Om te zorgen dat scholen optimale informatie uit toetsen te kunnen halen verzorgt ze trainingen over het gebruik van toetsresultaten.

E-mail: [Judith.hollenberg@cito.nl](mailto:Judith.hollenberg@cito.nl)



**Herman van Boxel** is ontwikkelingspsycholoog en als UHD verbonden aan de afdeling Ontwikkelingspsychologie van de Universiteit Utrecht. Hij is in 1993 gepromoveerd in de sociale wetenschappen en heeft als belangstellingsgebieden onder andere cognitieve ontwikkeling, intelligentie, (hoog)begaafdheid en test- en toetsconstructie. Hij is parttime gedetacheerd bij Cito. Daar houdt hij zich vooral bezig met de ontwikkeling en wetenschappelijke verantwoording van instrumenten voor het onderwijs op het gebied van leerprestaties, intelligentie en sociaal-emotioneel functioneren.

E-mail: [Herman.vanboxtel@cito.nl](mailto:Herman.vanboxtel@cito.nl)



**Jos Keuning** is onderwijskundige en in 2008 gepromoveerd in de sociale wetenschappen op een onderzoek dat zich richtte op de lees- en spellingontwikkeling van kinderen gedurende de basisschoolperiode. Na zijn promotie is hij als methodoloog gaan werken bij het Psychometrisch Onderzoekcentrum van Cito. In die functie is hij onder meer betrokken geweest bij projecten die gericht waren op de ontwikkeling van tests voor specifieke doelgroepen. Zijn huidige onderzoek richt zich op de equivalering van toetsen, rapportages in leerlingvolgsystemen, groeimodellen, en toegevoegde waarde en accountability.

E-mail: [Jos.keuning@cito.nl](mailto:Jos.keuning@cito.nl)