

De SVT's en de vaardigheidsscore

Johan Schokker, Barbara van der Valk en Teije de Vos

De Schoolvaardigheidstoetsen (SVT's) zijn bedoeld om, onafhankelijk van de didactische onderwijsmethode die de school gebruikt, het niveau van een leerling op de schoolse vaardigheden vast te stellen door middel van een landelijke norm. Door een leerling regelmatig te toetsen kan zijn of haar voortgang op een bepaalde vaardigheid worden bepaald. Eén van de manieren om de ontwikkeling van een leerling te volgen is door middel van een vaardigheidsscore.

Drie van de vijf SVT's bieden de mogelijkheid om een vaardigheidsscore te berekenen: de *SVT Begrijpend Lezen*, de *SVT Spelling* en de *SVT Rekenen-Wiskunde*.

Eerst leggen we uit wat we onder een vaardigheidsscore verstaan en hoe deze online in het Boom testcentrum kan worden opgevraagd. Vervolgens gaan we inhoudelijk in op het meten van de voortgang van een leerling op de drie verschillende SVT's door middel van de vaardigheidsscore.

Wat is een vaardigheidsscore?

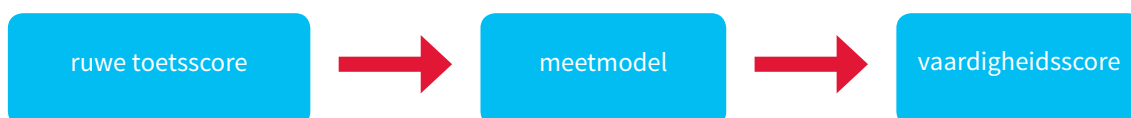
Een ruwe toetsscore is vaak gelijk aan het aantal goede antwoorden op een toets. Wanneer we steeds dezelfde toets afnemen, kunnen we de voortgang van de leerling afzetten tegen het aantal goede antwoorden; hoe groter het aantal goede antwoorden, des te groter de voortgang. Bij een leerlingvolgsysteem toetsen we de leerlingen regelmatig, maar meestal niet met dezelfde toets. We kunnen de ruwe toetsscores van de verschillende toetsen dan niet zomaar met elkaar vergelijken. Dit is niet mogelijk omdat de ene toets bijvoorbeeld meer opgaven heeft dan de andere toets, of omdat de opgaven van de ene toets moeilijker zijn dan die van de andere toets. Om de toetsscores op twee verschillende toetsen toch met elkaar te kunnen vergelijken, moeten we deze plaatsen op *dezelfde schaal*. Eén van de mogelijkheden is door gebruik te maken van een *vaardigheidsschaal*. Met behulp van een vaardigheidsschaal kunnen we de ruwe scores ('aantal goed') vertalen naar een *vaardigheidsscore*.

Wanneer we gebruik maken van *dezelfde* onderliggende vaardigheidsschaal kunnen we de vaardigheidsscores van diverse toetsen met elkaar vergelijken en kunnen we de groei van de leerling in beeld brengen.

Het meetmodel

In deze paragraaf wordt ingegaan op de statistische achtergrond van het berekenen van de vaardigheidsscore. Dit is vooral bedoeld als achtergrondkennis.

Om een vaardigheidsschaal te maken wordt gebruik gemaakt van een meetmodel. Het verband tussen de vaardigheid van een leerling en het antwoord op een opgave is probabilistisch van aard. Dat wil zeggen dat 'toevallige' factoren een rol spelen bij het tot stand komen van een toetsscore. Een leerling kan bijvoorbeeld een opgave in de toets verkeerd lezen waardoor hij de opgave fout maakt, terwijl hij die opgave gezien zijn vaardigheidsniveau eigenlijk goed had kunnen maken. Het meetmodel berekent de kans op een correct antwoord gegeven het beheersingsniveau van de leerling: hoe groter zijn vaardigheid, des te groter de kans op een correct antwoord. Wanneer men over een passend meetmodel beschikt, kan men de ruwe toetsscore vertalen naar een vaardigheidsscore.

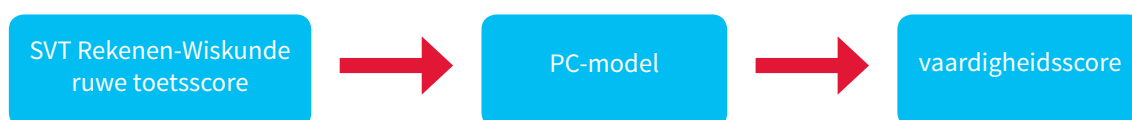


De vaardigheidsscore is een variabele die, omdat ze niet direct kan worden geobserveerd, moet worden geschat. Hoe meer opgaven een leerling heeft gemaakt, des te meer informatie beschikbaar is om het vaardigheidsniveau van een leerling te kunnen schatten. De vaardigheidsscore van de leerling kan dan ook preciezer worden geschat. Een relatief eenvoudig en in de testconstructie veel gebruikt meetmodel is het Rasch-model of het 1PL-model. Dit model is gebruikt voor de SVT Spelling.



Het Rasch-model wordt ook wel aangeduid als het 1PL-model, omdat elke opgave slechts door één parameter wordt gekarakteriseerd: de itemmoeilijkheid.

Voor de SVT Rekenen-Wiskunde is een soortgelijk model gehanteerd. De toetsen van de SVT Rekenen-Wiskunde (ook wel 'blokken' genoemd) zijn opgebouwd uit een reeks van opgaven. Elke opgave bestaat meestal uit een reeks subopgaven die een beroep doen op dezelfde rekenoperaties. Er bestaat dus een afhankelijkheid tussen de subopgaven. Om hiermee rekening te houden is een partial-credit model (PC-model) gehanteerd, waarbij de opgaven zijn opgevat als polytoom. (We spreken van 'polytoom' wanneer een opgave uit verschillende subopgaven bestaat. Daardoor kan men de opgave gedeeltelijk goed maken. De score wordt dan bepaald door het aantal goed gemaakte subopgaven.)



Het PC-model is een variant op het Rasch-model. Deze modellen hebben de prettige eigenschap dat de totaalscore op een toets direct kan worden omgezet in een vaardigheidsscore. Een vereiste daarbij is dat de opgaven van de toets hetzelfde discriminerend vermogen hebben (dat wil zeggen dat elke opgave in de toets leerlingen met een verschillend vaardigheidsniveau even goed kunnen onderscheiden). Daaraan wordt nooit helemaal voldaan, want dit is statistisch een zware eis.

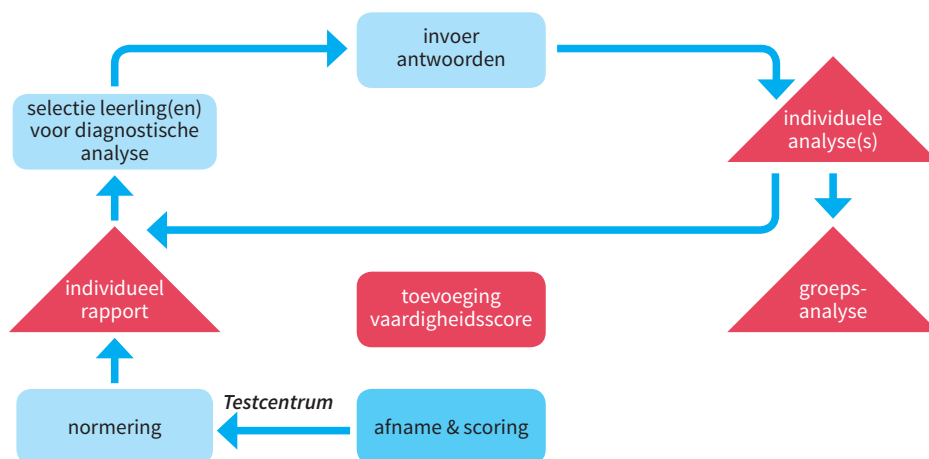
Bij de SVT Begrijpend Lezen was het discriminerend vermogen van de opgaven zo sterk verschillend, dat een ander meetmodel moest worden gebruikt. Daarom is voor de SVT Begrijpend Lezen het 2PL-model gebruikt. We spreken van het 2PL-model omdat elk item wordt gekarakteriseerd door twee parameters: een moeilijkheidsparameter en een discriminatieparameter. Omdat bij het 2PL-model het uitgangspunt is dat de opgaven van de toets verschillen in discriminerend vermogen, volstaat voor het berekenen van de vaardigheidsscore niet de ruwe toetsscore, maar is het antwoordenpatroon van de leerling nodig.



Dit heeft wel een belangrijke consequentie: *omdat bij de SVT Begrijpend Lezen de opgaven worden gewogen, kan eenzelfde totaalscore tot verschillende vaardigheidsscores leiden, terwijl bij de SVT Spelling en de SVT Rekenen-Wiskunde dezelfde totaalscore altijd tot dezelfde vaardigheidsscore leidt.*

Het opvragen van de vaardigheidsscore in het Boom testcentrum

Het gebruik van de verschillende meetmodellen voor de drie SVT's om een vaardigheidsscore te berekenen, heeft ook praktische gevolgen voor de wijze waarop de vaardigheidsscore in het Boom testcentrum kan worden opgevraagd. Bij de SVT Spelling en de SVT Rekenen-Wiskunde is alleen de ruwe toetsscore nodig voor het berekenen van de vaardigheidsscore en deze wordt dan ook tegelijk met de normscores in het Individueel (en Groeps-) Rapport getoond.



Figuur 1. Opvragen van de vaardigheidsscore van de SVT Begrijpend Lezen via de diagnostische module

Bij de SVT Begrijpend Lezen is voor de berekening van de vaardigheidsscore als gezegd ook het antwoordpatroon van de leerling nodig. Daarom moet de vaardigheidsscore worden verkregen *via* de diagnostische module. Het onderstaande schema maakt dit duidelijk.

Om een diagnostische analyse voor de SVT Begrijpend Lezen te kunnen uitvoeren moet altijd eerst de ruwe score van de leerling in het Boom testcentrum worden genormeerd. Na het normeren kan vervolgens de leerling worden geselecteerd, van wie men het toetsresultaat nader wil analyseren en voor wie men een vaardigheidsscore wil opvragen. Nadat de gegeven antwoorden zijn ingevoerd, verschijnt er in het overzicht van genormeerde leerlingen een vinkje. Daarmee wordt bevestigd dat de diagnostische analyse is uitgevoerd. Het individuele (of groeps-)rapport kan opnieuw worden opgevraagd. Daarop is dan naast de norm-scores ook de vaardigheidsscore zichtbaar.

Bij de vaardigheidsscore wordt ook het 90%-betrouwbaarheidsinterval van de vaardigheidsscore gerapporteerd. Met dit interval kan de meetprecisie van de vaardigheidsscore worden bepaald: wanneer men een leerling zeer vaak (met dezelfde toets) zou toetsen, zou de vaardigheidsscore in 90% van de gevallen tussen de grenzen van het interval komen te liggen. Men weet dus als gebruiker vrij zeker dat de vaardigheidsscore van de leerling tussen de grenzen van het gegeven betrouwbaarheidsinterval ligt.

De groei van de leerling in beeld: DLE of vaardigheidsscore?

Om de groei van de leerling in beeld te brengen, kennen de leerlingvolgsystemen in het (speciaal-)basis-onderwijs twee methoden: het DLE en de vaardigheidsscore. We kunnen niet zeggen dat de ene methode beter is dan de andere; beide methoden kennen hun voor- en nadelen. Het hangt van de voorkeuren van de gebruiker af, voor welke methode hij of zij kiest.

Het DLE heeft het onmiskenbare voordeel dat het direct te interpreteren is. Bijvoorbeeld: een leerling die in januari groep 6 (DL = 35) op een toets een score heeft behaald die gelijk is aan een DLE van 37 heeft de toets gemaakt op het niveau van een leerling van maart groep 6 (DL = 37). Men kan dan concluderen dat de leerling een voorsprong heeft van 2 onderwijsmaanden.

Een nadeel van het DLE is dat de schaal tot 60 loopt. Men kan dus moeilijker de groei van de heel goede leerlingen in groep 8 in beeld brengen.

Bij het DLE wordt de schaal in 60 onderwijsmaanden onderverdeeld. We spreken immers niet van een halve onderwijsmaand. Daardoor kunnen kleine afrondingsfouten ontstaan.

Een ander punt van aandacht is dat men moet vermijden de DLE los van de getoetste inhoud te interpreteren. Wanneer een leerling op een rekentoets voor groep 4 bijvoorbeeld een DLE van 21 (= september leerjaar 5) haalt, moet niet de conclusie zijn dat deze leerling het rekenen op het niveau van een groep 5-leerling beheerst, maar preciezer, dat deze leerling *de rekenstof van groep 4 op het niveau van een groep 5-leerling* beheerst. De rekenstof van groep 5 heeft men immers bij deze leerling nog niet getoetst.

Een vaardigheidsscore heeft het voordeel dat men de groei van de leerling op de vaardigheidsschaal vrij precies kan bepalen. Daarentegen is een vaardigheidsscore lastiger te interpreteren. Daarmee is de vaardigheidsscore ook niet vrij van valkuilen als hierboven beschreven. Ook voor de vaardigheidsscore geldt dat men altijd in het oog moet blijven houden op welke toets de betreffende score betrekking heeft.

Een voordeel van de vaardigheidsscore is dat men een betrouwbaarheidsinterval van de vaardigheidsscore kan berekenen. In de regel geldt dat bij extreme scores – leerlingen die bijna geen opgaven op de toets goed hebben gemaakt of leerlingen die bijna alle opgaven op de toets goed hebben gemaakt – de vaardigheidsscore moeilijker precies kan worden geschat en dat dus het betrouwbaarheidsinterval breder is. Voor welke methodiek men ook kiest, – DLE of vaardigheidsscore – het is verstandig om voor één methode te kiezen om de leerling te volgen. In principe leiden beide methoden tot dezelfde conclusies, maar soms zijn er subtiele verschillen die de vergelijkbaarheid van de groei uitgedrukt in DLE's en vaardigheidsscores compliceren. Bij de SVT Begrijpend Lezen worden de opgaven voor berekening van de vaardigheidsscore bijvoorbeeld gewogen, terwijl dit bij de bepaling van de DLE op SVT Begrijpend Lezen niet het geval is.

Interpretatie van de vaardigheidsscore

De vaardigheidsscore is geen normscore. De normscore, zoals de percentielscore of niveau A-E, niveau I-V of DLE, geeft aan welk niveau een leerling heeft ten opzichte van de groep. Zoals gezegd, de vaardigheidsscore is met name nuttig om de groei van de leerling in beeld te brengen. Daarbij dient men zich te realiseren dat de vaardigheidsschaal arbitrair is. Dit heeft als consequentie dat *alleen de vaardigheidsscores uit dezelfde toetsreeks met elkaar kunnen worden vergeleken*. Het is dus onzinnig om de vaardigheidsscore op een toets van de SVT Spelling te vergelijken met een vaardigheidsscore op een toets van de SVT Begrijpend Lezen of de SVT Rekenen-Wiskunde. Het is evenmin mogelijk de vaardigheidsscores op de SVT's te vergelijken met de toetsen van het Cito LOVS.

Dat gezegd hebbende, dient zich de vraag aan hoe we de groei van de vaardigheidsscore moeten interpreteren. Wanneer spreken we bij een leerling van gelijkmatige groei en wanneer kunnen we spreken van stagnerende groei?

Dit lijkt een simpele vraag, maar dat is zij zeker niet. In het ideale geval beschikken we over longitudinale gegevens van een representatief cohort van leerlingen. In dat geval zou het mogelijk zijn om op basis van modellen een groeicurve van een leerling op te stellen en daardoor – wanneer de leerling van de curve afwijkt – op verantwoorde wijze stagnatie of bovenmatige groei van de leerling vast te stellen. Op het ogenblik staat deze ontwikkeling nog in de kinderschoenen en worden leercurves of andersoortige modellen van de leerontwikkeling in het basisonderwijs nog niet breed toegepast¹. Bovendien is het de vraag of dergelijke modellen op een werkbare wijze in het basisonderwijs kunnen worden toegepast. Uit recent onderzoek van het Cito blijkt dat de ontwikkeling van de leerlingen in het basisonderwijs heel grillig kan verlopen en daardoor soms moeilijk voorspelbaar is². Dit geldt met name voor de zwakkere leerlingen, terwijl juist voor deze leerlingen, die de meeste stagnatie vertonen, het opstellen van een leercurve relevant zou kunnen zijn. De conclusie die uit het onderzoek kan worden getrokken, luidt daarom dat de groeiverwachtingen in termen van de vaardigheidsscore zich moet beperken tot een periode van een leerjaar.

Hoewel we niet over een cohort van longitudinale gegevens beschikken, beschikken we dankzij de normeringssteekproef wel over cross-sectionele gegevens. Op basis van deze gegevens kunnen we een uitspraak doen over *de verwachte gemiddelde groei in vaardigheidsscore*. Let wel, het gaat daarbij om een *gemiddelde* groei. Een individuele leerling kan aan het einde van de basisschool een gemiddeld eindniveau hebben, maar op weg daarnaar toe zowel perioden van stagnatie als perioden van bovengemiddelde groei hebben doorgemaakt, terwijl andere leerlingen een heel gelijkmatige ontwikkeling laten zien. Uiteraard speelt daarom bij de interpretatie van de toetsresultaten en de lijn die de opeenvolgende vaardigheidsscores laat zien, niet alleen de statistiek een rol, maar ook de inschatting van de leerkracht.

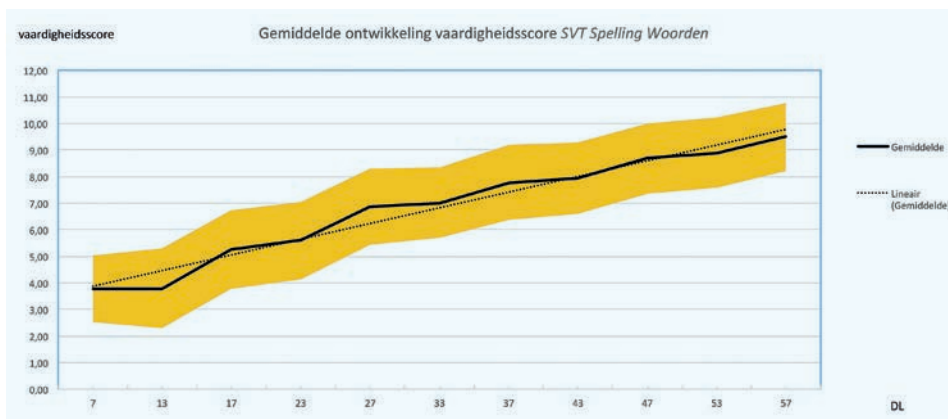
In de figuren 1 tot en met 4 worden de gemiddelde vaardigheidsscores op de SVT's weergegeven. Bij de keuze voor de subtoets (blokken) is uitgegaan van het gebruikelijke afnameschema, zoals beschreven in de betreffende handleidingen. We rapporteren de gemiddelden die zijn gebaseerd op de gewogen steekproef. Daarbij verbindt de zwarte lijn de gemiddelde vaardigheidsscores in de normeringssteekproef. Het gekleurde gebied geeft de spreiding van deze scores weer. Het spreidingsgebied is een standaarddeviatie boven en standaard-

1 Zie voor een uitwerking van mogelijke modellen de pilot naar leerwinst, die door OCW is uitgevoerd; Janssens, Rekers-Mombarg & Lacor, 2014.

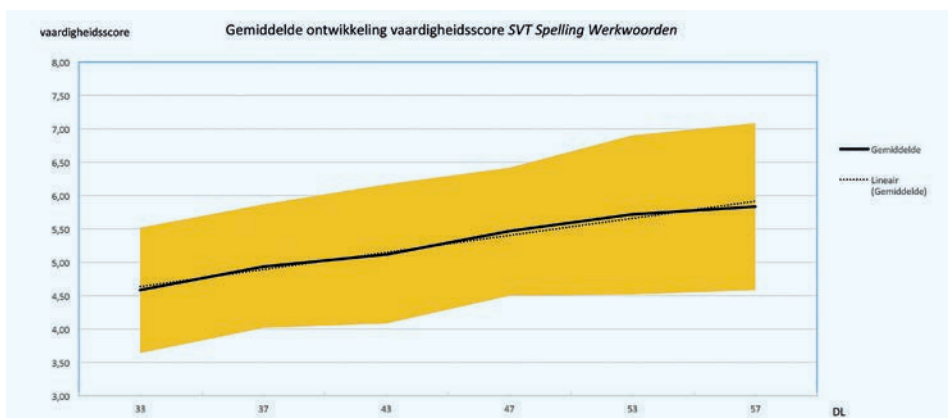
2 Zie Keuning & Visser, 2013.

deviatie onder het gemiddelde. Daarmee wordt zo'n twee derde van de mogelijke vaardigheidsscores in het gekleurde gebied weergegeven. Een vaardigheidsscore die buiten het gekleurde gebied valt, moet dus worden beschouwd als een relatief hoge dan wel een relatief lage score.

De stippellijn in de grafieken geeft de ontwikkelingstrend weer. Bij alle SVT's is er sprake van een duidelijk lineaire ontwikkeling van de vaardigheid. Dat wil zeggen dat leerlingen zich vanaf groep 3 met gelijke groei ontwikkelen. Bij een vaardigheid als technisch lezen is dat waarschijnlijk niet het geval; daar is de groei in groep 3 en groep 4 groot, maar vlakt deze af in latere leerjaren. Bij de vaardigheden die hier in het geding zijn, blijven leerlingen zich ontwikkelen. Let wel, het gaat hier om een *gemiddelde groei*. Zoals gezegd, zullen de meeste leerlingen zich naar alle waarschijnlijkheid met meer of minder grote schokken ontwikkelen. Wanneer we naar de zwarte lijnen kijken, zien we in de grafieken lichte schommelingen rond de stippellijn. In dat geval wijken de empirische gemiddelde vaardigheidsscores dus af van de lineaire trend (= de stippellijn in de grafiek).

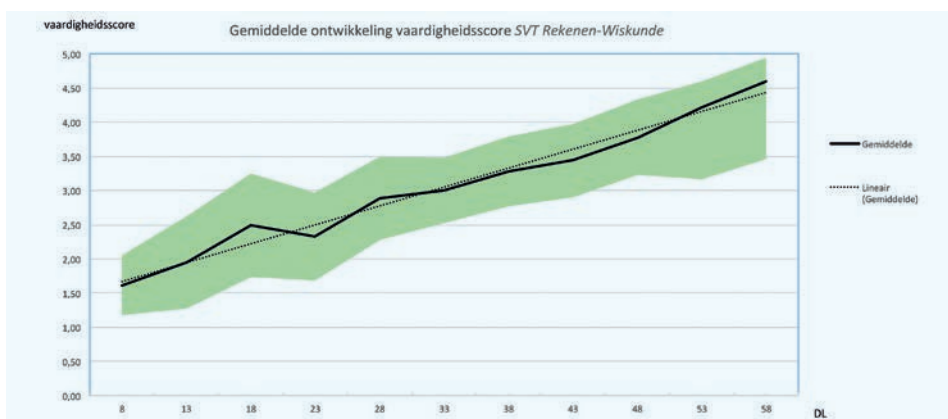


Figuur 1. Ontwikkeling van de spellingsvaardigheid gedurende het basisonderwijs – SVT Spelling Woorden



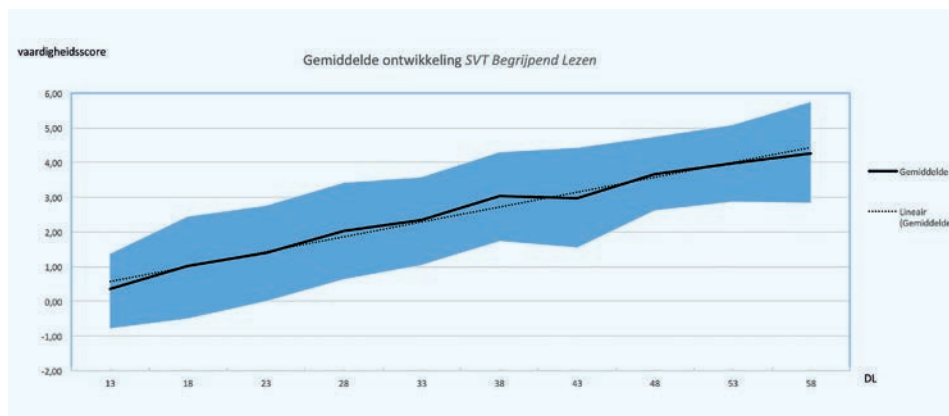
Figuur 2. Ontwikkeling van de spellingsvaardigheid gedurende het basisonderwijs – SVT Spelling Werkwoorden

Met name bij de SVT Spelling Woorden en de SVT Rekenen-Wiskunde treden afwijkingen op. Zo blijkt uit figuur 1 dat de vaardigheidsscore tussen DL = 7 en DL = 13 voor de SVT Spelling Woorden licht daalt. Uit figuur 1 kunnen we aflezen dat in de gewogen steekproef de gemiddelde score aan het begin van leerjaar 4 laag uitvalt in vergelijking met de verwachte gemiddelde score.



Figuur 3. Ontwikkeling van de rekenvaardigheid gedurende het basisonderwijs – SVT Rekenen-Wiskunde

Een soortgelijke daling van de vaardigheidsscore zien we bij SVT Rekenen-Wiskunde tussen DL = 18 en DL = 23. De meest waarschijnlijke verklaring is dat het hier om steekproeffluctuaties gaat³. De verschillen zijn ook niet zo groot dat we hieraan een belang moeten toekennen. Het lijkt daarom een rechtvaardige conclusie om bij de interpretatie van de vaardigheidsscore uit te gaan van de lineaire trendlijn.



Figuur 4. Ontwikkeling van de vaardigheid begrijpend lezen gedurende het basisonderwijs – SVT Begrijpend Lezen

Vanuit deze trendlijnen kunnen we een meer precieze uitspraak doen over de verwachte ontwikkeling. Daartoe is een eenvoudige regressieanalyse uitgevoerd op de empirische gemiddelden van de steekproef⁴. Deze regressievergelijkingen maken het mogelijk een uitspraak te doen over de verwachte groei in de vaardigheidsscore.

Tabel 1. Verwachte lineaire gemiddelde groei in vaardigheidsscore per onderwijsmaand⁵

Schoolvaardigheidstoets	Verwachte groei 1 onderwijsmaand	Verwachte groei 5 onderwijsmaanden	Verwachte groei 1 leerjaar
SVT Spelling Woorden	0,12	0,60	1,20
SVT Spelling Werkwoorden	0,05	0,25	0,50
SVT Rekenen-Wiskunde*	0,06	0,30	0,60
SVT Begrijpend Lezen	0,09	0,45	0,90

* Deze gemiddelde verwachte groei geldt zowel voor de afnamevorm van één blok als voor de afnamevorm van twee blokken.

In tabel 1 wordt de verwachte groei gegeven per onderwijsmaand, per 5 onderwijsmaanden en per leerjaar (= 10 onderwijsmaanden). Gegeven de meetprecisie van de vaardigheidsscore, en ook gegeven de praktijk waarin het leerlingvolgsysteem wordt toegepast, zal men in de regel eenmaal per jaar toetsen en zal men niet vaker dan halfjaarlijks toetsen. Er kunnen natuurlijk legitieme redenen zijn om hiervan af te wijken, bijvoorbeeld in het geval van een intensivering van het onderwijs in een bepaalde schoolse vaardigheid, bij een hele klas dan wel bij een groepje leerlingen, waardoor de verwachte groei in de vaardigheidsscore groter dan gemiddeld zal zijn. Maar de 'reguliere' groei in de vaardigheidsscore per onderwijsmaand zal men door middel van een toets niet kunnen vaststellen, omdat deze groei te gering is en de toetsen niet precies genoeg meten om zo'n relatief kleine groei betrouwbaar te kunnen vaststellen. Dat is overigens ook helemaal niet nodig; in het basisonderwijs volstaat een jaarlijkse of een halfjaarlijkse toetsing om een beeld te krijgen van de voortgang van de leerling.

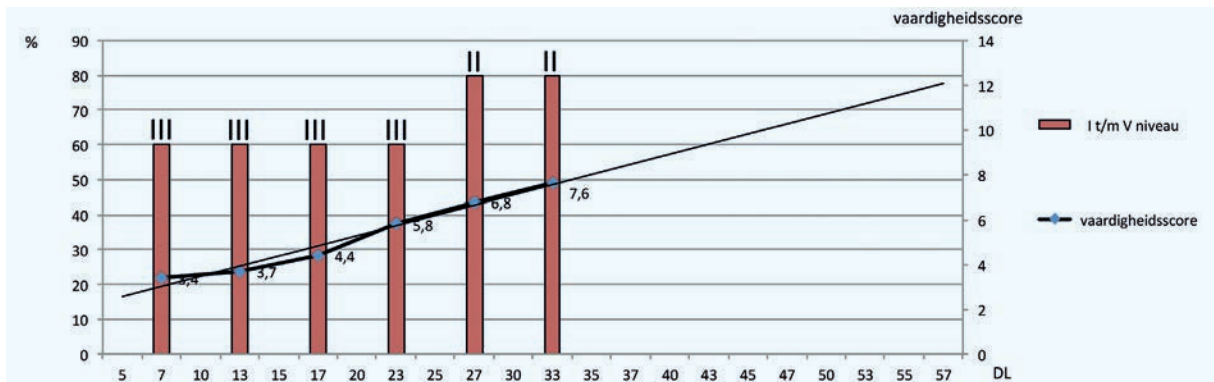
3 Hoewel bij de normering de representativiteit van de steekproef natuurlijk is gewaarborgd door de selectie van de scholen in de steekproef en de weging op de voornaamste kenmerkende variabelen van de leerlingenpopulatie (zie de betreffende handleiding voor een nauwkeurige beschrijving) zijn kleine fluctuaties in de steekproef nooit uit te sluiten. Een andere verklaring zou een meer structureel effect kunnen zijn. Wanneer leerlingen bijvoorbeeld in groep 4 in november getoetst worden op de spellingsvaardigheid zou een lange zomervakantie (en dus een gebrek aan oefening) bij sommige leerlingen tot een kleine en tijdelijke terugval in de spellingsvaardigheid kunnen leiden. Ten derde kan een onbedoeld effect van de toetsconstructie een rol spelen, bijvoorbeeld doordat blok 3 van de SVT Rekenen-Wiskunde een aantal lastige opgaven heeft, die demotiverend zouden kunnen werken waardoor de leerlingen begin groep 5 onder hun kunnen presteren. Het is natuurlijk ook niet uit te sluiten dat al deze effecten samen optreden; elk voor zich nagenoeg verwaarloosbaar, maar gezamenlijk net tot een zichtbaar effect leidend.

4 De R^2 -waarden, als maat voor beschrijving van de data door de opgestelde vergelijkingen, variëren van .96 tot .98.

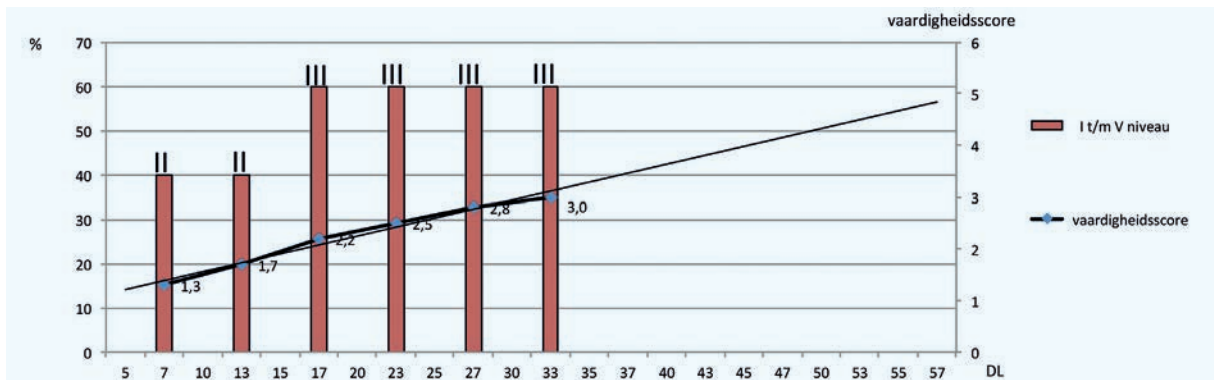
5 In het basisonderwijs gaan we uit van 10 onderwijsmaanden in een leerjaar. Daarbij worden de maanden juli en augustus niet meegerekend.

Voorbeeld van de interpretatie van de ontwikkeling bij leerling Kees

In de figuren 5 en 6 wordt een voorbeeld gegeven van de wijze waarop men grafisch de leergroei van een leerling in beeld kan brengen. Het gaat om de fictieve leerling Kees die vanaf halverwege groep 3 tweemaal per jaar wordt getoetst met de SVT Spelling Woorden en de SVT Rekenen-Wiskunde. Tot en met begin groep 6 zijn er gegevens verzameld. In de grafiek wordt steeds het verloop van de vaardigheidsscores gegeven. Daarnaast worden in dezelfde grafiek de normscores I-V niveau afgebeeld.



Figuur 5. Ontwikkeling van de spellingsvaardigheid van Kees uit groep 6 – SVT Spelling Woorden – niveau I - V en vaardigheidsscore



Figuur 6. Ontwikkeling van de rekenvaardigheid van Kees uit groep 6 – SVT Rekenen-Wiskunde – niveau I - V en vaardigheidsscore

Uit de normscores I-V blijkt dat Kees een redelijk stabiele leerling is. In spelling is hij aanvankelijk een gemiddelde leerling, maar in groep 6 lijkt hij zich te ontwikkelen tot een bovengemiddelde leerling. In groep 3 en begin groep 4 is Kees een zwakke leerling in het rekenen, maar in de latere leerjaren haalt hij zijn achterstand in.

Naast de normscores zijn ook de vaardigheidsscores in de grafiek uitgezet. Deze scores laten voor zowel de SVT Spelling als voor de SVT Rekenen-Wiskunde een lineaire ontwikkeling zien. Kees is steeds in het najaar en het voorjaar getoetst. Uit tabel 1 kunnen we aflezen dat we bij een 'normale' ontwikkeling in de rekenvaardigheid na vier onderwijsmaanden, een groei in de vaardigheidsscore verwachten van ($4 \times 0.06 = 0.24$) en na zes onderwijsmaanden een groei van ($6 \times 0.06 = 0.36$). Tussen november groep 4 (DL = 13) en maart groep 4 (DL = 17) laat Kees een bovengemiddelde groei zien in de rekenvaardigheid. Terwijl we een toename van 0.24 verwachten, is de werkelijke groei 0.5. Kees laat dus een dubbele groei in vaardigheidsscore zien. Vanaf dat moment bevindt Kees zich dan ook in de normgroep III en lijkt hij zich te hebben hersteld van zijn aanvankelijke zwakke start.

Tot slot

De leerlingvolgsystemen in het basisonderwijs kennen twee methoden om de groei van een leerling in beeld te brengen. Er kan gebruik worden gemaakt van een DLE-schaal of een vaardigheidsschaal. Deze schalen kunnen worden opgevat als een meetlat waarop men de toetsprestatie van de leerling kan plaatsen. Hoe hoger op de meetlat de toetsprestatie van de leerling kan worden geplaatst, des te groter is de beheersing van de vaardigheid. Door in het kader van een leerlingvolgsysteem herhaaldelijk te toetsen kan men zien waar de leerling zich op de schaal bevindt. Zo krijgt men zijn voortgang in beeld.

Voor welke methode men kiest, is afhankelijk van de voorkeuren van de gebruiker, maar het is aan te bevelen voor alle domeinen waarop men de leerlingen wil volgen, dezelfde methodiek te kiezen.

Het gebruik van een vaardigheidsscore biedt het voordeel van een betrouwbaarheidsinterval, maar de interpretatie is niet zo direct als bij het DLE. Bij de interpretatie van een vaardigheidsscore of een DLE dient men eveneens de normscore te betrekken (niveau A - E of niveau I - V).

Daarnaast wordt in dit paper voor elke SVT aangegeven welke gemiddelde groei men kan verwachten per onderwijsmaand.

Door de vaardigheidsscores van twee toetsafnames met elkaar vergeleken, kan men bekijken of de groei voldoet aan de verwachting. Wanneer er eventueel stagnatie optreedt, kunnen daarvoor allerlei redenen zijn.

Vaak is het onverstandig om op basis van een enkel tegenvallend toetsresultaat te concluderen dat er sprake is van stagnatie, zeker wanneer die stagnatie in de eerste leerjaren van het basisonderwijs plaatsvindt. Dan zijn er nog niet zoveel toetsresultaten beschikbaar van de leerling en heeft men op school vaak nog geen duidelijk beeld over wat een leerling kan. Bovendien is bekend dat sommige leerlingen in het basisonderwijs zich op heel grillige wijze kunnen ontwikkelen, en dat is met name in de eerste leerjaren het geval.

Men kan eventueel een leerling na een aantal maanden extra toetsen, om na te gaan of er daadwerkelijk sprake is van stagnatie in het leren. De SVT's bieden daarvoor ook de mogelijkheden. Mocht de stagnatie inderdaad duurzaam zijn, dan kan men vervolgens (o.a. met behulp van de diagnostische analyse) een didactisch plan opstellen om die achterstand terug te dringen en beter op de behoeften van de leerling in te spelen.

Literatuur

- Janssens, F., L. Rekers-Mombarg & E. Lacor (2014). *Leerwinst en toegevoegde waarde in het primair onderwijs: Eindrapportage*. Rotterdam: CED-Groep.
- Keuning, J., & J. Visser, (2013). De (on)nauwkeurigheid van een leerrendementsverwachting: Ervaringen met de LVS-toetsen Rekenen-Wiskunde. *Tijdschrift voor Orthopedagogiek*, 52, 164-176.

Auteurs

Johan Schokker is methodoloog bij Boom uitgevers Amsterdam en gespecialiseerd in toetsontwikkeling en -normering.

Barbara van der Valk is methodoloog en coördinator van normeringsonderzoeken bij Boom uitgevers Amsterdam.

Teije de Vos is oud-onderwijzer en orthopedagoog. Hij heeft jarenlange ervaring in het ontwikkelen van testen en toetsen, waaronder de verschillende schoolvaardigheidstoetsen (SVT Technisch Lezen, SVT Begrijpend Lezen, SVT Spelling, SVT Hoofdrekenen, SVT Rekenen-Wiskunde en Tempo Test Automatiseren).

© 2017 Johan Schokker, Barbara van der Valk, Teije de Vos en Boom test onderwijs. Dit whitepaper is vrij te gebruiken voor niet commerciële doeleinden. Voor commercieel gebruik dient u contact op te nemen met Boom test onderwijs, info@boomtestonderwijs.nl of (020) 524 45 14.