

De (on)nauwkeurigheid van een leerrendementsverwachting: Ervaringen met de LVS-toetsen Rekenen-Wiskunde

SAMENVATTING

Louwe (2013) maakt zich zorgen over de consequenties van automatisch tot stand gekomen ontwikkelingsperspectieven. Hij pleit voor een werkwijze op maat, waarin ruimte is voor individuele afwegingen en tussentijdse bijstellingen. In dit artikel wordt het pleidooi van Louwe ondersteund. Cito doet al jaren onderzoek naar de mogelijkheid om op basis van toetsresultaten betrouwbare leerrendementsverwachtingen op te stellen. De ontwikkeling van sommige leerlingen blijkt dermate grillig te verlopen dat het uitstroomniveau niet nauwkeurig voorspeld kan worden. Voor de kortere termijn kunnen op basis van toetsresultaten wel betekenisvolle leerrendementsverwachtingen worden geformuleerd. Deze leerrendementsverwachtingen kunnen de leerkracht houvast geven in de planningscyclus. De leerkracht kan het onderwijsaanbod op basis van de leerrendementsverwachting afstemmen op het niveau van de leerling, en kan de leervorderingen evalueren door de leerrendementsverwachting te vergelijken met daadwerkelijk behaalde toetsresultaten.

Enerzijds kan de beslissing om al dan niet een ontwikkelingsperspectief voor een leerling uit te werken, gebaseerd worden op een signalering van achterblijvende leerprestaties en aanvullend onderzoek naar verklarende beperkingen. Anderzijds kunnen scholen kiezen voor een meer ecologisch gefundeerde besluitvorming, zoals die ook in het speciaal basisonderwijs plaatsvindt

1 Introductie

Internationaal onderzoek laat zien dat er een sterk verband bestaat tussen de opbrengstgerichtheid van scholen en de leerresultaten van leerlingen (Fielding, Kerr & Rosier, 2007). Opbrengstgerichte scholen hebben hoge verwachtingen

van hun leerlingen, werken doelgericht om die verwachtingen te realiseren, maken veelvuldig gebruik van evaluatie-instrumenten en interpreteren evaluatiegegevens in het licht van onderwijsaanpassingen. Om de opbrengstgerichtheid van scholen voor het speciaal basisonderwijs te vergroten is op initiatief van de onderwijsinspectie het ontwikkelingsperspectief ontwikkeld. Het idee was dat het werken vanuit een ontwikkelingsperspectief zou leiden tot (a) een doelgerichte planning van het onderwijs, (b) beredeneerde keuzes in leerlijnen en onderwijsaanbod, (c) een betere afstemming tussen school, ouders en leerlingen, en (d) een betere overgang van het basisonderwijs naar het voortgezet onderwijs. In de praktijk blijkt het opstellen van een ontwikke-

lingsperspectief echter niet eenvoudig te zijn. Om scholen en leerkrachten meer houvast te geven, is daarom in 2009 een aanzet gegeven voor een gezamenlijk kader door middel van de publicatie: *‘Werken vanuit een ontwikkelingsperspectief in het speciaal basisonderwijs’* (Clijsen, Pieterse, Spaans & Visser, 2009). Vanaf 1 augustus 2012 staat het ontwikkelingsperspectief centraal in de waarderingskaders voor het speciaal onderwijs (Inspectie van het onderwijs, 2012). Ook van scholen voor regulier basisonderwijs verwacht de onderwijsinspectie dat zij een ontwikkelingsperspectief opstellen voor leerlingen die de minimumdoelen niet behalen.

Louwe (2013) stelt drie essentiële vragen bij de ontwikkelingsperspectieven die verplicht door scholen worden opgesteld. Ten eerste is onduidelijk op basis van welke procedures en argumenten scholen beslissen om een ontwikkelingsperspectief voor een leerling te formuleren. Ten tweede is het de vraag op basis van welke gegevens de school vaststelt dat het basis-uitstroomniveau niet haalbaar is voor een leerling. Ten derde is onduidelijk hoe nauwkeurig de inschattingen van scholen over de uitstroomniveaus van leerlingen zijn. Louwe (2013) schetst twee wegen die scholen kunnen volgen bij het vaststellen en gebruiken van ontwikkelingsperspectieven. Enerzijds kan de beslissing om al dan niet een ontwikkelingsperspectief voor een leerling uit te werken, gebaseerd worden op een signalering van achterblijvende leerprestaties en aanvullend onderzoek naar verklarende beperkingen. Anderzijds kunnen scholen kiezen voor een meer ecologisch gefundeerde besluitvorming, zoals die ook in het speciaal basisonderwijs plaatsvindt (Clijsen, Pieterse, Spaans & Visser, 2009). Er zijn verschillende kritische kanttekeningen te plaatsen bij de manier waarop het ontwikkelingsperspectief de laatste jaren vorm heeft gekregen in het onderwijs (cf. Louwe,

2013; Wieberdink & Kuster, 2011):

- 1 Het leerrendement, de intelligentie-ontwikkeling en het onderwijsaanbod beïnvloeden elkaar wederkerig. Dit brengt het risico met zich mee dat de beslissing om het leerstofaanbod voor een leerling te beperken, gebaseerd kan zijn op achterblijvende leerprestaties, die mogelijk het gevolg zijn van ondermaats onderwijs.
- 2 Er is weinig tot geen onderzoek gedaan naar de kwaliteit en nauwkeurigheid van de voorspelling bij leerlingen die doorstromen naar het praktijkonderwijs of het voortgezet speciaal onderwijs, terwijl juist voor deze leerlingen een ontwikkelingsperspectief wordt opgesteld. De validiteit van de schooladviezen die volgen uit de Eindtoets Basisonderwijs is wel onderzocht (Hakkenes & De Wijs, 2012; Stroucken, Tackenberg & Béguin, 2008). Hoewel het schooladvies het succes in het vervolgonderwijs goed kan voorspellen, gaat het bij de Eindtoets Basisonderwijs om een kortetermijnvoorspelling bij oudere leerlingen. Bij het vaststellen van het ontwikkelingsperspectief en het uitstroomniveau gaat het in de regel om jongere leerlingen en langere tijdsperiodes.
- 3 De instroomcompetenties die aangeven waarover een leerling na afloop van het basisonderwijs ten minste zou moeten beschikken om succesvol de verschillende vormen van voortgezet onderwijs te kunnen doorlopen, missen elke vorm van wetenschappelijke evidentie. De keuzes die scholen maken bij het opstellen van een ontwikkelingsperspectief en een uitstroomprofiel berusten daardoor slechts op *common sense*.
- 4 Het ontwikkelingsperspectief en het uitstroomprofiel worden vaak al op jonge leeftijd opgesteld. Als er te vroeg beslissingen worden genomen over het uitstroomprofiel en het onderwijsaanbod kan er een *self fulfilling prophecy* ontstaan. Ook het pedagogische en

didactische gedrag van de leerkracht kan (mede) gestuurd worden door de (toets)gegevens die hij op basis van eigen observaties en beoordelingen verzamelt. Het negatieve effect van impliciet of expliciet aanwezige lage verwachtingen is al in diverse onderzoeken naar schooleffectiviteit aangetoond (Benner & Mistry, 2007; Gill & Reynolds, 1999; Rist, 2000).

Louwe (2013) concludeert dat het vaststellen van een meerjarig toekomstperspectief en het nemen van beslissingen over het vervolg van het onderwijs niet eenvoudig is. Daarom zet hij vraagtekens bij geautomatiseerde procedures om het ontwikkelingsperspectief vorm te geven. Deze procedures houden volgens hem namelijk geen rekening met wederzijdse beïnvloeding van sociaal-emotioneel functioneren, leerwerkhouding en schoolvorderingen. Bovendien gaan de procedures uit van lineaire ontwikkelingslijnen en wordt er geen rekening gehouden met individuele verschillen in ontwikkelingssnelheid en -kwaliteit. Onderzoek dat Cito heeft uitgevoerd rond leerrendementsverwachtingen en uitstroomniveaus ondersteunen deze conclusie. Zelfs bij toepassing van een dynamische procedure die optimaal rekening houdt met ontwikkelingsverschillen bij individuele leerlingen blijkt het opstellen van een nauwkeurige leerrendementsverwachting niet altijd mogelijk te zijn. In de volgende paragraaf gaan we eerst in op de methodologische problemen die spelen bij het formuleren van leerrendementsverwachtingen. Vervolgens presenteren we een methode die zo goed mogelijk aan deze problemen tegemoetkomt. Ten slotte laten we aan de hand van enkele praktijkvoorbeelden uit het *Cito Volgstelsel primair en speciaal onderwijs* zien hoe nauwkeurig we het ontwikkelingsverloop en uitstroomniveau van individuele leerlingen kunnen voorspellen.

2 Vaststellen van een leerrendementsverwachting

Het volgen van leerresultaten en het voorspellen van groei is niet eenvoudig

2.1 Methodologische problemen

Om te kunnen werken vanuit een ontwikkelingsperspectief is het een voorwaarde dat scholen via een gestandaardiseerde procedure, op basis van meetbare gegevens, het verwachte leerrendement van een leerling kunnen vaststellen. Het volgen van leerresultaten en het voorspellen van groei is niet eenvoudig (cf. Kamphuis & Engelen, 1993). We noemen hier vier aandachtspunten. In de eerste plaats moeten we beschikken over een longitudinaal meetsysteem waarin de verschillende toetsen de beoogde onderliggende vaardigheid steeds op dezelfde manier, en in even sterke mate, reflecteren. De validiteit van een longitudinale meting wordt bedreigd indien er sprake is van multidimensionaliteit tussen groepen personen of items (cf. Keuning, 2008). Bij multidimensionaliteit tussen groepen personen worden de prestaties van leerlingen op de items in de toets mede beïnvloed door iets anders dan de vaardigheid die de toets beoogt te meten. Uit een studie van Kaplan en Walpole (2005) blijkt bijvoorbeeld dat leesvaardigheid in de lagere leerjaren van het basisonderwijs uit drie opeenvolgende ontwikkelingsfasen bestaat die kwalitatief van elkaar verschillen. De leesvaardigheid van leerlingen in groep 3 kan dus iets anders betekenen dan de leesvaardigheid van leerlingen in groep 4 of 5. Bij multidimensionaliteit tussen groepen items is er ook binnen één enkele groep sprake van twee of meer dimensies. Volgens Notenboom en Reitsma (2003) bestaat spellingvaardigheid in groep 3 bijvoorbeeld uit een fonologische en een lexicale component en ligt er pas vanaf groep 4 één enkele dominante factor ten grondslag aan spel-

lingvaardigheid. Bij beide vormen van multidimensionaliteit wordt het doen van voorspellingen over het niveau van leerlingen in latere ontwikkelingsfasen lastiger, zeker als de tijdsperiode waarop de voorspelling betrekking heeft, langer wordt.

Een tweede aandachtspunt bij het vaststellen van leerrendementsverwachtingen heeft betrekking op de betrouwbaarheid van meetresultaten voor individuele leerlingen (Bock, 1976; Kamphuis & Engelen, 1993). De betrouwbaarheid van het meetresultaat is niet bij elke toetsafname hetzelfde. Dat komt doordat de informatie die we verkrijgen over de vaardigheid van een leerling afhangt van de mate waarin de toets aansluit bij het vaardigheidsniveau (e.g., Lord, 1984; Hambleton, Swaminathan & Rogers, 1991; Wilson, 2005). Als een leerling alle items in de toets correct beantwoordt, dan weten we dat de leerling goed is, maar niet precies hoe goed. Als een leerling een aantal items fout maakt en een aantal goed, dan weten we precies wat de leerling wel kan en wat niet. In het tweede geval weten we dus meer over de vaardigheid van de leerling dan in het eerste geval en kunnen we de vaardigheid nauwkeuriger bepalen. Doordat de betrouwbaarheid van meetresultaten niet constant is over verschillende toetsafnames fluctueert de vaardigheid van leerlingen soms sterk in de tijd. Vooral bij zwakke en excellente leerlingen vertoont het groeipatroon in veel gevallen een zeer grillig verloop als gevolg van de onnauwkeurigheid van bepaalde metingen. Het is de vraag hoe we bij het vaststellen van een ontwikkelingsperspectief voor een leerling om moeten gaan met onbetrouwbare meetresultaten. Daarnaast is het de vraag hoe fluctuaties in groeipatronen van individuele leerlingen geïnterpreteerd moeten worden. Vanuit statistisch oogpunt is het bijvoorbeeld niet correct om een verandering tussen twee opeenvolgende metingen te interpreteren als een voor-

of achteruitgang in vaardigheid als de betrouwbaarheidsintervallen voor de metingen overlappen.

Een derde aandachtspunt heeft betrekking op de vraag welke informatie meegenomen moet worden bij het vaststellen van een leerrendementsverwachting. In de meest eenvoudige methodiek worden alleen de resultaten die een leerling behaalt binnen een bepaald vakgebied gebruikt om een toekomstperspectief te schetsen. Er kan echter sprake zijn van individuele verschillen in de manier waarop en de snelheid waarmee een vaardigheid tot ontwikkeling komt. In de literatuur wordt er bijvoorbeeld op gewezen dat de ontwikkeling van leerlingen met een hoog aanvangsniveau sneller verloopt dan de ontwikkeling van leerlingen met een lager aanvangsniveau (Francis et al., 1996; Merton, 1968; Stanovich, 1986). Uit onderzoek bij kinderen in het speciaal basisonderwijs blijkt dat factoren die betrekking hebben op de leerpotentie, de sociaal-emotionele ontwikkeling of de onderwijsleersituatie van invloed kunnen zijn op de ontwikkeling van leerlingen (Moelands, 2007). Als dergelijke factoren genegeerd worden bij het vaststellen van een leerrendementsverwachting bestaat het risico dat we de prestaties van een leerling systematisch onderschatten of juist overschatten. Om deze reden richt onderzoek zich vaak op de vraag welke correctievariabelen in een voorspellingsmodel moeten zitten om tot een valide leerrendementsverwachting te kunnen komen. Aan het gebruik van correctievariabelen kleeft echter ook een belangrijk nadeel. Het verzamelen van relevante leerling- en schoolkenmerken blijkt in de praktijk namelijk problematisch, tijdrovend en duur te zijn. Bovendien bestaat er momenteel geen overeenstemming over de vraag welke variabelen meegenomen moeten worden in het model en kan het ethisch bezwaarlijk zijn om bepaalde groepen leerlingen op voorhand hoger of lager in te schatten.

Een vierde en laatste aandachtspunt heeft betrekking op de procedure die gebruikt wordt om een voorspelling te doen over de ontwikkelingsmogelijkheden van individuele leerlingen. In de praktijk zien we dikwijls dat het leerrendement en de eindverwachting bepaald worden met behulp van *didactische leeftijdsequivalenten* (DLE). Wanneer de DLE kleiner is dan de didactische leeftijd (DL) loopt de leerling achter bij het gemiddelde van zijn groep. Het leerrendement is in dat geval kleiner dan 1.00. In geval een leerling een hogere score dan het gemiddelde van zijn groep behaalt, is het omgekeerde het geval. De eindverwachting wordt bepaald door het leerrendement in een bepaalde periode te vermenigvuldigen met het totale aantal maanden onderwijs dat een leerling krijgt aangeboden gedurende de basisschoolperiode. Hoewel de procedure zeer eenvoudig is toe te passen in volgsystemen is het gebruik ervan kwetsief (Evers & Resing, 2007). De procedure gaat bijvoorbeeld uit van een lineaire groeisnelheid voor alle leerlingen en houdt op geen enkele manier rekening met ontwikkelingsverschillen bij individuele leerlingen. Het gevolg kan zijn dat het toekomstperspectief dat uit de procedure volgt uiteindelijk zeer onrealistisch blijkt te zijn. In feite vragen longitudinale meetresultaten om een groeimodel (cf. Keuning, 2008). Een dergelijk model beschrijft de ontwikkeling van een groep personen op een bepaald gebied in relatie tot één of meer verklarende variabelen op een dynamische manier. Dit betekent dat het model metingen bij dezelfde personen op verschillende tijdstippen expliciet aan elkaar relateert. Enerzijds biedt een groeimodel mogelijkheden om het fluctueren van de vaardigheid in de tijd op het individuele niveau te onderdrukken (Kamphuis & Engelen, 1993). Anderzijds ontstaan er mogelijkheden om op basis van een model dat empirisch getoetst is, uitspraken te doen over

toekomstige prestaties van individuele leerlingen.

2.2 Growth mixture benadering

Cito heeft de afgelopen jaren gewerkt aan een procedure die de eerdergenoemde methodologische kanttekeningen bij het genereren van leerrendementsverwachtingen (gedeeltelijk) kan wegnemen. De procedure bestaat uit drie fasen. In de eerste fase wordt een meet-schaal ontwikkeld die scholen in staat stelt om de vaardigheid van leerlingen over een langere periode op een betekenisvolle manier te volgen. De meetschalen worden ontwikkeld met behulp van een meetmodel uit de item respons theorie (zie bijvoorbeeld Embretson & Reise, 2000; Van der Linden & Hambleton, 1997). In de meeste gevallen maakt Cito gebruik van het OPLM (Verhelst & Glas, 1995). De item respons functie voor het OPLM wordt gegeven door de volgende vergelijking:

$$P(X_j = 1 | \theta) = \frac{\exp [a_j (\theta - \beta_j)]}{1 + \exp [a_j (\theta - \beta_j)]},$$

waarin θ de vaardigheid van een leerling weergeeft, $a_j > 0$ de discriminatie-index voor item j is, en β_j de moeilijkheidsgraad van een item representeert. X is een random variabele met een waarde gelijk aan 0 of 1. Zoals we kunnen zien, geeft het model de kans op een correct antwoord ($X_j = 1$) op item j met discriminatie-index a_j en moeilijkheidsparameter β_j weer als functie van θ . Als het model in voldoende mate bij de data past kan de vaardigheid van een leerling met elke willekeurige deelverzameling met geschaalde items op dezelfde meet-schaal geschat worden. Op elk tijdstip verkrijgen we dus een score die dezelfde betekenis heeft, ook als niet op elk tijdstip dezelfde toets wordt afgenomen. In de tweede fase worden de verschillende metingen in de tijd aan elkaar gerateerd met behulp van een groeimodel. Er worden in de literatuur verschillende

modellen beschreven waarmee we de gemiddelden μ en de variantie-covariantiematrix Σ van de latente vaardigheid op populatieniveau kunnen schatten (e.g., Bollen & Curran, 2004; Delsing & Oud, 2008; Hamaker, 2005; Jöreskog, 1979; Meredith & Tisak, 1990; Molenaar & Campbell, 2008; Werts, Linn & Jöreskog, 1978). Het voert te ver om hier uitgebreid op in te gaan. In de procedure die Cito hanteert wordt nadrukkelijk onderzocht in hoeverre de aanname van één enkele multivariaat normale populatieverdeling verdedigbaar is. Als de longitudinale data die de afgelopen jaren op grote schaal bij scholen verzameld zijn aanleiding geven om meerdere groepen leerlingen te definiëren, wordt een mixture van multivariaat normale verdelingen aangenomen (zie Muthén, 2004). In dat geval heeft elke latente klasse C een eigen vector met gemiddelden μ_C en een eigen variantie-covariantiematrix Σ_C . Op deze manier kan in het groeimodel rekening gehouden worden met mogelijke ontwikkelingsverschillen tussen leerlingen. Vanwege de praktische (en mogelijk ethische) bezwaren die kleven aan het gebruik van achtergrondvariabelen worden er geen voorspellers aan het groeimodel toegevoegd.

In de derde en laatste fase worden zowel het meetmodel als het groeimodel gebruikt om de longitudinale prestaties van leerlingen dynamisch te monitoren en te voorspellen. De procedure werkt als volgt (zie ook Kamphuis, 1993):

- 1 Schat de vaardigheid van een leerling op tijdstip 1 (θ_1) gegeven het toetsresultaat s_1 en de marginale verdeling van θ_1 met gemiddelde μ_1 en variantie σ_1^2 – ofwel bepaal de *Expected A Posteriori* (EAP) schatter voor θ_1 .
- 2 Voeg $\hat{\theta}_1$ toe aan het groeimodel en schat θ_2 op basis van θ_1 , s_2 , μ_2 en σ_2^2 .
- 3 Voeg $\hat{\theta}_1$ en $\hat{\theta}_2$ toe aan het groeimodel en schat θ_3 op basis van θ_1 , θ_2 , s_3 , μ_3 en σ_3^2 .
- 4 Herhaal de laatste stap totdat de vaardigheid voor alle tijdstippen (θ_t)

conditioneel geschat is op basis van de voorgaande vaardigheidsschattingen ($\hat{\theta}_t$).

Dit betekent dat we de informatie gebruiken die verzameld is bij eerdere meetmomenten om de vaardigheid op latere meetmomenten nauwkeuriger te schatten. Als er voor een bepaald tijdstip geen toetsresultaat beschikbaar is, wordt de vaardigheid voor dat tijdstip voorspeld op basis van het groeimodel en *alle* toetsresultaten die al wel beschikbaar zijn. Als er bijvoorbeeld twee toetsafnames hebben plaatsgevonden (m_3 en e_3) en we willen weten hoe een leerling naar verwachting zal presteren bij de vierde toetsafname (e_4), dan wordt θ_1 geschat met behulp van de EAP-schatter, en θ_2 op basis van de EAP-schatter én θ_1 . De score voor θ_3 is niet bekend, dus die wordt conditioneel geschat op basis van $\hat{\theta}_1$ en $\hat{\theta}_2$. Vervolgens wordt θ_4 geschat op basis van $\hat{\theta}_1$, $\hat{\theta}_2$ en $\hat{\theta}_3$.

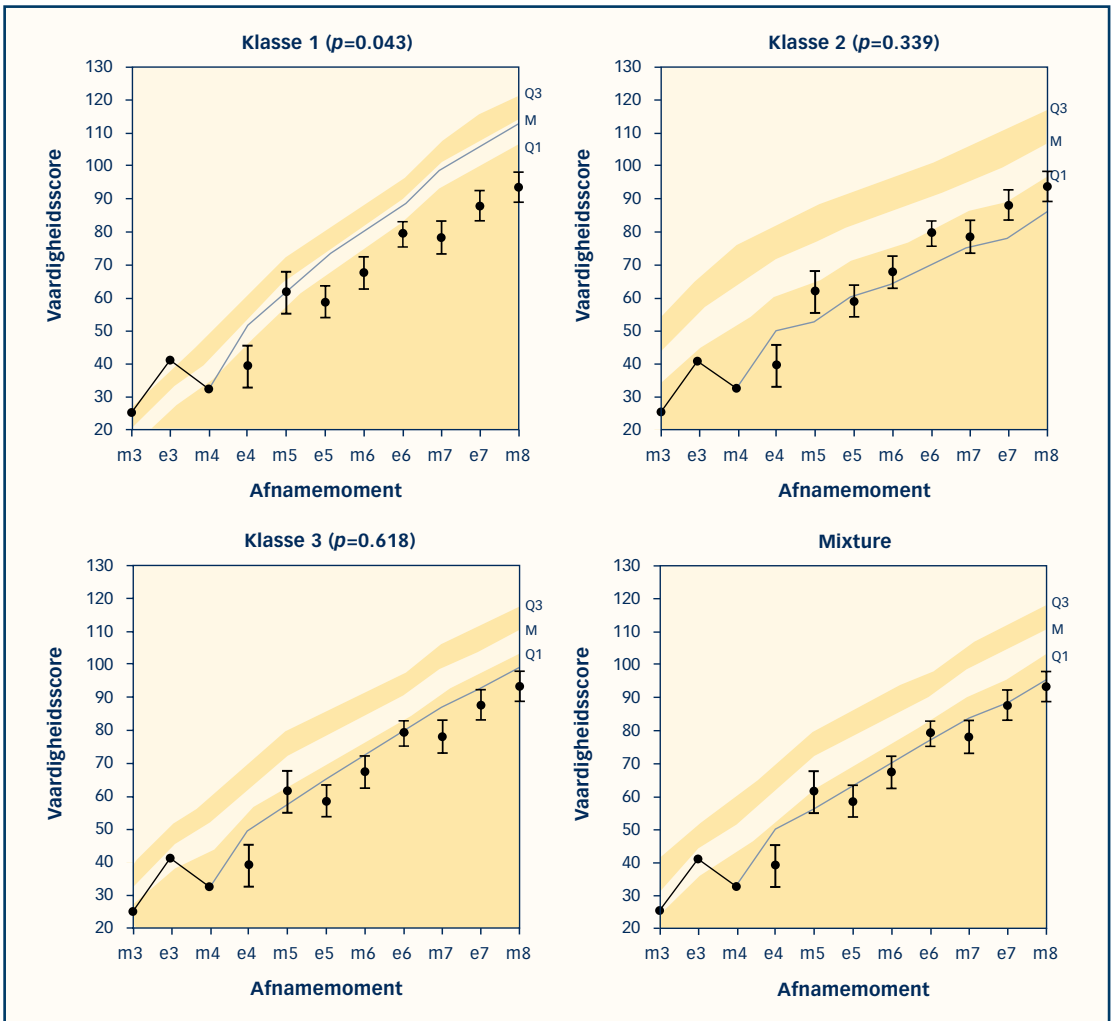
Bij de beschrijving van de laatste fase is voor het gemak aangenomen dat een groeimodel met één latente klasse een goede beschrijving geeft van de data. Als het groeimodel meerdere latente klassen bevat, worden de genoemde stappen voor elke klasse doorlopen. Figuur 1 laat zien hoe de procedure uitpakt bij een leerling die op 11 verschillende tijdstippen de toetsen Rekenen-Wiskunde uit het *Cito Volgsysteem primair en speciaal onderwijs* heeft gemaakt. De zwarte punten in Figuur 1 geven de vaardigheidsscores weer zoals die daadwerkelijk geobserveerd zijn. De blauwgekleurde lijn geeft de leerrendementsverwachting weer als we net doen alsof er slechts drie metingen hebben plaatsgevonden. De voorspelling is beter naarmate de zwarte punten dichter bij de blauwe lijn liggen. We zien dat de nauwkeurigheid van de voorspelling samenhangt met de kans dat een leerling tot een bepaalde klasse behoort. Als we de gemiddelden en covarianties van klasse 1 gebruiken, is de leerrendementsverwachting weinig

adequaat. De kans dat de ontwikkeling van deze leerling grote gelijkens zal vertonen met de ontwikkeling die andere leerlingen in klasse 1 laten zien, is dan ook erg klein ($p = .043$). Het is aannemelijker dat de leerling tot klasse 3 behoort ($p = .618$). De belangrijkste vraag is van welke leerrendementsverwachting we uit moeten gaan. Een praktische oplossing is om uit te gaan van de meest waarschijnlijke leerrendementsverwachting (in dit geval klasse 3). Een alternatief is om een ‘gewogen’ leerrendementsverwachting te maken op basis van de *posterior class probabilities*. Het resultaat hiervan staat rechtsonder

in Figuur 1. Deze aanpak leidt in dit voorbeeld tot de meest nauwkeurige leerrendementsverwachting.

2.3 Praktijkvoorbeelden

Om zicht te krijgen op het functioneren van de beschreven methodiek voor het voorspellen van het maximale leerrendement van individuele leerlingen is een dataset geanalyseerd met de gegevens van ruim 3000 leerlingen met 11 herhaalde metingen (m3, ... m8) op de LVS-toetsen Rekenen-Wiskunde van Cito. Uit de dataset zijn structureel datapunten verwijderd om die vervolgens op basis van het eerder beschreven model te voorspel-



Figuur 1 Leerrendementsverwachting op basis van een growth mixture model

Er is een dataset geanalyseerd met de gegevens van ruim 3000 leerlingen met 11 herhaalde metingen op de LVS-toetsen Rekenen-Wiskunde van Cito. Uit de dataset zijn structureel datapunten verwijderd om die vervolgens op basis van het eerder beschreven model te voorspellen. [...] Aangezien de echte waarden voor alle tijdstippen bekend waren, kon de accuraatheid van de leerrendementsverwachting gemakkelijk geëvalueerd worden door de voorspelde waarden met de echte waarden te vergelijken

len. Het groeiemodel bevatte drie latente klassen. Aangezien de echte waarden voor alle tijdstippen bekend waren, kon de accuraatheid van de leerrendementsverwachting gemakkelijk geëvalueerd worden door de voorspelde waarden met de echte waarden te vergelijken. Figuur 2 laat voor twee verschillende leerlingen de leerrendementsverwachting zien. Eerst is verondersteld dat beide leerlingen op drie achtereenvolgende meetmomenten getoetst zijn. Op basis van de eerder beschreven *growth mixture* aanpak is vervolgens voorspeld hoe de leerlingen naar verwachting zullen presteren bij volgende toetsafnames. Het gekleurde gebied geeft weer in hoeverre de leerrendementsverwachting verandert als er meer informatie beschikbaar zou zijn (bijvoorbeeld vier metingen in plaats van drie). De zwarte punt op tijdstip m8 representeert het eindniveau dat de leerling uiteindelijk heeft weten te realiseren. We zien dat de leerrendementsverwachting voor leerling A zeer stabiel is. We kunnen op basis van drie toetsafnames en het onderliggende groeiemodel goed voorspellen hoe de rekenvaardigheid van deze leerling zich zal ontwikkelen. Bij leerling B is het beeld tegengesteld. De ontwikkeling van deze leerling verloopt bij de eerste drie metingen zeer grillig en de leerrendementsverwachting die

gemaakt wordt op basis van deze drie metingen wijzigt nog sterk als informatie toegevoegd wordt. De enige zekerheid die er is op basis van drie metingen is dat het eindniveau van deze leerling ergens tussen percentiel 10 en 60 zal liggen. In werkelijkheid blijkt de leerling op tijdstip m8 net onder het twintigste percentiel te scoren.

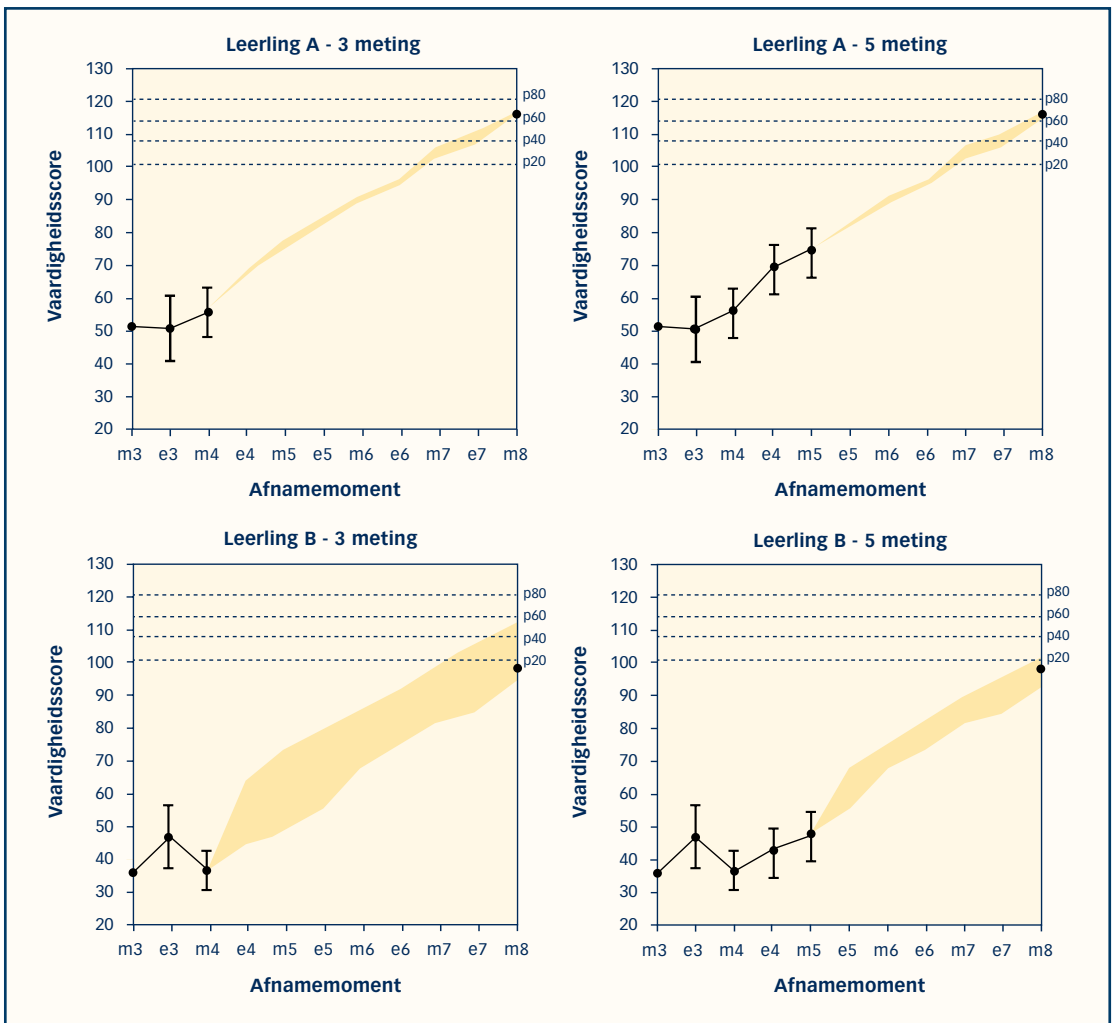
In een vervolgstap is gekeken in hoeverre de leerrendementsverwachtingen voor deze twee leerlingen nog variatie vertonen als er in plaats van drie opeenvolgende metingen, vijf beschikbaar zijn. De resultaten zijn weergegeven aan de rechterkant van Figuur 2. Voor leerling A is de leerrendementsverwachting onverminderd stabiel. Het toevoegen van twee toetsresultaten heeft ook nauwelijks effect op het verwachte eindniveau. Dit betekent dat er voor leerling A (achteraf gezien) geen enkele aanleiding zou zijn geweest om na drie toetsafnames nog te wachten met het maken van een leerrendementsverwachting. Bij leerling B zien we daarentegen dat de onzekerheid in de leerrendementsverwachting sterk afneemt als er twee extra toetsafnames plaatsvinden. Waar we op basis van drie toetsafnames slechts konden gissen naar het eindniveau van de leerling, weten we op basis van vijf toetsafnames behoorlijk zeker dat de leerling aan het einde van de basisschoolperiode tot de zwakste twintig procent zal behoren. De onzekerheidsmarges laten zien dat er over de leerrendementsverwachting van leerling B wel minder zekerheid bestaat dan over de leerrendementsverwachting van leerling A. Dit komt doordat leerling B na een voortvarende start in groep 3 een sterke terugval in vaardigheid laat zien bij de toetsafname op tijdstip m4. Hoewel de terugval niet significant is op het 90-procent niveau, is het met een dergelijk grillig groeipatroon zeer problematisch om nauwkeurige voorspellingen te doen over de ontwikkelingsmogelijkheden van de leerling. Bij de toetsafna-

mes op de tijdstippen e4 en m5 zien we dat het vaardigheidsniveau van leerling B zich stabiliseert onder percentiel 20. Met deze nieuwe informatie wordt het mogelijk om de leerrendementsverwachting met een grotere precisie op te stellen.

Met name bij de zwakkere leerlingen verloopt de ontwikkeling soms erg grillig, terwijl scholen juist voor die leerlingen een ontwikkelingsperspectief moeten opstellen. Het onderzoek laat het effect van het grillige ontwikkelingsverloop op de leerrendementsverwachting duidelijk zien

3 Conclusies en discussie

Het onderzoek van Cito laat zien dat het opstellen van leerrendementsverwachtingen niet eenvoudig is. We moeten rekening houden met een mogelijk grillige ontwikkeling in vaardigheid, de (on)betrouwbaarheid van metingen, en de lengte van de periode waarvoor de leerrendementsverwachting wordt opgesteld. Met name bij de zwakkere leerlingen verloopt de ontwikkeling soms erg grillig, terwijl scholen juist voor die leerlingen een ontwikkelingsperspectief moeten opstellen. Het on-



Figuur 2 **Nauwkeurigheid van een leerrendementsverwachting voor een gemiddelde (A) en een zwakke leerling (B)**

derzoek laat het effect van het grillige ontwikkelingsverloop op de leerrendementsverwachting duidelijk zien. Op basis van drie toetsresultaten kunnen we voor leerling B in Figuur 2 halverwege groep 4 hooguit voorspellen dat zijn rekenvaardigheid in groep 8 uitkomt tussen het 10^{de} en 60^{ste} percentiel. De leerling zou qua rekenvaardigheid aan het einde van de basisschoolperiode dus gemiddeld kunnen presteren, maar ook op vaardigheidsniveau E uit kunnen komen. Een jaar later kan het eindniveau nauwkeuriger voorspeld worden. We beschikken dan namelijk over twee extra toetsresultaten en bovendien is de periode waarover voorspeld wordt korter. Ook dan is de spreiding echter niet verwaarloosbaar. De rekenvaardigheid van de leerling zal naar verwachting op vaardigheidsniveau D of E uitkomen. Het voorbeeld in Figuur 2 bevestigt de vraagtekens die Louwe (2013) plaatst bij geautomatiseerde procedures om ontwikkelingsperspectieven op te stellen, veelal gebaseerd op sterk vereenvoudigde modellen van de werkelijkheid. We zien dat het al lastig is om voor één vakgebied een betrouwbare leerrendementsverwachting te formuleren. Het combineren van de leerrendementsverwachtingen voor meerdere vakgebieden tot één ontwikkelingsperspectief zal zeer zeker problematischer zijn, en dan is er nog niet eens gekeken naar het onderwijsaanbod, het sociaal-emotioneel functioneren van de leerling en de leerwerkhouding. Uit het onderzoek van Cito blijkt tevens dat het onwenselijk is om leerrendementsverwachtingen te formuleren die een lange tijdsperiode omvatten. Louwe (2013) gaf al aan dat naarmate kinderen jonger zijn en de problematiek complexer is, het lastiger wordt om met zekerheid voorspellingen te doen over de ontwikkelingsmogelijkheden. Leerling B in Figuur 2 is daarvoor illustratief. Ook met een model dat optimaal rekening houdt met interindividuele verschillen

en de (on)betrouwbaarheid van metingen is het eindniveau voor leerling B dermate onnauwkeurig te voorspellen, dat het onverstandig zou zijn om op basis van zijn leerprestaties te besluiten om het onderwijsaanbod te beperken. In het voorbeeld wéten we welk resultaat de leerling uiteindelijk heeft behaald in groep 8. In de praktijk weten we dat vanzelfsprekend niet. Het risico bestaat nu dat de leerkracht op basis van een onzekere leerrendementsverwachting het onderwijsaanbod wil beperken en besluit om bepaalde onderdelen - van in dit geval rekenen - niet meer aan te bieden. De intentie van de leerkracht zal zijn om het onderwijs af te stemmen op het vaardigheidsniveau van de leerling. Tegelijkertijd zal die beperking in het onderwijsaanbod er echter voor zorgen dat de rekenvaardigheid van de leerling ook niet hoger kán uitkomen. De *self fulfilling prophecy* is dan een feit; er wordt niet veel van de rekenvaardigheid van de leerling verwacht, we bieden hem daarom een beperkt aanbod, en (mede) daardoor blijkt de rekenvaardigheid uiteindelijk ook niet hoger uit te komen dan onze verwachting. Het is uitermate belangrijk om dit te voorkomen. Temeer omdat de grote onzekerheidsmarge in het model van Cito ontstaat doordat er andere leerlingen zijn met vergelijkbare toetsresultaten die uiteindelijk wél een hogere rekenvaardigheid hebben verworven dan leerling B. Het uiteindelijk behaalde leerrendement staat dus niet per definitie vast, gegeven het startniveau. Dezelfde voorzichtigheid geldt bij het gebruik van IQ-scores. Deze hangen over het algemeen tussen de 0.40 en 0.60 samen met schoolsucces, waardoor iets als uitstroomniveau of leerrendementsverwachting niet een-op-een afgeleid kan worden uit een IQ-score. In het licht van de onderzoeksresultaten zou een leerrendementsverwachting bij voorkeur maximaal één leerjaar moeten beslaan. Pas op het einde van het basis-

onderwijs zou dan een uitstroomprofiel opgesteld kunnen worden. Op basis van de leerrendementsverwachting kan de leerkracht onderwijs op maat aanbieden en kan hij tussentijds (bijvoorbeeld na een halfjaar) het tot dan toe gerealiseerde leerresultaat vergelijken met de verwachting. Feitelijk volgt de leerkracht hiermee de *cyclus plan - do - check - act*. De leerkracht vergelijkt zijn verwachting met het daadwerkelijk behaalde resultaat, interpreteert een eventueel verschil op basis van het onderwijsaanbod en andere factoren, stelt zo nodig de verwachting bij, bepaalt het onderwijs voor de volgende periode en geeft kwalitatief hoogwaardig onderwijs (zie ook Parrett & Budge, 2009). Onderzoek laat zien dat scholen die op deze manier werken uiteindelijk meer uit hun leerlingen halen dan scholen die dat niet doen (cf. Fielding, Kerr & Rosier, 2007). Cito kiest er dan ook voor om in het *Computerprogramma LOVS* geen langetermijnvoorspellingen te doen over het leerrendement en het uitstroomniveau van individuele leerlingen. Wel zullen er op basis van het eerder gepresenteerde model in de leerlingrapportages van Cito per vaardigheid leerrendementsverwachtingen opgenomen worden. De verwachting zal betrekking hebben op maximaal één leerjaar en de onzekerheidsmarge die bij de verwachting hoort, zal in de rapportage worden weergegeven. Daarmee willen we benadrukken dat het van belang is dat leerkrachten telkens opnieuw het onderwijs en de opbrengsten evalueren, en de verwachtingen bijstellen. De keuze om maximaal één leerjaar vooruit te voorspellen lost het risico op een *self fullfilling prophecy* niet op. Het blijft voor leerkrachten immers lastig om te bepalen hoe de meer objectieve informatie die verkregen wordt uit

toetsen gecombineerd kan worden met andere, veelal subjectieve, observaties. Daar komt bij dat leerlingen met eenzelfde startniveau in de praktijk soms op een ander eindniveau terechtkomen. Wat is in dat geval een realistische leerrendementsverwachting? Naar de oorzaken van de verschillen is het gissen. Verschillende onderzoeken laten zien dat goed onderwijs, het hebben van hoge verwachtingen en het systematisch gebruik van een leerlingvolgsysteem positief bijdragen aan de leerresultaten van leerlingen. Een dergelijke aanpak vraagt echter veel van de professionaliteit van leerkrachten. In de eerste plaats veronderstelt de aanpak dat leerkrachten uitgaan van leerstandaarden en lesdoelen, dat zij informatie verzamelen tijdens het leerproces, dit vastleggen voor nadere analyse en interpretatie, en op basis hiervan beslissingen nemen over het vervolg van het onderwijs (Parrett & Budge, 2009). In de tweede plaats is het belangrijk dat leerkrachten een positieve attitude hebben tegenover het omgaan met verscheidenheid in leren (zie Meijer, 2003). Ten slotte dienen zij te beschikken over inhoudelijke kennis omtrent de ontwikkeling van leerlingen in de tijd en op basis van die kennis moeten zij antwoord kunnen geven op de vraag welke leerstof en didactische benadering het meest geschikt is om leerlingen optimaal te begeleiden (Reezigt, Houtveen & Grift, 2002). Louwe (2013) stelt zeer terecht dat dit niet volautomatisch via een computersysteem kan en mag gebeuren. Leerrendementsverwachtingen kunnen slechts houvast geven bij het analyseren van leerresultaten. Uiteindelijk is het de leerkracht die besluit welke gegevens relevant zijn, die de gegevens analyseert en tegen elkaar afweegt om te komen tot een onderwijsaanbod afgestemd op zijn leerlingen.

GERAADPLEEGDE LITERATUUR

- Benner, A.D. & Mistry, R.S. (2007). Congruence of mother and teacher educational expectations and low-income youth's academic competence. *Journal of Educational Psychology, 99*, 140-153.
- Bock, R.D. (1976). Basic issues in the measurement of change. In D.N.M. De Gruijter & L.J.T. Van der Kamp (Eds.), *Advances in psychological and educational measurement* (pp. 75-96). London: Wiley.
- Bollen, K.A. & Curran, P.J. (2004). Autoregressive Latent Trajectory Models: A synthesis of two traditions. *Sociological Methods & Research, 32*, 336-383.
- Clijisen, A., Pieterse, E., Spaans, G. & Visser, J. (2009). *Werken vanuit een ontwikkelingsperspectief in het speciaal basisonderwijs - naar een gezamenlijk kader*. Utrecht: PO-raad / SBO Werkverband.
- Delsing, M.J.M.H. & Oud J.H.L. (2008). Analyzing reciprocal relationships by means of the continuous-time autoregressive latent trajectory model. *Statistica Neerlandica, 62*, 58-82.
- Embretson, S.E. & Reise, S. (2000). *Item response theory for psychologists*. Mahwah, NJ: Erlbaum Publishers.
- Evers, A. & Resing, W. (2007). Het drijfzand van didactische leeftijdsequivalenten. *De Psycholoog, 9*, 466-472.
- Fielding, L., Kerr, N. & Rosier, P. (2007). *Annual growth for all students*. Kennewick, WA: New Foundation Press.
- Francis, D.J., Shaywitz, S.E., Stuebing, K.K., Shaywitz, B.A. & Fletcher, J.M. (1996). Developmental lag versus deficit models of reading disability: A longitudinal, individual growth curves analysis. *Journal of Educational Psychology, 88*, 3-17.
- Gill, S. & Reynolds, A.J. (1999). Educational expectations and school achievement of urban African American children. *Journal of School Psychology, 37*, 403-424.
- Hakkenes, A. & Wijs, A. de (2012). Van Citotoets naar brugklas en door naar diploma. *Sociaal-economische trends, 1*, 65-79.
- Hamaker, E.L. (2005). Conditions for the equivalence of the autoregressive latent trajectory model and a latent growth curve model with autoregressive disturbances. *Sociological Methods Research, 33*, 404-416.
- Hambleton, R.K., Swaminathan, H. & Rogers, H.J. (1991). *Fundamentals of item response theory*. Newbury Park, CA: Sage.
- Inspectie van het Onderwijs (2012). *Toezichtkader (voortgezet) speciaal onderwijs*. Utrecht: Inspectie van het onderwijs. Download d.d. 8 februari 2013. http://www.onderwijsinspectie.nl/binaries/content/assets/Actueel_publicaties/2012/brochure-toezichtkader_ec_2_web_def.pdf.
- Jöreskog, K.G. (1979). Statistical models and methods for analysis of longitudinal data. In K.G. Jöreskog & D. Sörbom, *Advances in Factor Analysis and Structural Equation Models* (pp. 129-169). Cambridge, MA: Abt Books.
- Kamphuis, F. (1993). Estimation and prediction of individual ability in longitudinal studies. In J.H.L. Oud & R.A.W. van den Blokland-Vogeesang (Eds.), *Advances in long and multivariate analysis in the behavioral sciences*. Nijmegen: ITS.
- Kamphuis, F.H. & Engelen, R.J.H. (1993). Het meten van veranderingen. In T.J.H.M. Eggen & P.F. Sanders (Eds.), *Psychometrie in de praktijk*. Arnhem: Cito.
- Kaplan, D. & Walpole, S. (2005). A Stage-Sequential Model of Reading Transitions: Evidence From the Early Childhood Longitudinal Study. *Journal of Educational Psychology, 4*, 551-563.
- Keuning, J. (2008). *Monitoring growth in reading and spelling: Applications of item response theory and covariance structure analysis*. Nijmegen: Expertisecentrum Nederlands.
- Lord, F.M. (1984). Standard errors of measurement at different ability levels. *Journal of Educational Measurement, 21*, 239-243.
- Louwe, J.J. (2013). Werken vanuit een ontwikkelingsperspectief in het primair onderwijs: een zoektocht door het drijfzand. *Tijdschrift voor Orthopedagogiek, 52*, 591-602.
- Meijer, C.J.W. (2003). *Inclusive Education and Effective Classroom Practices*. Denmark: European Agency for Development in Special Needs Education.
- Meredith, W. & Tisak, J. (1990). Latent curve analysis. *Psychometrika, 55*, 107-122.
- Merton, R.K. (1968). The Matthew effect in science. *Science, 159*, 56-63.

- Moelands, A.H.J. (2007). *Stand van zaken leerrendementsverwachting per september 2007*. Arnhem: Cito.
- Molenaar, P.C.M. & Campbell, C.G. (2008). Discussion of the special issue on growth models for longitudinal data in educational research. *Educational Research and Evaluation*, 4, 377-390.
- Muthén, B.O. (2004). Latent variable analysis: Growth mixture modeling and related techniques for longitudinal data. In D. Kaplan (ed.), *Handbook of quantitative methodology for the social sciences* (pp. 345-368). Newbury Park, CA: Sage Publications.
- Notenboom, A. & Reitsma, P. (2003). Investigating the dimensions of spelling ability. *Educational and Psychological Measurement*, 6, 1039-1059.
- Parrett, W. & Budge, K. (2009). Tough questions in assessment. *Educational Leadership*, 67, 22-27.
- Reezigt, G.J., Houtveen, A.A.M. & Grift, W. van de (2002). *Ontwikkelingen in en effecten van adaptief onderwijs in de klas en integrale leerlingenzorg op schoolniveau*. Groningen: GION.
- Rist, R.C. (2000). HER classic: Student social class and teacher expectations: The self-fulfilling prophecy in ghetto education. *Harvard Educational Review*, 70, 257-301.
- Stanovich, K.E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly*, 26, 7-29.
- Stroucken, L., Takkenberg, D. & Béguin, A. (2008). Cito-toets en de overgang van basisonderwijs naar voortgezet onderwijs. *Sociaaleconomische trends*, 2, 7-16.
- Van der Linden, W.J. & Hambleton, R.K. (1997). *Handbook of Modern Item Response Theory*. New York: Springer.
- Verhelst, N.D. & Glas, C.A.W. (1995). The one-parameter logistic model. In G.H. Fischer & I.W. Molenaar (Eds.), *Rasch models. Foundations, recent developments, and applications*. New York: Springer-Verlag.
- Werts, C.E., Linn, R.L. & Jöreskog, K.G. (1978). A simplex model for analyzing academic growth. *Educational and Psychological Measurement*, 37, 745-756.
- Wieberdink, H. & Kuster, H. (2011). De uitzichtloosheid van het ontwikkelingsperspectief. *Orthopedagogiek: onderzoek en praktijk*, 50, 173-180.
- Wilson, M. (2005). *Constructing measures: An item response modeling approach*. Mahwah, NJ: Erlbaum.

OVER DE AUTEURS



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 betrokken geweest bij projecten die gericht waren op de ontwikkeling van tests voor specifieke doelgroepen. Daarnaast doet hij onderzoek naar leerrendementsverwachting en schooleffectiviteit.
E-mail: Jos.Keuning@cito.nl



Jacqueline Visser is onderwijskundige en organisatieadviseur. In 1997 is zij gepromoveerd in de sociale wetenschappen, op de validering en normering van de AVI-toets. Sinds 2007 werkt zij bij Cito, eerst als marktgroepmanager speciaal (basis)onderwijs en vanaf 2010 als marktgroepmanager basisonderwijs. Zij is in deze functie eindverantwoordelijk voor alle producten van Cito voor het basisonderwijs, van de kleutertoetsen tot en met de Eindtoets Basisonderwijs.
E-mail: Jacqueline.Visser@cito.nl