

Het effect van een jaar onderwijs op de wiskundeprestaties in groep 5 en 6 en de verschillen tussen scholen

H. Luyten

Samenvatting

Dit artikel laat zien hoe men meerniveau analyse, de standaardmethode om “schooleffecten” in kaart te brengen, kan combineren met onderbroken regressie (‘regression discontinuity’). Terwijl gebruikelijke meerniveau analyses uitsluitend een schatting van relatieve effecten opleveren (verschillen tussen scholen), kan met onderbroken regressie het effect van een jaar onderwijs gerelateerd worden aan een betekenisvol nulpunt. Met een combinatie van beide benaderingen wordt zowel het absolute effect als de spreiding ervan tussen scholen in kaart gebracht. Toepassing in Nederland vergt vanwege het hoge percentage zittenblijvers wel aanvullende assumpties. Secundaire analyses op data van TIMSS-95 (Third International Mathematics and Science Study) laten zien dat deze benadering verrassende inzichten oplevert over de effectiviteit van het Nederlandse basisonderwijs en de verschillen tussen scholen.

1 Inleiding

Al vier decennia lang, sinds het verschijnen van het beroemde *Coleman Report* (Coleman, Campbell, Hobson, McPartland, Mood, Weinfeld, & York, 1966), doelt men met de term *schooleffecten* doorgaans op het percentage variantie in leerprestaties dat kan worden toegeschreven aan verschillen tussen scholen. Destijds trok de bevinding dat in de VS slechts 10 à 15% van alle variantie gesitueerd bleek op het schoolniveau veel aandacht. Het leeuwendeel van de spreiding betrof verschillen tussen leerlingen binnen scholen. Algemeen werd dit opgevat als een teleurstellende uitkomst. Het leek te impliceren dat het effect van school op leerprestaties beperkt is. De bevinding dat slechts een bescheiden percentage van de totale spreiding is gesitueerd op schoolniveau is sindsdien

echter talloze malen bevestigd (Scheerens & Bosker, 1997).

We moeten niet uit het oog verliezen dat met schooleffecten bedoeld wordt op de relatieve omvang van verschillen tussen scholen. Een bescheiden omvang hiervan hoeft nog niet te betekenen dat ook de bijdrage van het onderwijs aan de ontwikkeling van leerlingen bescheiden is. Geringe verschillen kunnen ook duiden op een hoge onderwijskwaliteit in het merendeel van de scholen. Een consequentie van de focus op relatieve verschillen is dat men altijd scholen vindt met resultaten boven het gemiddelde en scholen die dat niveau niet halen.

2 Onderbroken regressie

De grote moeilijkheid bij het bepalen van de bijdrage die het onderwijs levert aan de ontwikkeling van leerlingen is het gegeven dat vrijwel iedereen naar school gaat. De ideale situatie – methodologisch gezien – zou zijn dat men op toevalsbasis leerlingen kan indelen in een experimentele groep en een controlegroep, waarbij de laatste van onderwijs verstoken blijft. Dat is natuurlijk geen reële optie. Toch worden in dit artikel schattingen van het effect van een jaar onderwijs op de leerprestaties gerapporteerd. Deze zijn verkregen door een aanpak die gebaseerd is op het onderbroken-regressieontwerp. Dit onderzoeksdesign is van toepassing als de criteria op basis waarvan “proefpersonen” in de experimentele groep dan wel in de controlegroep geplaatst worden, precies bekend zijn. Na controle voor de invloed van deze criteria is het mogelijk het effect van de experimentele variabele zeer exact te schatten (Rossi, Freeman & Lipsey, 2004; Shadish, Cook & Campbell, 2002; Swanborn, 1999; Trochim, 1984).

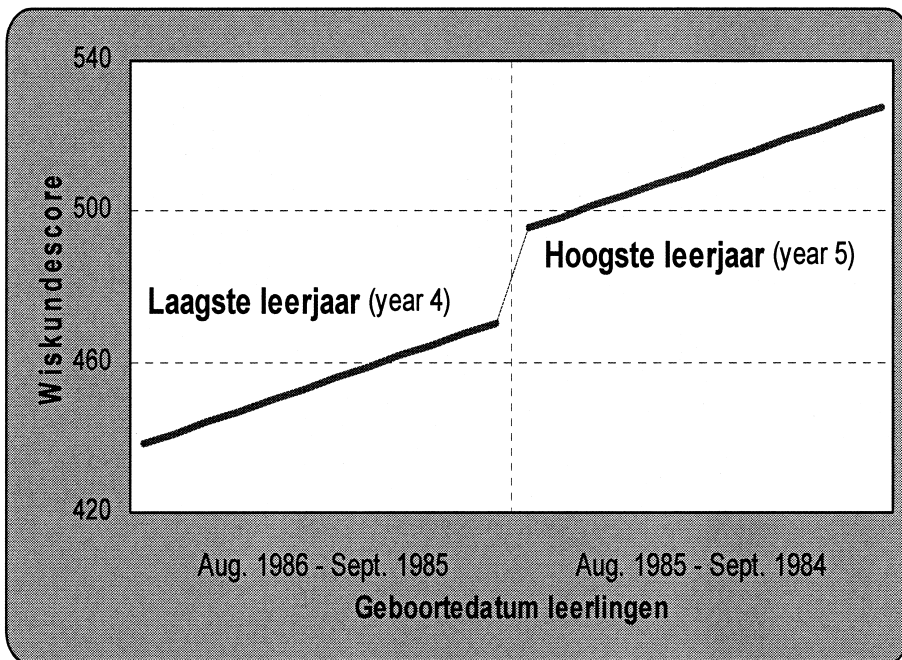
In de meeste landen bepaalt de geboortedatum in eerste instantie in welk leerjaar een

leerling terechtkomt. In Nederland zitten leerlingen die voor 1 oktober hun verjaardag vieren normaal gesproken in een hogere jaargroep dan degenen die later in het jaar zijn geboren. Indien een dergelijk leeftijds criterium voldoende strikt wordt toegepast kan op betrekkelijk eenvoudige wijze een zeer exacte schatting worden gemaakt van het effect van een jaar extra onderwijs. Deze aanpak vereist slechts dat men beschikt over gegevens uit twee (of meer) opeenvolgende jaargroepen (bijvoorbeeld toetsscores die betrekking hebben op een gemeenschappelijke schaal). Per leerjaar wordt het effect van leeftijd op leerprestaties geschat. Een eventuele onderbreking in de lijn die het effect van leeftijd op leerprestaties weergeeft tussen de oudste leerlingen in de laagste jaargroep en de jongste in de hoogste groep levert een valide schatting op van het effect van één jaar onderwijs. Een dergelijke discontinuïteit laat zien hoe groot het verschil in leerprestaties is tussen leerlingen die in twee opeenvolgende jaargroepen zitten, maar die verder niet van elkaar verschillen. Figuur 1 geeft een grafische weergave van de resultaten die deze aanpak kan opleveren. De figuur beschrijft de uitkomsten voor leerlingen in het Engel-

se basisonderwijs (Luyten, 2006).

De kracht van onderbroken regressie schuilt in het feit dat men controleert voor de invloed van het criterium (leeftijd) dat bepaalt in welke jaargroep een leerling terechtkomt. Net als bij het klassieke experiment is het selectiecriterium op basis waarvan proefpersonen in de experimentele of de controlegroep worden geplaatst bekend (Shadish et al., 2002; p. 224). In het klassieke experiment is dit toeval. Bij onderbroken regressie hebben we te maken met een ander selectiecriterium, maar kunnen we de invloed ervan in de data-analyse verdisconteren. Alleen als een derde factor precies samenvalt met de leeftijdsgrens leidt het onderbroken-regressieontwerp tot een vertekende schatting van het effect van een jaar onderwijs. Het is wel van cruciaal belang dat de relatie tussen het selectiecriterium en de afhankelijke variabele correct gemodelleerd wordt. Als een lineaire functie wordt geschat, terwijl het verband in werkelijkheid kromlijinig is, kan dit een vertekende schatting opleveren.

Overigens is de sterkte van de relatie tussen het selectiecriterium en de afhankelijke variabele niet van belang. Afwezigheid van samenhang leidt evenmin tot vertekeningen.



Figuur 1. Relatie tussen geboortedatum en wiskundescore in Engeland

In dat geval komt het onderzoeksdesign neer op een vergelijking tussen de gemiddelde scores in beide leerjaren. Indien de correlatie tussen de afhankelijke variabele en het selectie criterium gelijk is aan nul, wordt een situatie benaderd waarbij toewijzing aan de controlegroep of de experimentele groep plaatsvindt op basis van toeval (Trochim, 1984, pp. 82-83). Het enige wat telt is duidelijkheid omtrent het selectie criterium op basis waarvan leerlingen aan een bepaalde jaargroep worden toegewezen. In de data-analyse corrigeert men dan voor de eventuele invloed van het selectie criterium.

In Nederland (en andere landen) wordt leeftijd echter flexibel toegepast als selectie criterium bij de toewijzing van leerlingen aan leerjaren. Met name het percentage leerlingen dat vertraging oploopt, is aanzienlijk. Bovendien is de kans op een afwijkende loopbaan aanmerkelijk groter voor leerlingen met een geboortedatum die dicht in de buurt van 1 oktober ligt. Dit levert serieuze complicaties op als men met onderbroken regressie het effect van een jaar onderwijs wil vaststellen, vooral omdat de criteria op basis waarvan besloten wordt een leerling te laten doubleren weinig helder zijn en van school tot school verschillen (Doornbos, 1969; Reezigt & Knuver, 1995). In dit artikel wordt aangegeven hoe men op basis van realistische assumpties toch gebruik kan maken van onderbroken regressie.

3 Eerder onderzoek naar het effect van onderwijs met onderbroken regressie

Tot dusver is zeer sporadisch gebruik gemaakt van onderbroken regressie om de bijdrage van onderwijs aan de ontwikkeling van leerlingen vast te stellen. Cahan en Davis (1987) hebben dit als eerste gedaan. Hun onderzoek had betrekking op de eerste twee leerjaren van het Israëliësch basisonderwijs in 1973. Tymms, Merrell en Henderson (1997) hebben in hun studie, die zich richt op de eerste leerjaren in het Engelse basisonderwijs, een methode toegepast die veel verwantschap vertoont met de aanpak van Cahan en Davis. Beide studies rapporteren

uitkomsten die sterk afwijken van wat doorgaans in schooleffectiviteitsonderzoek naar voren komt. Zo concludeerden Cahan en Davis dat ongeveer tweederde van het verschil in toetscores tussen twee opeenvolgende leerjaren is toe te schrijven aan het gevolgde onderwijs en eenderde aan het verschil in leeftijd. Als het effect van één jaar verschil in leeftijd (10 punten) werd afgetrokken van het verschil in leerprestaties tussen beide jaargroepen (29 punten) nam het verschil met eenderde af. Omdat beide jaargroepen afgezien van het genoten onderwijs alleen in leeftijd van elkaar verschilden werd de rest van het verschil aan het onderwijs toegeschreven. Ook in de studie van Tymms en anderen kon maar een beperkt deel van het verschil in leerprestaties tussen twee jaargroepen (60 punten) worden toegeschreven aan het effect van één jaar verschil in leeftijd (18 punten).

In een recente studie van Luyten (2006) is een benadering gevolgd die onderbroken regressie combineert met meerniveau analyse. Leerprestaties zijn in navolging van Cahan en Davis (1987) gemodelleerd als een functie van leeftijd en jaargroep. Meerniveau analyse biedt de mogelijkheid om het effect van een jaar onderwijs tussen scholen te laten variëren. Zodoende is niet alleen de bijdrage berekend die een jaar onderwijs levert aan de ontwikkeling van leerlingen, maar werd eveneens nagegaan in hoeverre deze verschilde tussen scholen. De analyses zijn, evenals die waarvan in dit artikel verslag wordt gedaan, uitgevoerd op data afkomstig van TIMSS-95 (Third International Mathematics and Science Study). Aan dit project hebben in totaal 45 landen deelgenomen. Zowel in het primair als secundair onderwijs zijn toetsen afgenomen die betrekking hebben op het wiskunde- en natuuronderwijs. In de meeste landen werden de gegevens verzameld in mei of juni 1995, maar in de landen op het zuidelijke halfrond vond de dataverzameling eind 1994 plaats. De studie van Luyten heeft betrekking op het primair onderwijs. Dit deelproject van TIMSS-95 was gericht op de leerjaren met de hoogste percentages 9- en 10-jarige leerlingen. In totaal zijn in 26 landen tests afgenomen bij leerlingen in het basisonderwijs

(Mullis, Martin, Beaton, Gonzales, Kelly, & Smith, 1997; Martin, Mullis, Beaton, Gonzales, Smith, & Kelly, 1997; Gonzales & Smith, 1997).

Het onderzoek van Luyten (2006) is beperkt gebleven tot acht landen (zie Tabel 6). In de overige landen stuitte toepassing van onderbroken regressie op complicaties, omdat leeftijd onvoldoende strikt als selectie-criterium is toegepast. In veel landen loopt een aanzienlijk deel van de leerlingen al in het basisonderwijs vertraging op. Zolang het percentage “misclassificaties” beperkt blijft tot 5% kan men de analyses beperken tot de leerlingen die zich gezien hun leeftijd in de “juiste” jaargroep bevinden (Judd & Kennedy, 1981; Trochim, 1984; Shadish et al., 2002). Slechts in acht landen was dat het geval. Voor deze landen is vastgesteld dat er na correctie voor het effect van leeftijd sprake was van een significant leerjareffect. In de meeste gevallen was dit effect sterker dan het effect van leeftijd. Volgens de benadering van Cahan en Davis (1987) impliceert dit dat meer dan de helft van de cognitieve ontwikkeling van leerlingen moet worden toegeschreven aan het gevolgde onderwijs. Gemiddeld genomen was de bijdrage van het onderwijs ruim 55%, maar de percentages varieerden van 75% in Singapore tot 35% in Engeland. Verder bleek de bijdrage in alle gevallen significant tussen scholen te variëren. Dit verschilde echter sterk tussen landen. Voor Engeland en Singapore werd geconcludeerd dat in zeker 20% van de scholen het onderwijs geen positieve bijdrage leverde aan de ontwikkeling van leerlingen in de jaargroepen die in de analyses waren betrokken.

4 Complicaties bij toepassing in Nederland

In Nederland wordt leeftijd als selectie-criterium bij de toewijzing van leerlingen aan jaargroepen flexibel toegepast. Minder goed presterende leerlingen lopen vaak vertraging op en komen in een lagere groep terecht dan waar men ze gezien hun leeftijd zou verwachten. Het komt ook voor, zij het minder frequent, dat goed presterende leerlingen een groep overslaan. Bovendien lopen kinderen

die op latere leeftijd naar Nederland komen veelal vertraging op in hun schoolloopbaan. Deze groep vormt overigens slechts een kleine minderheid onder de vertraagde leerlingen (Reezigt & Knuver, 1995). Vooral het hoge percentage zittenblijvers leidt tot complicaties bij toepassing van onderbroken regressie.

Reezigt en Knuver (1995) hebben diverse studies naar omvang, oorzaken en effecten van zittenblijven in het basisonderwijs besproken. Zij concludeerden dat het percentage doubleurs circa 2% per jaar bedroeg. Dit zou een cumulatief percentage van 16% aan het eind van het basisonderwijs impliceren, maar sommige studies hebben nog hogere schattingen opgeleverd. Als een leerling blijft zitten, is dit doorgaans vanwege slechte leerprestaties. Voor leerlingen die in juli, augustus of september zijn geboren is de kans op zittenblijven duidelijk groter dan voor leerlingen die later in het schooljaar geboren zijn. Van duurzame effecten op de leerprestaties is nauwelijks sprake. Soms hebben leerlingen kort na het zittenblijven een voorsprong op hun klasgenoten, maar binnen een jaar is deze voorsprong doorgaans weer veranderd in een achterstand (Reezigt & Knuver, 1995; p. 117).

Omdat juist de slecht presterende leerlingen die in de laatste maanden van het schooljaar geboren zijn de meeste kans lopen op een vertraagde schoolloopbaan, leidt het verwijderen van deze leerlingen uit het databestand onvermijdelijk tot vertekeningen. Het gevolg zal zijn dat het effect van leeftijd op leerprestaties zwakker lijkt dan het in werkelijkheid is. Veel jonge leerlingen die minder goed presteren, worden dan immers niet meegeteld. Aangezien het percentage zittenblijvers veel hoger is dan 5%, gaat het om een aanzienlijke vertekening. Het gevolg daarvan is dat het effect van een jaar onderwijs – de onderbreking in de lijn die de relatie tussen leeftijd en leerprestaties weergeeft – eveneens vertekend wordt. Zoals eerder vermeld is het bij toepassing van onderbroken regressie juist van cruciaal belang dat de relatie tussen het selectie-criterium en de experimentele variabele correct gemodelleerd wordt. Indien men de analyse uitsluitend zou betrekken op de leerlingen met een standaard loopbaan, is dat niet het geval (vgl. Swan-

born, 1999; p. 270). Idealiter zou men in de data-analyses willen controleren voor de criteria die gehanteerd worden bij de beslissing om een leerling al dan niet te laten doubleren. Helaas zijn deze criteria onvoldoende doorzichtig.

Dit probleem is te ondervangen als men een realistische inschatting maakt van het prestatieniveau van de zittenblijvers indien ze wel waren overgegaan. Hierbij kunnen we gebruik maken van het gegeven dat het effect van zittenblijven op leerprestaties waarschijnlijk verwaarloosbaar is. Leerlingen blijven zitten omdat hun prestaties duidelijk lager liggen dan gemiddeld, maar na verloop van tijd komen ze in de nieuwe jaargroep weer op hetzelfde niveau terecht. In deze studie is nagegaan wat het effect van leeftijd zou zijn als we ervan uitgaan dat dit inderdaad het geval is. Daarnaast is berekend in hoeverre de resultaten veranderen als men aanneemt dat het effect van zittenblijven hiervan in positieve of negatieve zin afwijkt. Er wordt een aantal situaties gesimuleerd waarin de zittenblijvers wel zijn overgegaan. Zodoende wordt het mogelijk om met onderbroken regressie het effect van een jaar onderwijs te schatten.

De studie waarvan hier verslag wordt gedaan betreft leerlingen in de groepen 5 en 6. Over het prestatieniveau van de leerlingen die vertraging hebben opgelopen maar gezien hun leeftijd in groep 5 hadden moeten zitten, is niets bekend. Het prestatieniveau van de vertraagde leerlingen die eigenlijk in groep 6 en 7 hadden moeten zitten, is wel bekend. Deze zitten immers in groep 5 en 6. Hun prestatieniveau wordt gebruikt om een schatting te maken van het niveau dat de vertraagde leerlingen, die nu in groep 4 en 5 zitten, behaald zouden hebben, als ze "gewoon" in groep 5 en 6 zouden zitten. Als men aanneemt dat zittenblijvers na verloop van tijd toch weer op hetzelfde prestatieniveau terechtkomen, kan men ervan uitgaan dat het prestatieniveau van de vertraagde leerlingen in groep 5 overeenkomt met het niveau dat de vertraagde leerlingen die in groep 4 zitten behaald zouden hebben als ze geen vertraging hadden opgelopen. Concreet betekent dit dat aan de vertraagde leerlingen in groep 5 een virtuele leeftijd is toegekend die precies één

jaar lager ligt dan hun werkelijke leeftijd. Dezelfde operatie is uitgevoerd in groep 6. Daarnaast zijn nog vier andere scenario's in overweging genomen. In deze gevallen zijn aan de zittenblijvers – naast een virtuele leeftijd – juist iets hogere of lagere scores toegekend dan in werkelijkheid behaald. Op die manier zijn situaties gesimuleerd waarin zittenblijven een zwak of gematigd effect in positieve dan wel negatieve zin uitoefent op de leerprestaties. Leerlingen met een versnelde loopbaan of een sterk vertraagde (meer dan een jaar vertraging) zijn niet in de analyses opgenomen. Gezien de beperkte omvang van deze groep (nog geen 4% van het totaal) zal verwijdering niet tot grote vertekeningen leiden.

5 Modelleren van het onderwijs-effect

Vergelijking (1) geeft het basismodel weer. De coëfficiënten β_1 en β_2 hebben betrekking op het effect van leeftijd en het effect van een jaar onderwijs. Het model is van toepassing op een dataset met leerlingen uit twee opeenvolgende jaargroepen.

$$Y_i = \beta_0 + \beta_1(x_i - x_0) + \beta_2 z_i + R_i \quad (1)$$

Waarbij:

- Y_i = score op de afhankelijke variabele (in deze studie: de wiskundescore voor leerling i)
- x_i = leeftijd van leerling i
- x_0 = leeftijdsgrens (in deze studie: 9 jaar en 7 maanden op het moment van toetsafname)
- z_i = jaargroep van leerling i (in deze studie: 0 indien groep 5; 1 indien groep 6)
- β_0 = parameter voor het intercept van de controlegroep op de datumgrens
- β_1 = effect van leeftijd
- β_2 = effect van een jaar onderwijs
- R_i = random residu

Parameter β_1 geeft aan hoe sterk de lijn stijgt die de relatie tussen leeftijd en leerprestaties weergeeft (zie Figuur 1). Parameter β_2 geeft

aan hoe groot het verschil is tussen de oudste leerlingen in groep 5 en de jongste in groep 6. Het niveau van de oudste leerlingen in groep 5 wordt weergegeven met parameter β_0 . Het basismodel zoals weergegeven in vergelijking (1) kan worden uitgebreid met kwadratische en hogere-machtstermen voor het effect van leeftijd. Zo kan worden nagegaan of er sprake is van een kromlijng verband. Het is ook mogelijk dat de relatie tussen leeftijd en leerprestaties verschilt tussen beide leerjaren. Dit kan men modelleren door het basismodel uit te breiden met extra interactietermen (leeftijd * leerjaar).

Als we de uitgangspunten van onderbroken regressie toepassen in een meerniveau model kan het intercept (β_0) en het effect van een jaar onderwijs (β_2) variëren tussen scholen :

$$Y_{ij} = \beta_{0j} + \beta_1(x_{ij} - x_0) + \beta_2 z_{ij} + R_{ij} \quad (2)$$

Waarbij:

- J = index voor scholen
- i = index voor leerlingen binnen scholen

Het intercept en het onderwijseffect zijn nu schoolspecifiek. Deze schoolspecifieke coëfficiënten kunnen uiteengelegd worden in een gemiddeld effect en de schoolspecifieke afwijking:

$$\begin{aligned} \beta_{0j} &= \gamma_{00} + U_{0j} \\ \beta_{2j} &= \gamma_{20} + U_{2j} \end{aligned}$$

Substitutie levert het volgende model op:

$$\begin{aligned} Y_{ij} &= \gamma_{00} + \beta_{10}(x_{ij} - x_0) + \gamma_{20} z_{ij} + \\ &U_{0j} + U_{2j} z_{ij} + R_{ij} \end{aligned} \quad (3)$$

In deze vergelijking geeft γ_{20} het algemene effect weer van een jaar onderwijs, terwijl U_{2j} de schoolspecifieke afwijking weergeeft. De variantie hiervan is bijzonder interessant, omdat deze aangeeft in hoeverre het onderwijseffect varieert tussen scholen. Wanneer dit model gefit wordt, worden de varianties van R_{ij} en U_{0j} geschat, evenals de covariantie tussen U_{0j} and U_{2j} .

Om vast te stellen of het effect van een jaar onderwijs samenhangt met een derde

variabele (bijvoorbeeld het opleidingsniveau van de ouders), kan men een interactie-effect van deze variabele met het onderwijseffect in het model opnemen. Zie Luyten (2006) voor een voorbeeld.

6 Onderzoeksvragen

In dit artikel worden de resultaten besproken die betrekking hebben op analyses met de leerprestaties voor wiskunde in groep 5 en 6 van het basisonderwijs als afhankelijke variabele. De analyses hebben betrekking op de volgende onderzoeksvragen:

1. Wat is het effect van een jaar onderwijs op de leerprestaties? en
2. In hoeverre varieert het effect van een jaar onderwijs tussen scholen?

Bij de beantwoording van de onderzoeksvragen zal specifiek besproken worden hoe de bevindingen voor Nederland zich verhouden tot de resultaten die voor acht andere landen zijn gerapporteerd (Luyten, 2006). Tevens wordt nagegaan in hoeverre de resultaten afhankelijk zijn van de aannames die men maakt over het effect van zitten-blijven.

7 Methode

7.1 Data

De analyses hebben betrekking op data uit TIMSS-95 van leerlingen in groep 5 en 6 van het basisonderwijs. Voor de Nederlandse dataset geldt dat het percentage leerlingen met een standaard schoolloopbaan relatief klein is (circa 70%, zie Tabel 1). Verder valt op dat van veel leerlingen de geboortedatum onbekend is. De analyses hebben uitsluitend betrekking op de normaalvorderende leerlingen en de leerlingen die één jaar vertraging hebben opgelopen. Dit betreft 4.491 leerlingen verspreid over 131 scholen (2.262 in groep 5; 2.229 in groep 6). Van deze leerlingen heeft bijna 17% vertraging opgelopen. De gemiddelde wiskundescore van deze 4.491 leerlingen ligt iets boven het gemiddelde in de oorspronkelijke dataset. Opvallend is het iets lagere percentage vertraagde leerlingen in groep 6, terwijl men een cumulatieve

Tabel 1

Gemiddelde scores en correlaties met leeftijd naar type schoolloopbaan

Loopbaan	Scores wiskunde		Percentages loopbanen		Correlatie met leeftijd	
	Groep 5	Groep 6	Groep 5	Groep 6	Groep 5	Groep 6
Standaard	500,71	587,08	70,7%	69,3%	,055	,004
Vertraagd (1 jaar)	472,69	548,83	14,4%	13,6%	-,095	-,125
Vertraagd (2 jaar)	442,83	527,40	0,7%	0,9%	-,410	-,012
Versneld	509,23	592,77	2,4%	2,6%	-,176	-,033
Onbekend	470,16	551,46	11,9%	13,6%	---	---
Totaal						
Gemiddelde	492,88	576,66	---	---	---	---
Standaarddeviatie	64,89	70,54	---	---	---	---
Standaard loopbanen en 1 jr. vertraagde loopbanen						
Gemiddelde	496,01	580,61	---	---	---	---
standaarddeviatie	64,73	70,87	---	---	---	---

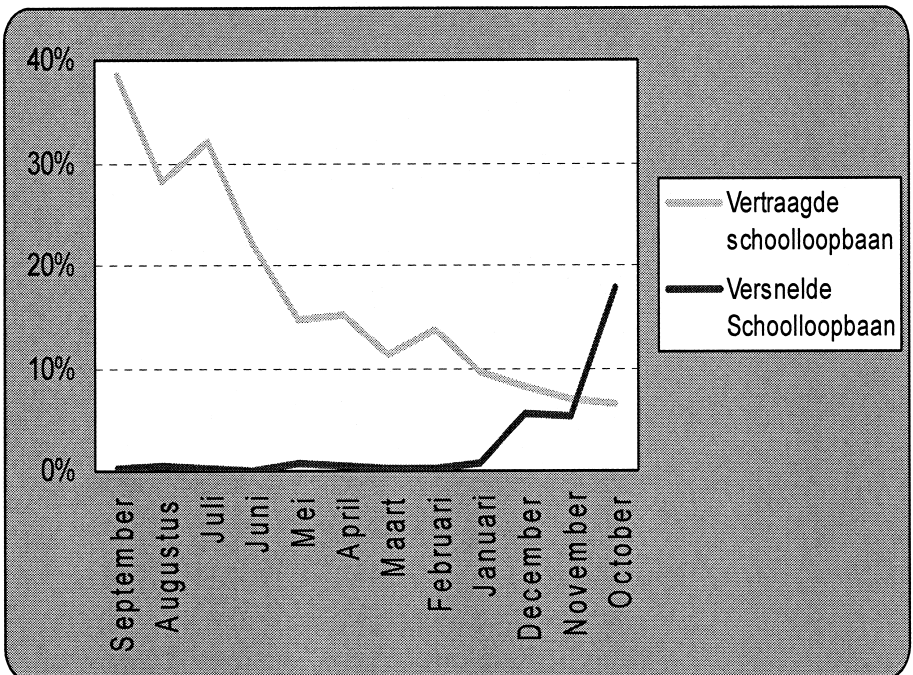
zou verwachten. Kennelijk bevat groep 5 in deze dataset relatief veel zittenblijvers en groep 6 relatief weinig.

De kans op een afwijkende schoolloopbaan hangt sterk samen met de geboortedatum van een leerling. In Figuur 2 wordt dit grafisch weergegeven. Leerlingen geboren in september lopen de meeste kans (38,5%) op vertraging. Voor leerlingen geboren in oktober is dit slechts 6,5%. Versnelde loopbanen zijn zeer zeldzaam bij leerlingen geboren na december (minder dan 1%). Bij leerlingen

geboren in de maanden november en december ligt het percentage versnelde loopbanen op ruim 5% en voor degenen die in oktober zijn geboren, is dit bijna 18%.

7.2 Variabelen

De leerlingenbestanden bevatten verschillende typen toetsscores voor wiskunde. In de analyses is gebruik gemaakt van de 'international proficiency score'. De 'user guide' voor de internationale database van TIMSS (Gonzales & Smith, 1997) beveelt het ge-

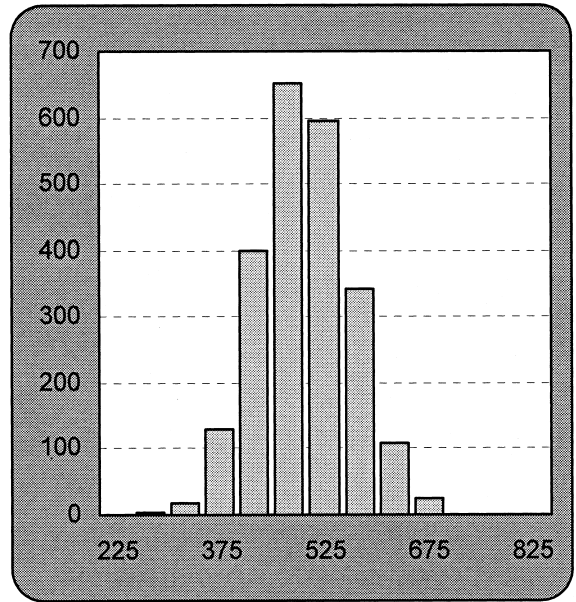


Figuur 2. Percentage vertraagde en versnelde schoolloopbanen per geboortemaand

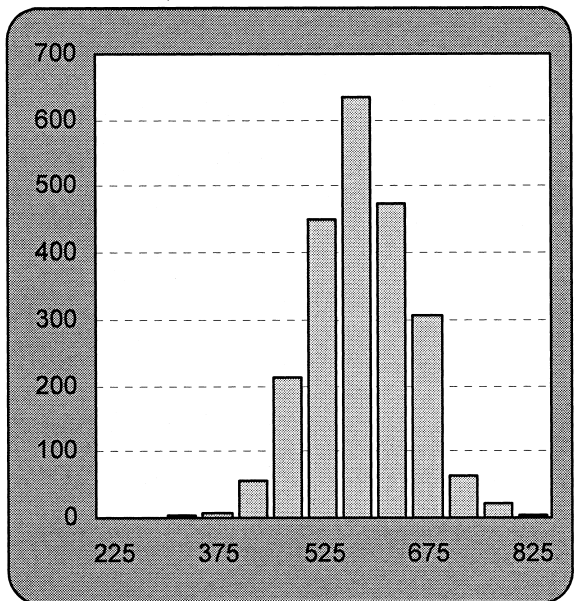
bruik van deze variabele aan voor zowel internationale als nationale studies. Hun frequentieverdelingen in groep 5 en 6 worden in Figuur 3a en 3b weergegeven. In beide groepen wordt de standaardnormale verdeling dicht benaderd. Er zijn geen aanwijzingen voor vloer- of plafondeffecten. De leerlingen in groep 5 en 6 hebben dezelfde toetsen gemaakt.

In Tabel 1 zijn de scores uitgesplitst naar leerjaar en type schoolloopbaan. De tabel rapporteert ook de correlaties tussen leeftijd en leerprestaties naar jaargroep en loopbaan-type. Voor leerlingen met een standaard schoolloopbaan is de correlatie positief maar (zeer) zwak, terwijl deze voor de overige leerlingen negatief is. Binnen de groep met een afwijkende loopbaan worden de laagste scores behaald door de oudste leerlingen. De meeste plausibele verklaring is dat leerlingen met een geboortedatum dicht in de buurt van 1 oktober bij gelijke prestaties eerder blijven zitten of een klas overslaan dan leerlingen met een geboortedatum die verder van de datumgrens verwijderd is. Hieruit blijkt dat corrigeren voor leeftijd op zich niet volstaat als men in Nederland het effect van een jaar onderwijs wil vaststellen. Het komt er op aan een realistische inschatting te maken van de relatie tussen leeftijd en leerprestaties indien leeftijd het enige selectie criterium zou zijn.

Jaargroep en leeftijd zijn de enige onafhankelijke variabelen in deze studie. De variabele die de jaargroep van een leerling aangeeft, is hergecodeerd, zodat leerlingen in groep 5 de score 0 krijgen en in groep 6 de score 1. De variabele die de leeftijd van een leerling aanduidt, is gebaseerd op geboortjaar en -maand. Elke leeftijd is omgezet in een decimaal getal. Zo is aan een leerling geboren in mei 1985 de score 10,00 toegekend en aan een leerling geboren in februari 1985 de score 9,75. Vervolgens is van deze scores de grenswaarde afgetrokken, zodat de oudste leerlingen in groep 5 (de "controlegroep") de score 0 kregen. De grenswaarde is 9,58, wat correspondeert met een geboortedatum in oktober 1985. Normaalvorderende leerlingen geboren in oktober 1985 of later zitten in groep 5, terwijl degenen die eerder zijn geboren in groep 6 zitten. Tabel 2 laat zien hoe geboortedata zijn omgezet in de scores zoals



Figuur 3a. Frequentieverdeling wiskundescores groep 5



Figuur 3b. Frequentieverdeling wiskundescores groep 6

gebruikt in de data-analyses. Het gaat hier om normaalvorderende leerlingen. In de analyses waarover gerapporteerd wordt, zijn de leeftijden van leerlingen met een vertraagde schoolloopbaan aangepast.

Tabel 2

Omzetting geboortedata naar leeftijd op moment van toetsafname

	Jaar en maand	Leeftijd in decimalen	Minus grenswaarde (9,58)
Groep 6	Oktober 1984	10,58 (10 + 7/12)	1,00
	November 1984	10,50 (10 + 6/12)	,92
	December 1984	10,42 (10 + 5/12)	,83
	Januari 1985	10,33 (10 + 4/12)	,75
	Februari 1985	10,25 (10 + 3/12)	,67
	Maart 1985	10,17 (10 + 2/12)	,58
	April 1985	10,08 (10 + 1/12)	,50
	Mei 1985	10,00 (10 + 0/12)	,42
	Juni 1985	9,92 (9 + 11/12)	,33
	Juli 1985	9,83 (9 + 10/12)	,25
	Augustus 1985	9,75 (9 + 9/12)	,17
September 1985	9,67 (9 + 8/12)	,08	
Groep 5	Oktober 1985	9,58 (9 + 7/12)	,00
	November 1985	9,50 (9 + 6/12)	-,08
	December 1985	9,42 (9 + 5/12)	-,17
	Januari 1986	9,33 (9 + 4/12)	-,25
	Februari 1986	9,25 (9 + 3/12)	-,33
	Maart 1986	9,17 (9 + 2/12)	-,42
	April 1986	9,08 (9 + 1/12)	-,50
	Mei 1986	9,00 (9 + 0/12)	-,58
	Juni 1986	8,92 (8 + 11/12)	-,67
	Juli 1986	8,83 (8 + 10/12)	-,75
	Augustus 1986	8,75 (8 + 9/12)	-,83
September 1986	8,67 (8 + 8/12)	-,92	

7.3 Analyses

Met meerniveau-analyse werden vijf modellen gefit. Het eerste model diende als vergelijkingsbasis. Hiermee werd het verschil in toetsscores tussen de leerlingen in groep 5 en 6 geschat en de mate waarin dit verschil tussen scholen varieerde. In de volgende drie modellen werd ook het effect van leeftijd opgenomen. In het eerste hiervan is het leeftijd-effect als een lineaire functie gemodelleerd. Dit model correspondeert met vergelijking (3). In de volgende twee modellen werden eveneens kwadratische en derdemachts-termen opgenomen. Zodoende is nagegaan of de relatie tussen leerprestaties en leeftijd gemodelleerd kan worden als een tweede- of derdegraads functie. De uitkomsten geven aan wat het effect van een jaar onderwijs is na controle voor het effect van leeftijd. In het vijfde model werd een interactie-effect opgenomen van leeftijd met leerjaar. Zodoende is nagegaan of relatie tussen leerprestaties en leeftijd verschilt tussen beide jaargroepen. De randomeffecten in de modellen geven aan in hoeverre het effect van een jaar onderwijs varieerde tussen scholen.

Elk van deze modellen is voor vijf verschillende scenario's doorgerekend. Elk scenario veronderstelt een ander effect van zit-

tenblijven op de leerprestaties. In het eerste scenario is uitgegaan van een "neutraal" effect, waarmee bedoeld wordt dat doubleurs in hun nieuwe jaargroep weer op hetzelfde relatieve niveau terechtkomen. Volgens Reezigt en Knuver (1995) komt dit effect naar voren uit de meeste studies naar het effect van zittenblijven. De benaming *neutraal* verwijst in dit verband naar de relatieve positie van de zittenblijvers in hun jaargroep. Die verandert uiteindelijk niet. Als gevolg van het doubleren, lopen de leerlingen natuurlijk wel een jaar achterstand op. Laten we als voorbeeld twee leerlingen nemen: een zittenblijver (Piet) die in groep 5 bij de laagste 10% behoorde en in zijn nieuwe groep 5 uiteindelijk op hetzelfde relatieve niveau terechtkomt. Voor Piet is het effect van een jaar zittenblijven "neutraal". In absolute zin is zijn prestatieniveau echter veel lager dan dat van een andere leerling (Hein) die in groep 5 op hetzelfde niveau zat als Piet, maar niet is blijven zitten. Hein zit nu in groep 6 bij de laagste 10%. Om dit scenario te simuleren is aan de zittenblijvers een virtuele leeftijd toegekend die precies een jaar lager ligt dan hun werkelijke leeftijd. In de andere vier scenario's worden "positieve" (zwak en gematigd) dan wel "negatieve" (zwak en gematigd)

Tabel 3

Gemiddelden vertraagde en normaalvorderende leerlingen in de vijf scenario's

		VERONDERSTELD EFFECT VAN ZITTENBLIJVEN				
		"neutraal"	zwak "positief"	zwak "negatief"	matig "positief"	matig "negatief"
GROEP 6	normaalvorderend	587,08	587,08	587,08	587,08	587,08
	vertraagd	548,83	535,83	561,83	516,33	581,33
GROEP 5	normaalvorderend	500,71	500,71	500,71	500,71	500,71
	vertraagd	472,69	459,69	485,69	440,19	505,19

Noot. De cijfers voor de vertraagde leerlingen geven aan wat hun verwacht prestatieniveau zou zijn als zittenblijven niet voor zou komen in het Nederlandse basisonderwijs.

effecten van doubleren verondersteld.

Met een "positief" effect wordt bedoeld dat de relatieve positie van de zittenblijvers in hun nieuwe jaargroep iets verbetert. Als deze leerlingen geen jaar hadden overgedaan zou hun relatieve positie iets slechter zijn geweest. Met een "negatief" effect wordt bedoeld dat de relatieve positie van de zittenblijvers in hun nieuwe jaargroep nog slechter is dan in hun oorspronkelijke jaargroep. De kwalificaties zwak en gematigd zijn ontleend aan Cohen (1988; pp. 19-27) en worden afgemeten aan de standaarddeviatie in groep 5. Een zwak effect houdt in dat de score met 0,2 standaarddeviatie omhoog dan wel omlaag gaat bij doubleren. Bij een gematigd effect is dit 0,5 standaarddeviatie. Het lijkt onwaarschijnlijk dat de effecten van zittenblijven in werkelijkheid sterker zijn. Volgens eerder onderzoek is een "neutraal" effect het meest waarschijnlijk.

De verschillende scenario's zijn gesimuleerd door aan de vertraagde leerlingen verlaagde dan wel verhoogde scores toe te kennen. In het "neutrale" scenario is alleen de leeftijd van deze leerlingen aangepast. In de scenario's met een zwak effect zijn daarnaast hun scores met 13 punten verhoogd dan wel verlaagd. In de gematigde scenario's zijn dat 32,5 punten. Deze cijfers komen overeen met respectievelijk 0,2 en 0,5 maal de standaarddeviatie in groep 5 (zie Tabel 1). Tabel 3 geeft een overzicht van de gemiddelde scores die zijn toegekend aan de vertraagde leerlingen in de diverse scenario's. Per scenario wordt aangegeven welk prestatieniveau bij de zittenblijvers verwacht wordt als doubleren niet aan de orde zou zijn. Het "neutrale" scenario veronderstelt dat zittenblijven geen verschil maakt voor de relatieve positie van een leerling in de jaargroep. Een positief effect be-

tekent dat bij doubleren de relatieve positie verbetert. Dat impliceert dat het relatieve prestatieniveau in de oorspronkelijke jaargroep van de vertraagde leerlingen lager zou zijn geweest, als ze wel waren overgegaan. Een negatief effect houdt in dat het relatieve prestatieniveau verslechtert. In dat scenario zouden de leerlingen een hoger prestatieniveau hebben behaald in de oorspronkelijke groep.

Omdat in Nederland een gestratificeerde steekproef (naar grootte, gemiddeld OVB-gewicht¹, denominatie, urbanisatiegraad) is getrokken van scholen (Knuver, Doolaard, & Matthijsen, 1997), is in de analyses is door middel van weging rekening gehouden met het steekproefontwerp. Zodoende wordt voorkomen dat uitkomsten vertekend worden door onder- of oververtegenwoordiging van bepaalde groepen. De gewichten zijn zodanig berekend dat hun som gelijk is aan de omvang van de steekproef, zodat standaardfouten en significantieniveaus correct geschat worden. Daarnaast wordt met de weging ook een correctie beoogd voor de vertekeningen die door selectieve non-respons zijn ontstaan. Gonzales en Smith (1997) verschaffen nadere details over de constructie van de gewichten.

8 Resultaten

De resultaten van de analyses worden weergegeven in de Tabellen 4 en 5. Deze tabellen laten de voornaamste output zien die men verkrijgt met behulp van de MLwiN-software (Rashbash, Browne, Goldstein, Yang, Plewis, Healy, Woodhouse, Draper, Langford, & Lewis, 2000). In de volgende paragraaf worden de uitkomsten en hun implicaties nader besproken. Eerst komen enkele

algemene conclusies die eenvoudig uit de tabellen afgeleid kunnen worden aan bod.

Tabel 4 geeft per scenario de vaste effecten weer die geschat worden als men de vijf modellen fit. De vaste leerjareffecten in model 1 geven een schatting van het verschil in leerprestaties tussen groep 5 en 6. Deze schattingen zijn nagenoeg gelijk in alle vijf scenario's. In model 2 wordt het lineaire effect van leeftijd toegevoegd. De uitkomsten laten duidelijk zien dat het geschatte effect sterk afhankelijk is van de aannames met betrekking tot het effect van zittenblijven. Als men een matig "negatief" effect veronderstelt is het leeftijds-effect nagenoeg gelijk aan nul. Hoe "positiever" het veronderstelde effect van zittenblijven des te groter het geschatte leeftijdeffect. Naarmate het leeftijdeffect toeneemt, neemt het effect van een jaar extra onderwijs af. Als men een matig "negatief" effect van zittenblijven veronderstelt, komt het leerjareffect uit op 83,63. Een matig "positief" effect levert een

leerjareffect op van 64,38. In alle scenario's is het leerjareffect echter significant.

Toevoeging van een kwadratische term voor het leeftijdeffect levert geen verbetering op. Alle kwadratische effecten zijn kleiner dan de corresponderende standaardfouten. De derdemachtstermen die in model 4 worden toegevoegd, zijn wel allemaal groter dan de standaardfout, maar geen van alle in voldoende mate om als statistisch significant te worden beschouwd. De meest extreme t-waarde bedraagt -1,64. Ook bij een eenzijdige toetsing is dit niet significant voor $\alpha = 0,05$. De interactietermen die in model 5 zijn toegevoegd, laten geen van allen een significant effect zien. Er is dus geen reden om aan te nemen dat het leeftijdeffect verschilt tussen jaargroepen. De toevoegingen in de modellen 3, 4 en 5 leveren weinig verbetering op ten opzichte van model 2, dat alleen een lineair leeftijdeffect bevat.

In Tabel 5 worden de varianties van de

Tabel 4

Vaste effecten onder verschillende aannames m.b.t. effect van zittenblijven

	VERONDERSTELD EFFECT VAN ZITTENBLIJVEN									
	"neutraal"		zwak "positief"		zwak "negatief"		matig "positief"		matig "negatief"	
	effect	s.e.	effect	s.e.	effect	s.e.	effect	s.e.	effect	s.e.
Model 1:										
Uitsluitend leerjaar effect										
Intercept	493,83	2,99	491,61	3,03	496,06	2,96	488,28	3,09	499,41	2,93
Leerjaar	84,55	3,00	84,66	3,05	84,43	2,95	84,83	3,14	84,25	2,91
Model 2:										
Uitbreiding met lineair leeftijdeffect										
Intercept	498,62	3,37	498,20	3,41	499,04	3,35	497,60	3,47	499,70	3,32
Leerjaar	74,03	4,88	70,17	4,94	77,87	4,83	64,38	5,08	83,63	4,81
Leeftijd	10,61	3,90	14,61	3,95	6,61	3,88	20,61	4,06	,63	3,89
Model 3:										
Uitbreiding met kwadratische term leeftijdeffect										
Intercept	499,20	3,50	498,95	3,52	499,47	3,48	498,58	3,57	499,89	3,47
Leerjaar	74,03	4,88	70,18	4,95	77,88	4,84	64,40	5,09	83,63	4,81
Leeftijd	10,75	3,87	14,79	3,91	6,72	3,85	20,86	4,02	,68	3,87
Leeftijd ²	-1,81	3,42	-2,30	3,46	-1,33	3,40	-3,05	3,56	-,61	3,39
Model 4:										
Uitbreiding met derdemachtsterm leeftijdeffect										
Intercept	501,56	3,79	501,33	3,82	501,80	3,77	501,00	3,88	502,17	3,76
Leerjaar	68,20	5,44	64,28	5,55	72,11	5,36	58,39	5,78	77,98	5,30
Leeftijd	23,83	8,30	28,03	8,38	19,64	8,28	34,31	8,60	13,35	8,36
Leeftijd ²	,05	3,45	-,42	3,49	,51	3,44	-1,14	3,59	1,19	3,46
Leeftijd ³	-14,86	9,11	-15,04	9,15	-14,68	9,14	-15,29	9,33	-14,39	9,29
Model 5:										
Interactie-effect leeftijd en leerjaar										
Intercept	499,08	3,73	498,74	3,75	499,44	3,71	498,24	3,80	499,99	3,71
Leerjaar	74,12	4,91	70,28	4,98	77,95	4,86	64,51	5,13	83,69	4,83
Leeftijd	11,63	5,00	15,79	5,05	7,48	4,99	22,04	5,18	1,29	5,04
Lft.*Leerj.	-2,06	7,15	-2,37	7,25	-1,75	7,08	-2,87	7,47	-1,32	7,06

randomeffecten gerapporteerd die model 2 oplevert. De varianties in de modellen 3, 4 en 5 wijken hier nauwelijks van af. De verschillen tussen de vijf scenario's zijn iets groter maar nog altijd beperkt in vergelijking tot de corresponderende standaardfouten. Het meest interessant is de mate waarin het leerjareffect (na controle voor het leeftijdseffect) varieert tussen scholen. Deze variantie is in alle scenario's ook bij strenge criteria significant (voor $\alpha = 0,001$ bij een tweezijdige toetsing). Dit impliceert dat het effect van een jaar extra onderwijs van school tot school verschilt. Normaal gesproken wordt in onderzoek naar schooleffectiviteit het schooleffect afgemeten aan de variantie van het intercept op schoolniveau. Deze geeft aan in hoeverre het *niveau* van de leerprestaties tussen scholen verschilt. Door zo veel mogelijk rekening te houden met achtergrondkenmerken en eerder behaalde leerprestaties wordt dan getracht een zo eerlijk mogelijke vergelijking te maken. In de analyses waarvan hier verslag wordt gedaan, is het effect van een jaar onderwijs geschat. Vervolgens is nagegaan in hoeverre scholen in dit opzicht van elkaar verschillen.

De negatieve covarianties geven aan dat in scholen waar het prestatieniveau in groep 5 hoog ligt, het leerjareffect relatief laag is. Aangezien de frequentieverdelingen van de scores in beide jaargroepen de standaardnormale verdeling dicht benaderen (zie Figuur 3a en 3b), kunnen de negatieve covarianties niet worden geweten aan een plafondeffect in de wiskundetoets. In het matig "positieve" scenario is de covariantie overigens niet significant voor $\alpha = 0,05$ bij een tweezijdige toetsing (t-waarde = $-1,88$).

8.1 Antwoorden op de onderzoeksvragen

Het effect van een jaar onderwijs op leerprestaties

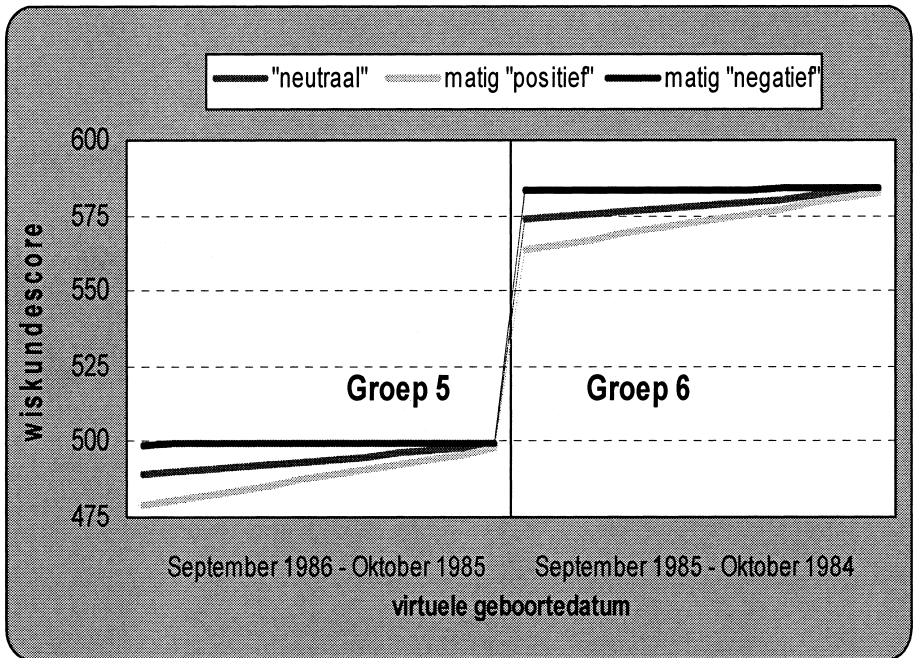
Als men controleert voor een lineair effect van leeftijd, loopt het geschatte effect van een jaar onderwijs – afhankelijk van het veronderstelde effect van zittenblijven – uiteen van ruim 64 tot bijna 84 punten (zie Tabel 4, model 2). In de scenario's die veronderstellen dat doubleren een "positief" effect heeft is het geschatte effect van een jaar onderwijs kleiner. Gaat men uit van een "negatief" effect van zittenblijven, dan wordt het leerjareffect juist groter. Figuur 4 biedt een grafische weergave van de resultaten en geeft de patronen weer voor drie van de vijf scenario's. Als men uitgaat van een matig "negatief" effect van zittenblijven, is het effect van leeftijd op de leerprestaties vrijwel gelijk aan nul. In dat geval is het effect van een jaar extra onderwijs nagenoeg gelijk aan het ruwe verschil tussen de gemiddelden in beide jaargroepen. Figuur 4 laat zien hoe bij een sterker effect van leeftijd op de leerprestaties het effect van een jaar onderwijs afneemt.

Door Cahan en Davis (1987) is het effect van een jaar onderwijs uitgedrukt als een percentage van het ongecorrigeerde verschil tussen twee opeenvolgende jaargroepen. Een dergelijke benadering levert in deze studie percentages die uiteenlopen van 76% (in het scenario met een matig "positief" effect van zittenblijven) tot 99% (bij een matig "negatief" effect). In het "neutrale" scenario ligt het percentage op 88%. In Tabel 6 worden deze percentages vergeleken met de resultaten zoals die zijn gerapporteerd voor de acht landen uit TIMSS-95 waar de onderbroken-

Tabel 5

Varianties van random effecten in model 2

	VERONDERSTELD EFFECT VAN ZITTENBLIJVEN									
	"neutraal"		zwak "positief"		zwak "negatief"		matig "positief"		matig "negatief"	
	effect	s.e.	effect	s.e.	effect	s.e.	effect	s.e.	effect	s.e.
SCHOOLNIVEAU										
Intercept	860,5	146,4	873,0	148,2	849,5	144,8	894,5	151,4	836,1	143,1
Leerjaar	566,1	132,4	582,5	136,4	550,9	129,2	609,0	144,1	529,6	126,0
Covariantie	-189,7	94,7	-189,6	96,8	-190,1	93,1	-189,5	100,9	-190,6	91,8
LEERLINGNIVEAU										
Intercept	3688,0	171,3	3802,5	177,1	3615,9	167,7	4053,7	190,4	3587,2	166,5



Figuur 4. Leeftijd en leerprestaties onder verschillende aannames m.b.t. effect van zittenblijven (model 2)

regressieaanpak niet op complicaties stuit. Ook als men een matig "positief" effect van zittenblijven veronderstelt, is het effect van een jaar onderwijs in Nederland sterker dan in deze acht landen, waar de percentages uiteenlopen van 38% in Engeland tot 75% in Singapore.

Het effect van een jaar onderwijs kan ook worden uitgedrukt in termen van de 'Effect size index' (Cohen, 1988). Deze grootheid (d) wordt berekend door het verschil tussen de experimentele groep en de controle groep te delen door de standaarddeviatie in de con-

Tabel 6

Implicaties van de resultaten en vergelijking met andere landen

	Effect van een jaar onderwijs		Range leerjareffect middelste helft van de scholen	Percentage scholen zonder positief leerjareffect	Correlatie leerjareffect en prestaties lage groep
	percentage van verschil beide groepen	effect size index			
Cyprus	57%	,53	22 – 59	7%	-,24
Engeland	38%	,25	2 – 42	23%	-,33
Griekenland	66%	,53	14 – 76	16%	-,54
IJsland	63%	,58	28 – 50	1%	-,42
Japan	55%	,45	25 – 40	0%	-,18
Noorwegen	72%	,81	38 – 78	2%	-,45
Schotland	55%	,43	15 – 56	12%	-,08
Singapore	75%	,56	11 – 101	20%	-,49
NEDERLAND (afhankelijk van verondersteld effect van zittenblijven)					
"neutraal"	88%	1,14	58 – 90	0%	-,27
zwak "positief"	83%	1,08	54 – 86	0%	-,27
zwak "negatief"	92%	1,20	62 – 94	0%	-,28
matig "positief"	76%	,99	48 – 81	0%	-,26
matig "negatief"	99%	1,29	68 – 99	0%	-,29

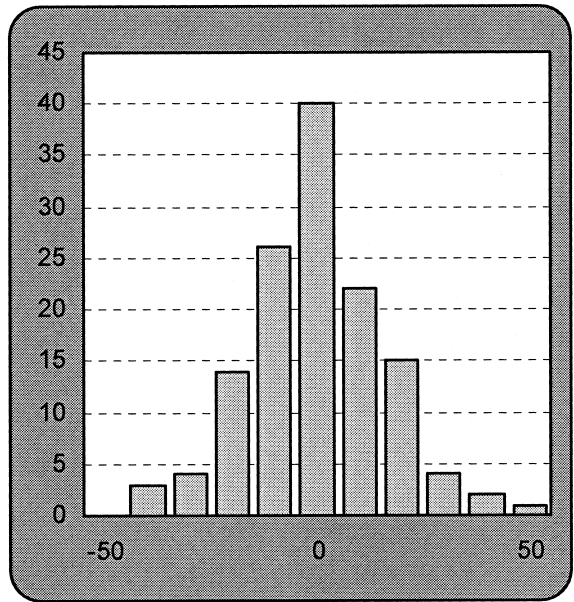
De cijfers voor Nederland zijn gebaseerd op de resultaten gerapporteerd in Tabellen 4 en 5,

trolegroep (in dit geval het laagste leerjaar). In deze studie wordt de 'Effect size index' berekend aan de hand van het verschil na correctie voor het leeftijdeffect (m.a.w. het leerjareffect). In Tabel 6 wordt aangegeven hoe sterk het onderwijseffect is in Nederland en de andere acht landen als men de resultaten op deze wijze beziet. Voor Nederland lopen de waarden uiteen van 0,99 tot 1,29, terwijl in de overige acht landen de uitkomst tussen 0,25 en 0,81 varieert. Deze laatste twee waarden corresponderen met correlaties van respectievelijk 0,12 en 0,38, terwijl de voor Nederland gevonden waarden corresponderen met correlaties tussen 0,44 en 0,54. Een 'effect size' groter dan 0,80 kan zonder meer als sterk worden aangemerkt. Naast Nederland is Noorwegen het enige andere land met een dusdanig sterk onderwijseffect. Voor bijna alle andere landen ligt de waarde rond de 0,50, wat als een effect van middelmatige omvang kan worden gezien (Cohen, 1988; pp. 19-27). Een dergelijk effect correspondeert met een correlatie van 0,24. Alleen voor Engeland is het effect veel minder sterk.

Eerdere rapportages (Mullis et al., 1997) lieten al zien dat in Nederland het verschil tussen beide groepen hoog is in vergelijking met andere landen. De hier gepresenteerde resultaten tonen aan dat hetzelfde geldt voor het effect van een jaar onderwijs.

Variantie tussen scholen met betrekking tot het effect van een jaar onderwijs

Uit de analyses blijkt dat het effect van een jaar onderwijs significant varieert tussen scholen. De varianties in Tabel 5 zijn minimaal vier keer zo groot als de corresponderende standaardfouten. Om een beter idee te krijgen van de betekenis van deze resultaten zijn de varianties omgerekend naar standaarddeviaties. Deze zijn vervolgens gebruikt om het bereik van het leerjareffect te berekenen voor de middelste helft van de scholen. In een standaardnormale verdeling is de helft van de scores minder dan 0,67 maal de standaarddeviatie van het gemiddelde verwijderd. In Tabel 6 is het bereik voor Nederland en de overige acht landen gerapporteerd². In vergelijking met de meeste andere landen blijken de verschillen tussen scholen in Nederland beperkt. Alleen in IJs-



Figuur 5. Frequentieverdeling residuen leerjareffect

land en Japan is het bereik nog smaller. De bereiken zijn berekend onder de aanname dat de leerjareffecten een standaardnormale verdeling volgen. Op basis van Figuur 5, waarin de verdeling van de residuen van het leerjareffect op schoolniveau wordt weergegeven, kan men concluderen dat dit een gerechtvaardigde assumptie is.

Figuur 6 laat de bereiken voor de middelste helft van de scholen zien in Nederland en de acht andere landen. Voor Nederland is de situatie weergegeven in drie van de vijf scenario's. De bereiken zijn in alle scenario's nagenoeg even groot. De assumpties met betrekking tot het effect van doubleren hebben wel consequenties voor de onder- en bovengrenzen van de bereiken. Uit Tabel 6 blijkt verder dat in sommige landen de ondergrens van het bereik voor de middelste helft van de scholen dicht bij nul ligt. Engeland is hiervan het meest extreme voorbeeld. Dit betekent dat in bijna 25% van de Engelse scholen het effect van een jaar onderwijs op het gebied van wiskunde, zoals gemeten in TIMSS-95, negatief is. In deze scholen halen de oudste leerlingen uit 'year 4' hogere scores dan de jongste leerlingen uit 'year 5'. Tabel 6 vermeldt per land een schatting van het percentage scholen zonder positief effect van een jaar onderwijs.² In Nederland zijn de percen-

tages zeer laag, in alle scenario's namelijk lager zijn dan 0,5%. Japan is het enige andere land met een dergelijk laag percentage.

Tabel 5 laat negatieve covarianties zien tussen het leerjaareffect en het intercept op schoolniveau. Dit impliceert dat het effect van een jaar onderwijs relatief sterk is in scholen waar het niveau van de leerprestaties in groep 5 aan de lage kant is. Ook in de acht overige landen zijn zonder uitzondering dergelijke negatieve covarianties gevonden. Als men deze covarianties omrekent naar correlaties vindt men voor Nederland waarden van $-0,26$ tot $-0,29$. Dit kan men beschouwen als een samenhang van middelmatige sterkte (Cohen, 1988). In de meeste andere landen zijn deze correlaties sterker.

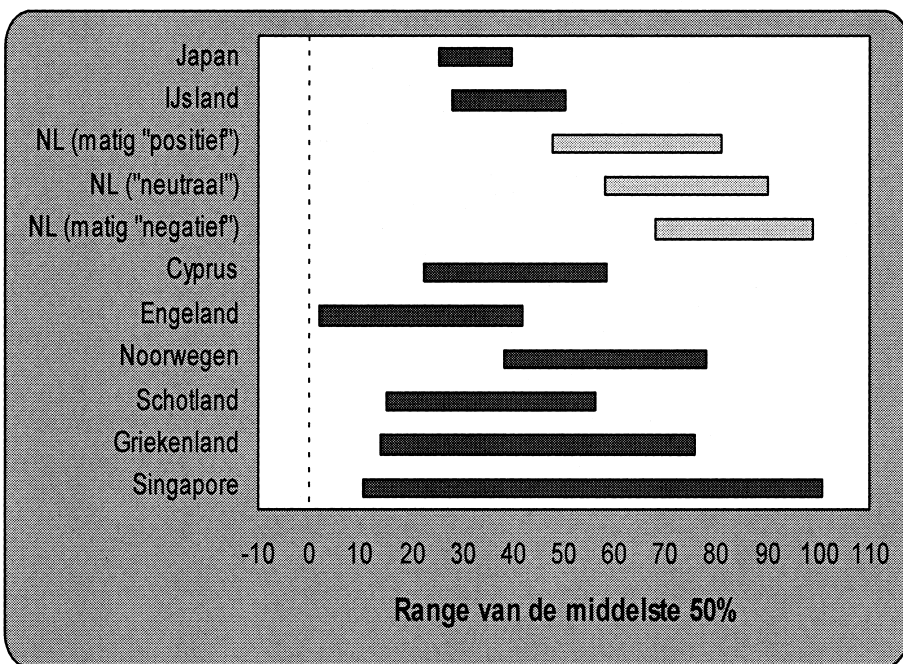
9 Conclusie en discussie

De gepresenteerde onderzoeksresultaten tonen de meerwaarde aan van een meerniveau-analyse gebaseerd op onderbroken regressie voor onderzoek naar schooleffectiviteit. Conceptueel gezien is het grote voordeel van deze aanpak dat de uitkomsten gerelateerd kunnen worden aan een betekenisvol

nulpunt. In de gebruikelijke aanpak kan alleen worden vastgesteld in hoeverre de resultaten op een school (rekening houdend met de achtergrond van de leerlingen en hun eerdere leerprestaties) afwijken van het gemiddelde. Met de hier gepresenteerde aanpak wordt het mogelijk om vast te stellen in hoeverre een jaar onderwijs een bijdrage heeft geleverd aan de ontwikkeling van de leerlingen. Tevens wordt het mogelijk om na te gaan in welke scholen een jaar onderwijs geen positief effect laat zien.

De bevindingen bieden een perspectief op de effectiviteit van onderwijs dat duidelijk afwijkt van wat doorgaans naar voren komt in studies naar schooleffectiviteit. Daar wordt het schooleffect afgemeten aan het percentage variatie in leerprestaties dat kan worden toegeschreven aan het schoolniveau. Omdat de verschillen in prestaties tussen leerlingen binnen een school meestal veel groter zijn dan de verschillen tussen scholen, komt men doorgaans uit op percentages van 10 à 15%. De benadering in dit artikel gaat uit van een ander perspectief.

De effectiviteit van het onderwijs wordt afgemeten aan het verschil in prestatieniveau van leerlingen uit twee opeenvolgende leer-



Figuur 6. Variatie in leerjaareffect tussen scholen

jaren. Vervolgens wordt nagegaan in hoeverre dit verschil kan worden toegeschreven aan het verschil in leeftijd. Als het verschil in leerprestaties alleen een kwestie van leeftijd zou zijn, voegt een jaar onderwijs kennelijk niets toe aan de cognitieve ontwikkeling. In landen waar leeftijd strikt wordt toegepast als criterium om leerlingen aan leerjaren toe te wijzen, kan men op basis van onderbroken regressie op betrekkelijk eenvoudige wijze het effect van een jaar onderwijs vaststellen. Het grote aantal zittenblijvers in Nederland zorgt echter voor complicaties. Het gaat in dergelijke gevallen in ieder geval niet aan om de afwijkende waarnemingen (de vertraagde leerlingen) uit de dataset te verwijderen (Swanborn, 1999; p. 270). In deze studie is de oplossing gezocht in het doorrekenen van een aantal scenario's waarin de leerlingen die in werkelijkheid een jaar zijn blijven zitten, toch zouden zijn overgegaan. Dit vereist realistische assumpties met betrekking tot de effecten van zittenblijven. Op basis van de onderzoeksliteratuur lijkt een "neutraal" effect (zittenblijvers komen in hun nieuwe groep uiteindelijk weer op hetzelfde relatieve niveau terecht) het meest waarschijnlijk. Alternatieve scenario's zijn echter eveneens in overweging genomen.

De analyses laten een sterk effect zien op de wiskunde-prestaties van leerlingen in groep 5 en 6. Uit eerder onderzoek – in het buitenland – blijkt dat doorgaans ruim 50% van het verschil in leerprestaties tussen twee opeenvolgende leerjaren kan worden toegeschreven aan het gevolgde onderwijs. De voor Nederland gevonden percentages variëren van 75% tot 99%. Daarnaast blijkt dat het effect van een jaar onderwijs minder dan in de meeste andere landen tussen scholen verschilt.

Naast voordelen uit conceptueel oogpunt biedt onderbroken regressie ook belangrijke praktische voordelen. In deze aanpak is geen controle voor scores op een pre-test of andere achtergrondvariabelen vereist om een valide schatting te maken van het effect van onderwijs. Evenals in een gerandomiseerd experiment kunnen op basis van cross-sectionele data sterke causale conclusies getrokken worden. Onderzoek naar schooleffectiviteit richt zich in de eerste plaats op verschillen in

leerprestaties tussen scholen. In deze benadering is controle voor achtergrondkenmerken van de leerlingen en vooral prestaties op een eerder tijdstip cruciaal om vertekende resultaten te voorkomen. Als toewijzing aan een controlegroep of een experimentele groep plaatsvindt op basis van toeval, kan een verschil tussen beide groepen alleen worden toegeschreven aan de experimentele variabele of aan toevalsfluctuaties. In het onderbroken-regressieontwerp is het criterium voor toewijzing aan de controlegroep dan wel de experimentele groep bekend en wordt het effect ervan in de data-analyse vastgesteld. Wat er dan nog over blijft aan verschil tussen beide groepen kan met grote waarschijnlijkheid worden toegeschreven aan de experimentele variabele (afgezien van toevalsfluctuaties).

De praktische voordelen van onderzoek op basis van cross-sectionele data zijn overduidelijk. Het kost nu eenmaal veel meer tijd, geld en inspanning om een goede longitudinale dataset aan te leggen. Bovendien heeft men bij het verzamelen van longitudinale data veelal te kampen met selectieve uitval. Desondanks kunnen longitudinale data belangrijke aanvullende informatie bieden in vergelijking met cross-sectionele. Dit artikel laat wel zien dat op basis van cross-sectionele data veel sterkere conclusies getrokken kunnen worden dan doorgaans verondersteld.

Het hoge percentage zittenblijvers vormt wel een serieuze complicatie als men in Nederland met onderbroken regressie het onderwijseffect in kaart wil brengen. Nader onderzoek naar de effecten van zittenblijven kan hier uitkomst brengen. Hoe exacter men de effecten van doubleren kan specificeren des te realistischer de assumpties met betrekking tot het prestatieniveau dat de zittenblijvers behaald zouden hebben indien ze wel waren overgegaan. De kracht van het onderbroken-regressieontwerp is gelegen in het feit dat op een zeer specifiek punt van het leeftijdscontinuüm een effect wordt voorspeld. Daarom is het ook een veel sterker design dan alle overige quasi-experimentele onderzoeksdesigns (samen met onderbroken tijdsreeksen). Alternatieve verklaringen voor het gevonden effect zijn immers moeilijk te vinden. Alleen factoren die exact samenvallen met de leeftijds grens zouden het gevonden effect

eveneens kunnen verklaren. In verband met het leerjaareffect is het voorstelbaar dat de jonge en oude leerlingen binnen een jaargroep anders worden behandeld (door leerkrachten, medeleerlingen of ouders). Nader empirisch onderzoek zal moeten aantonen in hoeverre dergelijk processen de schatting van het onderwijseffect kunnen vertekenen. Daarnaast kan een onjuiste modellering van de relatie tussen leeftijd en leerprestaties de resultaten beïnvloeden. De analyses in deze studie leverden geen overtuigend bewijs voor een kromlijnig verband, maar Tabel 4 laat wel zien dat, indien de relatie tussen leeftijd en leerprestaties als een derdegraadsfunctie wordt gemodelleerd, de schatting van het leerjaareffect verandert. Een ander punt is dat in Nederland veel leerlingen hun schoolloopbaan al starten op de dag dat ze vier jaar worden. Dit betekent dat leerlingen die vroeg in het schooljaar zijn geboren meer (kleuter)onderwijs hebben ontvangen dan de jongere leerlingen. Leuven, Lindahl, Oosterbeek en Webbink (2003) rapporteren dat dit extra onderwijs geen effect heeft voor de meerderheid van de leerlingen, maar wel voor achterstandsleerlingen. Als dit effect in groep 5 en 6 nog doorwerkt, dan heeft dat wellicht tot een lichte overschatting van het leeftijdseffect geleid. Dat zou een lichte onderschatting van het leerjaareffect tot gevolg kunnen hebben. Verder zou men de gevonden discontinuïteit deels kunnen opvatten als een omgevingseffect. Leerlingen die een jaar extra onderwijs hebben genoten, hebben ook een jaar lang bij oudere leerlingen in de klas gezeten. Gezien de manier waarop onderwijs doorgaans wordt georganiseerd, is een dergelijk 'peer group'-effect onlosmakelijk verbonden met het volgen van onderwijs. De sociale interacties binnen een klas dragen op hun eigen wijze bij aan de ontwikkeling van de leerlingen. Met de hier gepresenteerde benadering wordt deze bijdrage niet geïsoleerd van de overige componenten die samen het effect van een jaar onderwijs vormen. Dat is overigens ook alleen noodzakelijk als men van mening is dat een 'peer group'-effect niet zou mogen meetellen als het effect van een jaar onderwijs in kaart wordt gebracht.

Het onderzoek waarvan hier verslag is gedaan heeft betrekking op twee opeenvolgen-

de leerjaren. Het is ook mogelijk om het model uit te breiden en meerdere jaargroepen te vergelijken. In dat geval moeten er ook meerdere leerjaareffecten worden geschat. Voor een nadere bespreking van een model met 'multiple cutoff points' wordt verwezen naar Trochim (1984, pp. 134-135). Schatting van een dergelijk model vereist wel dat men beschikt over testcores over een wijder bereik van jaargroepen, maar die toch op dezelfde schaal betrekking hebben. Dit is mogelijk als men gebruik kan maken van scores die met behulp van itemresponstheorie geëquivalet zijn.

De hier besproken methodologie lijkt ook interessant in verband met onderzoek naar de economische opbrengsten van onderwijs. In de traditie van de 'human capital'-theorie worden onderwijsopbrengsten veelal afgemeten aan het behaalde niveau in termen van behaalde diploma's (Becker, 1993; Schultz, 1962). Met onderbroken regressie is het mogelijk om op betrekkelijk eenvoudige wijze de opbrengst van onderwijs in kaart te brengen op een manier die betrekking heeft op het daadwerkelijk bereikte niveau van kennis en vaardigheden. Met name in internationaal vergelijkend onderzoek is dit van belang aangezien de prestaties die geleverd moeten worden om een diploma te bemachtigen op een bepaald niveau sterk van land tot land kunnen verschillen.

Al met al wordt in deze studie nogmaals het beeld bevestigd dat in het Nederlandse onderwijs goede resultaten behaald worden, zoals reeds in een groot aantal internationaal vergelijkende studies is gebleken (Beaton, Martin, Mullis et al., 1996; Beaton, Mullis, Martin et al., 1996; Martin, Mullis, Beaton et al., 1997; Martin, Mullis, Gonzales et al., 2000; Martin, Mullis, Gonzales, Chrostowski, 2004; Mullis et al., 1997; Mullis et al., 1998; Mullis, Martin, Beaton et al. 2000; Mullis, Martin, Gonzales et al., 2003; Mullis, Martin, Gonzales, et al., 2004; OECD, 2001; OECD, 2004; Postlethwaite & Wiley, 1992; Robitaille & Garden, 1989). Specifiek met betrekking tot TIMSS-95 is uit eerdere rapportages naar voren gekomen dat de leerlingen in het Nederlandse basisonderwijs op een hoog niveau presteren. Uit de hier gepresenteerde analyses blijkt bovendien dat in

Nederland het onderwijseffect groot is in vergelijking met andere landen. De verschillen tussen scholen met betrekking tot het effect van een jaar onderwijs zijn in Nederland relatief klein en scholen die geen positief effect laten zien zijn (voor wiskunde) uiterst zeldzaam.

Noten

- 1 OVB staat voor Onderwijsvoorrangsbeleid
- 2 Het bereik voor de middelste helft van de scholen in Nederland is als volgt berekend: De variantie van het leerjareffect in het "neutrale" scenario bedraagt 566,1. Dit komt overeen met een standaarddeviatie van 23,8. Het vaste effect (het gemiddelde over de scholen) is 74,0. De ondergrens voor de middelste helft van alle scholen is zodoende als volgt vastgesteld: $74,0 - 0,67 \cdot 23,8 = 58,1$. De bovengrens wordt dan: $74,0 + 0,67 \cdot 23,8 = 90,0$.
- 3 Om aan te geven hoe deze percentages berekend zijn, wordt het zwak "positieve" scenario in Nederland als voorbeeld genomen. De standaarddeviatie van het leerjareffect bedraagt 25,1. Het vaste effect is gelijk aan 70,2. Dit betekent dat een nuleffect 2,91 maal de standaarddeviatie onder het gemiddelde ligt. Uitgaande van een standaard normale verdeling betekent dit dat in 0,2% van de scholen het effect van een jaar onderwijs lager is dan nul.

Literatuur

- Beaton, A. E., Martin, M. O., Mullis, I. V. S., Gonzales, E. J., Kelly, D. L., & Smith, T. A. (1996). *Science achievement in the middle school years, IEA's Third International Mathematics and Science Study*. Chestnut Hill, MA, USA: The International Association for the Evaluation of Educational Achievement (IEA).
- Beaton, A. E., Mullis, I. V. S., Martin, M. O., Gonzales, E. J., Kelly, D. L., & Smith, T. A. (1996). *Mathematics achievement in the middle school years, IEA's Third International Mathematics and Science Study*. Chestnut Hill, MA, USA: The International Association for the Evaluation of Educational Achievement (IEA).
- Becker, G. S. (1993). *Human capital: A theoretical*

and empirical analysis, with special reference to education (3rd ed.). Chicago: Chicago University Press.

- Cahan, S., & Davis, D. (1987). A between-grades-level approach to the investigation of the absolute effects of schooling on achievement. *American Educational Research Journal*, 24, 1-12.
- Cohen, J. (1988). *Statistical power analysis for the behavioral science* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates Publishers.
- Coleman, J. S., Campbell, E., Hobson, C., McPartland, J., Mood, A., Weinfeld, F., & York, R. (1966). *Equality of education opportunity*. Washington, D.C.: Government Printing Office.
- Doornbos, K. (1969). *Opstaan tegen het zitten-blijven*. Den Haag: Staatsuitgeverij.
- Gonzalez, E. J., & Smith, T. A. (1997). *User guide for the TIMSS international database. Primary and middle school years*. Chestnut Hill, MA: TIMSS international study center, Boston College.
- Judd, C. M. & Kennedy, D. (1981). *Estimating the effects of social interventions*. New York: Cambridge University Press.
- Knuver, J.W. M., Doolaard, S., & Matthijssen, C. (1997). *Rekenen-wiskunde en natuuronderwijs op de basisschool. Nederlands aandeel in TIMSS populatie 1*. Enschede, Nederland: OCTO/Universiteit Twente.
- Leuven, E., Lindahl, M., Oosterbeek, H. & Web-bink, D. (2003). *The effect of potential time in school on early test scores*. Opgehaald op 3 februari 2006, van <http://ecodip.net/pdf/holiday.pdf>
- Luyten, H. (2006). An empirical assessment of the absolute effect of schooling: regression discontinuity applied to TIMSS-95, *Oxford Review of Education*, 32, 397-429.
- Martin, M. O., Mullis, I. V. S., Beaton, A. E., Gonzales, E., Smith, T. A., & Kelly, D. (1997). *Science achievement in the primary school years: IEA's Third International Mathematics and Science Study (TIMSS)*. Chestnut Hill, MA: TIMSS international study center, Boston College.
- Martin, M. O, Mullis, I. V. S., Gonzales, E. J., Gregory, K. D., Smith, T. A., Chrostowski, S. J., Garden, R. A., & O'Connor, K. M. (2000). *TIMSS 1999 international science report. Findings from IEA's repeat of the Third Inter-*

national Mathematics and Science Study at the eighth grade. Boston, MA: The International Association for the Evaluation of Educational Achievement (IEA).

Martin, M. O., Mullis, I. V. S., Gonzales, E. J., & Chrostowski, S.J. (2004). *TIMSS 2003 international science report. Findings from IEA's trends in International Mathematics and Science Study at the fourth and eighth grades*. Boston, MA: TIMSS& PIRLS International Study Center.

Mullis, I. V. S., Martin, M. O., Beaton, A., Gonzales, E. J., Kelly, D., & Smith, T. A. (1997). *Mathematics achievement in the primary school years: IEA's Third International Mathematics and Science Study (TIMSS)*. Chestnut Hill, MA: TIMSS international study center, Boston College.

Mullis, I. V. S., Martin, M. O., Gonzales, E. J., Gregory, K. D., Garden, R. A., O'Connor, K. M., Chrostowski, S. J., & Smith, T. A. (2000). *TIMSS 1999 International mathematics report. Findings from IEA's repeat of the Third International Mathematics and Science Study at the eighth grade*. Boston, MA: The International Association for the Evaluation of Educational Achievement (IEA).

Mullis, I. V. S., Martin, M. O., Gonzalez, E. J., & Kennedy, A. M. (2003). *PIRLS 2001 International report. IEA's study of reading literacy achievement in primary schools in 35 countries*. Boston, MA: The International Association for the Evaluation of Educational Achievement (IEA).

Mullis, I. V. S., Martin, M. O., Gonzales, E. J., & Chrostowski, S. J. (2004). *TIMSS 2003 international mathematics report. Findings from IEA's trends in International Mathematics and Science Study at the fourth and eighth grades*. Boston, MA: TIMSS& PIRLS International Study Center.

OECD (2001). *Knowledge and skills for life. First results from the OECD Programme for International Student Assessment (PISA) 2000*. Paris: OECD.

OECD (2004). *Learning for tomorrow's world. First results from PISA 2003*. Paris: OECD.

Postlethwaite, T. N., & Wiley, D. E. (Eds.) (1992). *Science achievement in Twente-Three countries: The IEA study of science II*. Oxford: Pergamon Press.

Rasbash, J., Browne, W., Goldstein, H., Yang, M., Plewis, I., Healy, M., Woodhouse, G., Draper, D., Langford, I., & Lewis, T. (2000). *A user's guide to Mlwin*. London: Multilevel Models Project, Institute of Education, University of London.

Reezigt, G. J., & Knuver, A. W. M. (1995). Zittenblijven in het basisonderwijs. *Pedagogische Studiën*, 72, 114-132.

Robitaille, D. F., & Garden, R. A. (Eds.) (1989). *The IEA study of mathematics II: Context and outcomes of school mathematics*. Oxford: Pergamon Press.

Rossi, P. H., Freeman, H. E., & Lipsey, M. W. (2004). *Evaluation, a systematic approach (7th Ed)*. Thousand Oaks, CA: SAGE Publications.

Scheerens, J., & Bosker, R. J. (1997). *The foundations of educational effectiveness*. Oxford: Pergamon.

Schultz, T. W. (1962). Reflections on investment in education. *Journal of Political Economy*, 70, 1-8.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston, MA: Houghton Mifflin Company.

Swanborn, P. G. (1999). *Evalueren*. Amsterdam: Boom.

Tymms, P., Merrell, C., & Henderson, B. (1997). The First year at school: A quantitative investigation of the attainment and progress of pupils. *Educational Research and Evaluation*, 3, 101-118.

Trochim, W. M. K. (1984). *Research design for program evaluation, the regression-discontinuity approach*. London: SAGE Publications.

Manuscript aanvaard: 21 september 2006

Auteur

Hans Luyten is als universitair docent verbonden aan de Faculteit Gedragswetenschappen (afdeling onderwijsorganisatie & -management) van de Universiteit Twente in Enschede.

Correspondentieadres: H.Luyten, Universiteit Twente, GW/O&M, Postbus 217, 7500 AE Enschede, e-mail: j.w.luyten@gw.utwente.nl

Abstract

The effect of education and the school effects in the Netherlands

This article shows how multilevel analysis, the standard method for assessing “school effects”, can be combined with the regression-discontinuity design. Whereas usual multilevel analyses only estimate relative school effects (differences between schools), the regression-discontinuity approach allows for an assessment of the absolute effect of one year schooling. A combination of both approaches yields an estimate of this absolute effect, but also of its variation between schools. Using this approach in the Dutch context, however, requires additional assumptions regarding the effects of grade retention. Application of this method in a secondary analysis on data from TIMSS-95 shows that this approach produces surprising new perspectives on the effectiveness of Dutch primary education and on the differences between schools.