

De studie van Allen et al. (2009) wijst er vooral op dat we voor het inschatten van de effecten van zittenblijven enkel mogen kijken naar methodologisch sterke studies. In de systematische literatuurstudie van Goos et al. (2013) werden 37 van die quasi-experimentele studies, gepubliceerd tussen 2001 en 2013 samengevat. Er werden ook enkel studies geselecteerd die de leerlingen meer dan twee jaar volgen (dus niet enkel in het bisjaar). De algemene conclusie uit die 37 studies is dat zittenblijven, gemiddeld genomen, een negatief effect heeft. Vanuit dat overzicht lijkt het mij nog steeds gepast om te stellen dat voor de gemiddelde zwakke leerling zittenblijven doorgaans minder gunstig is dan overgaan.

Het is normaal dat ik veel meer gewicht toeken aan de meta-analyse van Allen et al. (2009) dan aan de overzichtsstudie van Goos et al. (2013). De studie van Allen et al. is de meest recente kwantitatieve meta-analyse. Die studie is heel objectief en deskundig uitgevoerd, zonder een zweem van bias in de één of andere richting. De belangrijkste conclusie uit deze meta-analyse (zie citaat in mijn rapport) werd niet eens vermeld in het OBPWO-rapport, nl. dat als de studies beter worden, de effecten minder negatief zijn. Het begrip 'methodologisch sterke studie' is hier van belang. Voor Goos e.a. betekent dit eenvoudig dat er gebruik werd gemaakt van bepaalde statistische technieken (bv. propensity scores). Ik denk dat dit een veel te enge interpretatie is van methodologische kwaliteit. Om te kunnen spreken van een goede studie moet de vergelijkbaarheid van de groepen gewaarborgd zijn (dat is niet hetzelfde als ongeveer gelijk, of gelijk op propensity scores). Dit kan streng klinken, maar anders zijn valide conclusies onmogelijk (ik kom er dadelijk nog op terug). Bovendien is de studie van Goos e.a. geen kwantitatieve meta-analyse. Voor elke sub-onderzoeksvraag komen ze tot het besluit dat de resultaten niet eensluidend zijn. Wel besluiten ze dat daaruit blijkt dat de effecten van zittenblijven alleszins negatiever zijn *dan wat leerkrachten doorgaans denken*. Dit is dus slechts een vage en zeker geen kwantitatieve conclusie. Ik wees er al op dat deze studie bestaat uit heel wat onderzoeken die niet in wetenschappelijke tijdschriften gepubliceerd werden. Daardoor is er nauwelijks controle mogelijk op de kwaliteit van de studies, noch op de selectiecriteria van de auteurs.

Allen et al. (2009) schatten enkel het effect van het aantal jaren na de interventie apart voor de leeftijdsvergelijking en de leerjaarvergelijking. Ze modelleren niet het directe onderscheid tussen de leeftijdsvergelijking en leerjaarvergelijking.

Dat is correct, maar dat brengt de conclusies van Allen et al. in geen enkel opzicht in het gedrang.

Maar tussen het schrijven van het OBPWO-rapport (december 2011) en nu hebben we nog bijkomend en sterker onderzoek uitgevoerd, heb ik nog meer recente literatuur gelezen en met wat ik nu weet, zou ik niet meer adviseren om het zittenblijven eenvoudigweg af te schaffen. Als een onderzoeker op basis van extra onderzoeksresultaten haar inzichten nuanceert, getuigt dat van een juiste onderzoekshouding.

Ik kan het zeer waarderen dat je je mening hebt bijgesteld op grond van argumenten en onderzoek. Maar ik zie nog altijd niet op welk 'sterk onderzoek' je dan doelt. Het kan niet het onderzoek zijn van Goos et al. (2013), want die studie was al in press toen het rapport geschreven werd.

Daarbij aansluitend: van het OBPWO-rapport is enkel de [beleidssamenvatting](#) publiek gemaakt en daarin wordt de aanbeveling om het zittenblijven af te schaffen niet vermeld.

Dit is niet correct. In de beleidssamenvatting staat expliciet op pag. xvi het advies dat in het decreet basisonderwijs doorlopende leerlijnen zouden moeten worden opgenomen en dat dit het afschaffen van zittenblijven in het basisonderwijs impliceert.

De prestaties wiskunde en technisch lezen zijn inderdaad gemeten in het eerste leerjaar in het SiBO-onderzoek, maar dat zijn *pre-treatment* variabelen waarmee rekening gehouden is (meetmoment: juni 2004, wanneer alle leerlingen nog samen in L1 zitten).

Ik neem aan dat je bedoelt met 'rekening gehouden met' dat ze opgenomen werden in de propensity score. Daar gaat het hier niet om. Het gaat erom of de prestatiescores werkelijk gelijkend waren tussen de twee groepen.

De extrapolatiehypothese geldt zowel voor de leeftijdsvergelijking en de leerjaarvergelijking. De schatting van het intercept is hier dus een (mogelijks vertekende) extrapolatie.

Het gehanteerde model schat de intercept op grond van de beschikbare data en de aannames van het model. In het geval van de leerjaarvergelijking kan je inderdaad spreken van een extrapolatie omdat er nog geen data zijn voor de schoolprestatie begin 1^{ste} leerjaar. Voor de leeftijdsvergelijking zijn die data er wel en wordt de intercept ook geschat,

maar hier dan mede op grond van geobserveerde data (dus geen extrapolatie). Je probeert de ongelijke beginsituatie, zoals af te lezen uit figuur 2 en tabel 6 toe te schrijven aan ofwel weinig plausibele zomereffecten, ofwel aan de extrapolatie van de modelschatting (dus in werkelijkheid geen verschil, maar wel in het model). In de veronderstelling dat dit laatste klopt, dan is het model fout gespecificeerd, wat dan meteen vragen oproept bij de gehele analyse. Maar hoe dit ook zij, je draait hier de bewijslast om. Ik moet niet aantonen dat de groepen initieel ongelijk waren, de auteurs moeten aantonen dat de groepen gelijk waren. Op grond van de gepresenteerde data en analyses is daar echter geen grond voor.

Beste Wim, in tegenstelling tot wat je veronderstelt, verschillen de twee groepen niet op het einde van schooljaar 0304 (wanneer alle leerlingen nog samenzitten in het eerste leerjaar) voor wat betreft hun prestaties wiskunde, technisch lezen en de psychosociale variabelen. Je zou kunnen stellen dat de tijdscoording in de studie in *Journal of School Psychology* (2013) verwarrend is, en leidt tot extrapolatie, maar de conclusies blijven dezelfde.

Ik veronderstel niets, ik stel alleen vast dat de scores van de zittenblijvers bij het begin van het eerste leerjaar bis al substantieel lager waren dan die van de doorstromers begin tweede leerjaar. Jullie veronderstellen dan dat dit niet ligt aan het feit dat ze op het einde van het eerste leerjaar ook al verschillend waren (wat wel de meest waarschijnlijke aanname is). In het artikel worden daarover echter geen gegevens verstrekt! Jij verzekert me nu dat de twee groepen niet verschillend waren. Ik wil je wel geloven, maar een fout die hier gemaakt wordt (en helaas heel frequent voorkomt in gepubliceerd onderzoek) is dat je niet moet aantonen dat de nulhypothese niet kan verworpen worden, maar dat de nulhypothese waar is (wat niet hetzelfde is). Kortom, de match tussen beide groepen moet kunnen worden hard gemaakt.

Maar als u de conclusies van haar studie ongeldig vindt, dan stel ik voor dat u een commentaarstuk instuurt naar *Journal of School Psychology*. Zo'n commentaarstuk in een wetenschappelijk tijdschrift is het juiste kanaal voor kritiek.

Je weet heel goed dat de kans om een commentaarstuk in een journal geplaatst te krijgen heel klein is (de meeste tijdschriften zijn daar niet tuc op). Een beter idee is, zoals ik al ettelijke keren heb aangeboden, om samen een nieuw onderzoek uit te voeren waarin tegemoet gekomen wordt aan de hier geformuleerde bezwaren in de proefopzet. In een publicatie die daaruit kan voortvloeien, kan dan begonnen worden met aan te geven wat de problemen waren in eerder onderzoek. En verder zijn er geen 'juiste' kanalen voor kritiek, alle kanalen zijn legitiem.

De kost van zittenblijven staat los van de effecten... Maar het is niet verantwoord om kosten te doen voor een 'behandeling' die geen positieve bewezen effecten heeft.

Hier valt m.i. weinig zinnigs over te zeggen. Uiteraard heeft zittenblijven een kost, net als het hele onderwijssysteem een kost heeft voor de samenleving (maar ook baten). Opnieuw doe je alsof zittenblijven kan worden beschouwd als een geïsoleerde interventie ('behandeling') in het onderwijssysteem. Dat is niet reëel. De echte vraag wordt dan wat de kost is van het afschaffen van zittenblijven of het sterk problematiseren van zittenblijven. In heel wat studies van de groep van Hanushek lees ik dat de opbrengsten van het onderwijs voor de samenleving vooral afhangen van de cognitieve uitkomsten. Als dat klopt (maar ik ben geen econoom), dan doet Vlaanderen het op dat vlak nog altijd zeer goed.

Mijn onderzoekscentrum heeft de laatste jaren sterk ingezet op quasi-experimentele designs, vooral om het zittenblijven in de derde kleuterklas en het eerste leerjaar te onderzoeken. Dat type onderzoek (met propensity scores) is momenteel de *state of the art*. Als er betere designs en machtingsmethoden ontwikkeld worden, zullen we die zéker toepassen.

Hier zeg je, wij gebruiken de best mogelijke methoden die er momenteel zijn. Ja en neen. Ja, als je bedoelt dat je de methoden hanteert die in recent onderzoek gangbaar zijn, en dan vooral de statistische technieken. Dat begrijp ik perfect en is op zich te verdedigen. Maar dat is nog altijd geen vrijgeleide om niet kritisch te blijven nadenken over de kwaliteit van het gehanteerde design. En daarom zeg ik neen. De principes van het kunnen uitsluiten van storende variabelen zijn zo oud als de wetenschap zelf en gelden onverkort. Het is niet voor niets dat alle toonaangevende auteurs die onderzoek gedaan hebben naar de effecten van zittenblijven waarschuwen voor de valkuil van de selectie-bias. Ik begrijp heel goed dat jullie ervan uit gingen dat door te werken met propensity scores die selectie-bias in voldoende mate onder controle is. Maar ook in de gespecialiseerde statistische literatuur over propensity scores waarschuwt men voor het gevaar van interventie gerelateerde selectie (Austin, 2008).

Ik weet dat ik hier aanzienlijk strenger ben, maar m.i. met recht en reden. Mijn wetenschappelijk werk gaat in grote mate precies over de relatie tussen de kwaliteit van onderzoeksdesigns en de al dan niet correcte conclusies die daaruit kunnen worden getrokken. Mijn recente publicaties in *Developmental Psychology* en *Cognitive Psychology*

gaan over een design waarbij men zwakke lezers vergeleek met jongere lezers van eenzelfde leesniveau (dus ook een soort matching). Dit design is letterlijk honderden keren toegepast en leverde stevast (ook in meta-analyses) een zelfde conclusie op betreffende leesstoornissen (dyslexie), nl. dat dyslectici een specifiek probleem zouden hebben met het lezen van nonsens woorden (het 'nonword reading deficit'). Ook dit werd beschouwd als een soort van 'established fact'. In deze twee publicaties heb ik samen met Astrid Geudens kunnen aantonen dat dit effect een zuiver artefact was van de methode. Maar er is meer. We zijn nu volop bezig aan te tonen dat in massaal veel klinisch onderzoek de gerapporteerde effecten vaak ook een artefact kunnen zijn van een inadequate matchingprocedure. Al dit werk maakt me natuurlijk bijzonder alert voor mogelijke problemen met vergelijkingsgroepen (en dat is heus geen vitterij).

Verder kijken we uit naar het onderzoek dat collega Van den Broeck zelf zal uitvoeren over het thema zittenblijven.

Ik zou graag willen, maar heb daartoe momenteel niet de onderzoekscapaciteit. Mijn aanbod om hier samen verder onderzoek over te doen, blijft dus gelden. Dit mag je ook zien als een blijf van vertrouwen. Onze vinnige discussie heeft dat niet ondermijnd.

Prof. Dr. Wim Van den Broeck
Klinische en Levensloopspsychologie
Faculteit Psychologie en Educatiewetenschappen
Vrije Universiteit Brussel
Pleinlaan 2 B-1050 Brussel