

Pygmalion in de klas: Over verwachtingen van leraren en de invloed van medeleerlingen¹

J. Terwel

Samenvatting

In deze discussiebijdrage bespreek ik het onderzoeksrapport van Paul Jungbluth, getiteld "De ongelijke basisschool". De uitkomsten van dit onderzoek worden allereerst geplaatst tegen de achtergrond van de researchtraditie naar verwachtingen van leraren. Vervolgens wordt ingegaan op de analyses en uitkomsten van Jungbluths onderzoek. Jungbluths conclusie luidt dat de "verwachtingen" (veronderstelde talenten van de leerlingen) van leraren van invloed zijn op het onderwijsaanbod en op de prestaties van de leerlingen. Een leerling van wie de leraar weinig verwacht, krijgt minder leerstof aangeboden, en presteert minder dan een leerling die hoger scoort op de schaal van de leraarsverwachting. Deze verwachtingen blijken gerelateerd aan het sociale milieu. In een kritische analyse vanuit een 'resource'-theoretisch perspectief, verdedig ik de stelling dat Jungbluth er niet in is geslaagd overtuigend aan te tonen dat het gaat om effecten van leraarsverwachtingen op de leerlingen. Zijn studie geeft eerder aanleiding om de causale richting om te keren: de leraar is geen "valse" profeet, maar een marionet die wordt bespeeld door de leerlingen in zijn klas. Deze stellingname wordt geplaatst in het bredere kader van het internationale onderzoek naar klascompositie en 'peer influences'. De discussie besluit met aanbevelingen voor verder onderzoek vanuit een breed-spectrummethodologie.

1 Introductie

Iedereen kent wel verhalen over voorspellingen van leraren en hoe die doorwerken in schoolprestaties en schoolloopbanen van kinderen. Die voorspellingen kunnen expliciet worden uitgesproken, maar ook onbewust doorwerken in houding en gedrag van leraren. Het zijn profetieën die zichzelf waarmaken. Er kraait geen haan meer naar. Soms

wordt echter een "valse" profetie ontmaskerd; luister eens naar Asuman Bayraktar: "Op de basisschool had ik geen flauw benul wat ik kon tot ik bij de Cito-toets de hoogste punten haalde van de hele klas." Toen ze zei dat zij naar het gymnasium wilde, zei haar leraar "dat lukt jou toch niet". Asuman zette door en ging daarmee tegen de voorspelling van haar leraar in (Broer, 2000). Als onderwijswetenschappers dit type processen eens zouden kunnen verhelderen, deze voorbeelden uit de anekdotische sfeer zouden kunnen halen ... Als we de causale mechanismen zouden kunnen blootleggen, dan is er een theoretische en empirische basis voor vernieuwing, zeker wanneer deze mechanismen ook nog beïnvloedbaar blijken te zijn.

Paul Jungbluth (2003a) heeft in een recente studie, "De ongelijke basisschool" onderzocht of leraren op de basisschool verschil maken tussen leerlingen die objectief gezien even talentvol zijn. Deze studie past in een onderzoekstraditie die bekend staat als het onderzoek naar leerkrachtverwachtingen of "Pygmalion in de klas". Dit type onderzoek is erop gericht relaties op te sporen tussen verwachtingen van leraren over de talenten van leerlingen, en hun leerprestaties op een later tijdstip. Hoewel dit denken een langere historie kent, werd het pas wereldwijd bekend door het onderzoek van Rosenthal en Jacobson in de 60-er jaren van de vorige eeuw. Een IQ-test werd afgenomen bij leerlingen. Uit deze groep werd volgens toeval een aantal leerlingen gekozen. Leerkrachten werd meegedeeld dat deze leerlingen volgens de test tot de 'bloomers' behoorden, zij zouden de komende periode een groeispurt in hun cognitieve ontwikkeling doormaken. Gedurende het lopende jaar werden alle kinderen opnieuw getest. De zogenaamde bloomers bleken significant hoger te scoren dan hun even knappe medeleerlingen. De foute verwachtingen van leraren (ze waren immers door de onderzoekers op het verkeerde been gezet) bleken een duidelijk effect te hebben

(Rosenthal, 1987; Rosenthal & Jacobson, 1968; Weinstein, 2002; Wineburg, 1987). Nu zou men willen weten hoe deze verwachtingen van leraren doorwerken in hun pedagogisch-didactisch handelen in de klas; dan pas kan iets zichtbaar worden van de causale keten: leerlingkenmerken - verwachtingen - interactie in de klas - leerresultaten. Met dat zoeklicht heb ik naar de studie van Jungbluth gekeken.

2 De ongelijke basisschool

Aan de studie van Jungbluth is in de pers veel aandacht besteed. Dat is terecht, want het gaat om een belangrijk thema. Het thema *ongelijkheid* is de afgelopen decennia naar de zijlijn van het onderwijswetenschappelijke onderzoek en het maatschappelijke debat geschoven. Daarom is het verheugend dat Jungbluth het telkens weer op de agenda weet te krijgen.

2.1 Voorgeschiedenis

In zijn rapport geeft Jungbluth een boeiende beschrijving van de geschiedenis van het debat en het beleid inzake onderwijsongelijkheid in Nederland. Hij geeft ook een beschrijving en analyse van twee vernieuwingsbewegingen, en dan met name de meer leerlinggerichte beweging (traditionele vernieuwingsscholen) versus de meer leerkrachtgestuurde beweging (effectieve scholen), waarbij hij een voorkeur aan de dag legt voor de tweede: de 'effective schools movement'. Dat onderscheid zou echter genuanceerd kunnen worden, waarbij een derde oplossingsrichting in beeld komt. In een meer genuanceerde benadering zou ook de recente theorievorming, geïnspireerd op onder andere Dewey, Vygotsky en Freudenthal, aan bod zijn gekomen. Dit type vernieuwingen laat zich niet in deze dichotomie vangen, omdat het de pedagogisch-didactische interactie in de klas naar vorm en inhoud centraal stelt. Die interactie wordt wel aangeduid met 'guided reinvention' en 'guided co-construction'. Het onderzoek van Carol Lee (2001) dat zich richt op het proces van 'cultural modeling' in de klas is hiervan een mooi voorbeeld (zie ook Lowyck & Terwel, 2003). Ik kom daar later nog op terug.

2.2 Vraagstelling en opzet van het onderzoek

Kort gezegd, is de vraagstelling van het onderzoek: "Welke formele en informele verschillen in het onderwijsaanbod ontwikkelen zich tussen en binnen basisscholen in relatie tot de individuele en collectieve sociaal-etnische achtergrond van de leerlingen?" (Jungbluth, 2003a, pag. 2). Vervolgens is er een toespitsing op de relatie tussen deze verschillen in aanbod en de leerprestaties (effecten) op individueel en schoolniveau (pag. 45). Deze vraagstelling plaatst Jungbluth tegen de achtergrond van de researchtraditie naar de effecten van verwachtingen van leraren op het onderwijsaanbod en de leerresultaten van leerlingen. In Jungbluths studie zijn de verwachtingen van leraren in "natuurlijke" situaties onderzocht; de leraren werden niet op het verkeerde been gezet in een experiment (zoals in de klassieke studie van Rosenthal & Jacobson, 1968). Hun werd gevraagd de talenten van hun leerlingen te schatten. Deze verwachtingen van leraren werden in Jungbluths onderzoek in kaart gebracht door leraren te vragen hun leerlingen ook nog in te delen naar veronderstelde capaciteiten of talenten (cognitieve capaciteiten). In een aanvullende vragenlijst, bij een beperkte en extreme selectie uit het PRIMA-bestand, vroeg Jungbluth de leraren om hun leerlingen ook nog per klas in te delen in drie statusgroepen: kopgroep, middengroep of staartgroep. Per statusgroep moesten de leraren aangeven hoe actueel of opportuun bepaalde leerstof was. Deze fictieve, relatieve groepsindeling is in het onderzoek onder meer gebruikt om relaties te bepalen met achtergrondvariabelen van leerlingen, en om vast te stellen wat per statusgroep en per individuele leerling het onderwijsaanbod is en wat de effecten hiervan zijn op de prestaties van de leerlingen.

2.3 Steekproef, analysemethoden en resultaten

De grootte van de steekproef varieert per type analyse: tussen meer dan 11000 en minder dan 150 leerlingen. Jungbluth gebruikt aanvankelijk vooral beschrijvende statistieken zoals kruistabellen met gemiddelden en percentages waarop significantietoetsen worden toegepast. Daarna volgen regressieanalyses

op individueel niveau, waarbij variabelen op groepsniveau worden toegeschreven aan individuele leerlingen. Hauser (1970) toonde aan dat de analyse van contexteffecten via kruistabellen niet adequaat is. Hij stelde voor om gebruik te maken van covariantieanalyse en regressieanalyse. Zijn kritiek was raak, maar zijn aanbeveling was onjuist, althans naar huidige inzichten. Het toeschrijven van groepskarakteristieken aan het individuele niveau en het uitvoeren van regressieanalyses op individueel niveau, zoals Hauser voorstelde, is nu in diskrediet geraakt. Hetzelfde geldt overigens voor het terugbrengen van het individuele niveau tot het groepsniveau via aggregatie (Bryk & Raudenbush, 1992). Hoewel Jungbluths kruistabellen en regressieanalyses wel enige beschrijvende waarde hebben, leveren zij dus geen solide basis voor het trekken van definitieve conclusies.

Ten slotte volgen vier ‘multilevel’-analyses. Bij het eerste tweetal gaat het om een analyse naar het effect van *opleiding moeder*, *herkomstland moeder*, en *non-verbale intelligentie* (individueel- en schoolniveau) op respectievelijk *Begrijpend Lezen* in groep 8 en de *Cito-eindtoets* op de uitgebreide dataset van PRIMA 2001, met respectievelijk 3000 en 2300 leerlingen. Zoals te verwachten, zien we bij de drie variabelen op individueel niveau bij beide analyses een significant effect. Op het schoolniveau blijkt alleen de gemiddelde non-verbale intelligentie een significante bijdrage te leveren.

In het tweede tweetal analyses wordt het model uitgebreid met twee variabelen (op beide niveaus): het *aanbodniveau* en de *cognitieve capaciteiten* (verwachtingen van de leraar). Ook hier gaat het om *Begrijpend Lezen* en de *Cito-eindtoets* als afhankelijke variabelen. Echter, hoewel het aantal variabelen in het model aanzienlijk is toegenomen, is het aantal leerlingen drastisch gereduceerd tot respectievelijk 183 en 149 leerlingen. De uitkomsten van dit tweetal analyses zijn merkwaardig: krachtige voorspellers op het individuele niveau (opleiding moeder, herkomstland moeder en non-verbale IQ) vallen weg als determinanten van leerprestaties, terwijl de cognitieve capaciteiten (leraarsverwachtingen) een significante bijdrage leveren. Bovendien is het aanbodniveau significant,

zij het alleen bij de Cito-eindtoets. Op het tweede niveau, het schoolniveau, zijn er nu nog nauwelijks variabelen met significante effecten over; zelfs het gemiddelde non-verbale IQ levert in dit model geen bijdrage meer.

2.4 Conclusies van de studie volgens Jungbluth

Volgens Jungbluth blijkt dat de “verwachtingen” (cognitieve capaciteiten of verondersteld talent van de leerling) van leraren van invloed zijn op het onderwijsaanbod en op de prestaties van de leerlingen. Een leerling waarvan de leraar weinig verwacht, krijgt minder leerstof aangeboden en presteert minder dan een leerling die hoger scoort op de schaal van de leraarsverwachting. Deze verwachtingen blijken gerelateerd aan het sociale milieu. Leraren verwachten minder van kinderen uit kansarme milieus, en zo doen zij er nog een schepje ongelijkheid bovenop. Met dit onderzoek probeert Jungbluth aan te tonen dat niet alleen buitenschoolse factoren zoals sociaal milieu en gezin, maar ook factoren in de school en in de klas bijdragen aan de kansongelijkheid in het onderwijs. De leraar is daarbij de spil waar alles om draait. Voor onderwijsvernieuwers en beleidsmakers lijkt dat belangwekkend nieuws; aan gezin en buurt kun je weinig veranderen, maar de leraar is te bereiken, bijvoorbeeld door voorlichting, opleiding en nascholing.

3 Kritische evaluatie

Het onderzoek is binnen de beperkte randvoorwaarden van het databestand en de onderzoekssituatie degelijk opgezet. De theoretische vooronderstellingen zijn expliciet vermeld en Jungbluth geeft de lezer ook alternatieve interpretaties van zijn bevindingen. Het eerste tweetal ML-analyses, op de uitgebreide dataset, levert conclusies op die overeenkomen met elders verricht onderzoek, namelijk dat het gemiddelde cognitieve niveau van de klas een effect heeft op de leerprestaties, bovenop de effecten van variabelen op het individuele niveau. Medeleerlingen hebben een significante invloed op de leerprestaties van een leerling.

Mijn kritiek richt zich primair op het tweede tweetal ML-analyses, waarbij het beperkte databestand is gebruikt, waarin de twee cruciale variabelen *leraarsverwachtingen* (cognitieve capaciteiten) en *aanbodniveau* aan het model zijn toegevoegd en waarop de hoofdconclusies zijn gebaseerd. Mijn kritiek is samen te vatten in vier punten: (1) het ontbreken van beschrijvende statistieken; (2) een te beperkte steekproef; (3) onvoldoende controle op voorkennis; en (4) een conclusie die niet past bij het gehanteerde model.

3.1 Ontbrekende statistieken

Jungbluth maakt het de kritische lezer niet gemakkelijk, omdat hij in zijn rapport niet systematisch verslag doet van beschrijvende statistieken zoals gemiddelden, standaarddeviaties, en correlaties tussen de variabelen. Een eenvoudig overzicht van onder meer de verdeling van die (183 of 149) leerlingen over de klassen (leraren), de gemiddelden en standaarddeviatie per klas, en de correlaties tussen de variabelen, zou een solide basis zijn voor een kritische evaluatie van het onderzoek. Misschien zou dan blijken dat er door de extreme selectie die Jungbluth heeft toegepast op bepaalde variabelen, tussen klassen nauwelijks nog verschillen bestaan, of dat het patroon van correlaties zodanig is dat opname van al deze variabelen in één model niet zinvol is. Die gegevens zijn ook essentieel om te kunnen bepalen wat de verschillen zijn tussen de data in het totale PRIMA-bestand en de gegevens in Jungbluths selectie. Die ontbrekende statistieken en correlaties bemoeilijken de interpretatie van de uitkomsten van het laatste tweetal multilevel-analyses, in respectievelijk Tabel 10.4 en 10.5, waarop de eindconclusies zijn gebaseerd. Daar komt nog bij dat de Tabellen 10.2 en 10.3 verkeerd zijn geplaatst, waardoor de tekst niet correspondeert met de getallen in de tabellen.

3.2 Beperkte, eenzijdige steekproef

Op basis van zijn beperkte en specifieke selectie uit het PRIMA-bestand concludeert Jungbluth: “Met zekerheid valt te concluderen dat de verwachtingen van leerkrachten van grote invloed zijn voor de ongelijkheid in prestaties tussen leerlingen, waaronder voor

de ongelijkheid die verband houdt met hun sociaal-etnische achtergrond. De manier waarop die verwachtingen werken loopt onder meer via het aanbodniveau.” (Jungbluth, 2003a, pag. 145).

In deze analyses is “ingezoomd” op extreem gesegregeerde scholen en vervolgens, voor zover mogelijk, binnen elk stratum nog weer op opmerkelijk effectieve, dan wel opmerkelijk ineffectieve scholen (in termen van de PRIMA-toets voor rekenen in groep 8). Het inzoomen op een kleine, extreme dataset is uit oogpunt van grootte en representativiteit problematisch. Het is een ongelukkige ingreep waardoor geen solide conclusies over het effect van leraarsverwachtingen te trekken zijn. De hierbij gehanteerde definitie van effectieve scholen is gebaseerd op de Cito-eindtoets, gecorrigeerd voor SES en type schoolbevolking. Overigens is het niet duidelijk waarom Jungbluth de verwachtingen van leraren niet heeft gebruikt in de multilevel-analyse op de uitgebreide PRIMA-dataset met enkele duizenden leerlingen en honderden leraren (dus in het eerste tweetal ML-analyses).

3.3 Onvoldoende controle op cognitief beginniveau

Wat ook opvalt, is de zeer beperkte controle op cognitief niveau (via non-verbaal IQ). Controle op voorkennis ontbreekt zelfs geheel. Non-verbaal IQ verklaart slechts een klein deel van de variantie. Dit maximaliseert de “effecten” van andere variabelen in het model, zoals de sociaal-etnische achtergrond en verwachtingen van leraren. Het is van belang om Ausubels adagium in het oog te houden: “If I had to reduce all of educational psychology to just one principle, I would say this: The most important single factor influencing learning is what the learner already knows.” (Ausubel, 1968).

3.4 Inadequaat model

Dat het effect van de leraarsverwachtingen *via* het aanbodniveau zou lopen, wordt niet aangetoond. Het gehanteerde model - met alleen directe effecten - laat deze conclusie niet toe. Het aanbod is in deze hoofdconclusie van Jungbluth een ‘mediator’-variabele. Voor het vaststellen van het effect van een me-

diërende variabele, zoals *aanbodniveau*, zijn echter andere modellen en andere analyses nodig (vgl. Holmbeck, 1997).

4 Reflectie: Over de invloed van medeleerlingen

Met de focus op leraarsverwachtingen en aanbodniveau is de invloed van medeleerlingen buiten beeld geraakt. Jungbluths onderzoek toont namelijk wél aan dat de gemiddelde capaciteiten (non-verbale IQ-scores) van de medeleerlingen in een klas een significant effect hebben op de prestaties van een leerling in die klas (Begrijpend Lezen en de Cito-eindtoets). Die conclusie is gebaseerd op de uitgebreide dataset met ruim voldoende leraren en leerlingen (ter relativering: ook bij deze conclusie geldt mijn kritiek op onvoldoende controle op het cognitief beginniveau van de leerlingen, maar dit is geen absoluut oordeel, omdat er wel voor non-verbale IQ op individueel niveau is gecontroleerd). Van belang is ook dat het effect van het klassengemiddelde geheel in lijn is met wat in internationale en vergelijkende studies is gevonden. Vaak gaan deze studies nog verder en worden naast directe effecten ook differentiële (interactie)effecten gevonden van klascompositie². Uit een grote studie in Engeland, waar op de basisschool soms al 'streaming' wordt toegepast, komt de volgende conclusie naar voren: "The data showed that students with high prior achievement did better in streams than did those in an unstreamed school. However, those of lower measured achievement did worse than predicted. They did better in mixed-ability situations in which schools were well mixed."³ En de conclusie uit een zeer grote, internationale, vergelijkende studie ligt in dezelfde lijn: "The authors found that the higher the mean test score of class mates, the higher the achievement level of students and that where high- and low-ability are in a mixed classroom, most of the gains were captured by low-ability students."³ Uit een recente studie van De Fraine (2003) bleek opnieuw: je bent als leerling beter af in een knappe klas dan in een klas met een laag gemiddelde (vgl. ook Opendakker, 2003; Westerbeek, 1999).

Een van de problemen waar Jungbluth mee worstelt, is het databestand. Het PRIMA-bestand waar hij zich op baseert, bevat geen gegevens over de interactieprocessen tussen leraar en leerlingen en over het feitelijk gerealiseerde curriculum. Jungbluth heeft bij een zeer beperkt aantal leraren (en hun leerlingen) uit het PRIMA-bestand wel het "aanbod" proberen te meten via een leerkrachtvragenlijst, maar het curriculum zoals feitelijk gerealiseerd in de klas blijft volledig in het duister. De indeling in *statusgroepen* en het daarbij geïndiceerde *aanbod* zijn slechts virtuele grootheden in de perceptie van de leraar. De statusgroepen zijn geen didactisch identificeerbare 'streams' of groepen die binnen de klas een gedifferentieerd aanbod krijgen. In de analyse en in zijn spraakgebruik wordt gesuggereerd dat het om (fysieke), observeerbare groepen zou gaan. In feite werd van de leraar, toen hem gevraagd werd zijn leerlingen in drie statusgroepen in te delen, een exercitie gevraagd die niet los te zien valt van zijn verwachtingen omtrent het talent van zijn leerlingen. Geen wonder dat later blijkt dat deze verwachtingen een hoge correlatie vertonen met de toewijzing aan statusgroepen, met het onderwijsaanbod en de afhankelijke variabele *leerprestaties*. Het opnemen, in één model, van zo'n groot aantal onafhankelijke variabelen die onderling hoog correleren (multicollineariteit) kan leiden tot moeilijkheden bij de interpretatie van de uitkomsten (Pedhazur, 1982, pag. 232). Die problemen worden nog versterkt bij een kleine steekproef van 149 leerlingen. Daar komt nog bij: het behoren tot een statusgroep in de ogen van de leraar lijkt minder relevant dan een statusindeling door medeleerlingen, maar dat is niet gemeten. Bovendien bleek er bij leraren op scholen met veel kansarme leerlingen weinig bereidheid om de vragenlijst in te vullen. Jungbluths eindconclusies betreffende de verwachtingen van leraren, het curriculaire aanbod, en de leerprestaties zijn dan ook gebaseerd op een te kleine en te eenzijdige steekproef om valide conclusies te kunnen trekken uit de multilevel-analyses.

Een ander probleem betreft de leraarsverwachting op zichzelf: leraren zullen hun verwachtingen toch ook baseren op de kennis en

vaardigheden die hun leerlingen in de klas meebrengen. Uit ander onderzoek blijkt dat voorkennis ('past performance') in de brede zin van het woord een krachtige voorspeller is van toekomstig schoolsucces. Geen wonder dat "verwachtingen" van leraren ertoe doen; zij baseren zich op dagelijkse ervaringen met de leerlingen.

Bryk en Raudenbush (1992) voerden een meta-analyse uit over 19 studies met als hypothese: "Hoe langer de leraren hun leerlingen kennen op het moment waarop zij de misleidende informatie ontvangen, des te kleiner het effect op de leerprestaties." Nieuwe informatie over de talenten van leerlingen die afwijkt van reeds gevestigde inzichten wordt minder snel opgenomen. In de experimentele groepen werden leraren aangemoedigd om hoge verwachtingen van hun leerlingen te koesteren. In de controlegroepen ontbrak die aanmoediging. Het effect van de kunstmatig opgeroepen leraarsverwachtingen bleek inderdaad af te nemen naarmate er langer contact was geweest tussen leraar en leerlingen.

Ten slotte is er de cruciale vraag: Is het de leraar die het verschil maakt of zijn het de medeleerlingen in de klas? Is de leraar wel de spil waar alles om draait? Er valt heel wat af te dingen op die autonome positie die Jungbluth aan de leraar toekent; de leraar is niet alleen een profeet die in houding en gedrag haar eigen voorspellingen waarmaakt, zij staat ook onder invloed van haar leerlingen. Daar komt nog bij dat leerlingen elkaar onderling sterk beïnvloeden. In vele onderzoeken komt naar voren dat de capaciteiten van medeleerlingen belangrijke voorspellers zijn van de leerprestaties van een leerling. In klassen met veel laagpresterende leerlingen is het moeilijk om onderwijs van goede kwaliteit te realiseren en leerlingen dezelfde kansen te bieden als hun tegenvoeters in de betere klassen (Lowyck & Terwel, 2003).

Jungbluth heeft een belangrijke studie verricht, maar de beperkingen van de data, het gehanteerde model en de analyse (qua selectie, steekproef, aard van de metingen, en geringe controle op relevante co-variabelen zoals voorkennis) maken het uiterst moeilijk om betrouwbare en valide uitspraken te doen omtrent de mate waarin en vooral de wijze

waarop verwachtingen van leraren doorwerken in onderwijsleerprocessen en leerresultaten. Naar mijn mening is Jungbluth er niet in geslaagd overtuigend aan te tonen dat het gaat om effecten van leraarsverwachtingen op de leerlingen. Samengevat en bewust chargerend: zijn studie geeft eerder aanleiding om de causale richting om te keren: de leraar is geen (valse) profeet, maar een marionet die wordt bespeeld door de leerlingen in zijn klas, zoals blijkt uit het effect van het klassengemiddelde van het non-verbale IQ op de prestaties van de leerlingen. Met andere woorden, het lijken eerder de medeleerlingen dan de leraar die de leereffecten bepalen. Met de omkering van de causale richting blijf ik echter binnen de hierboven genoemde dichotomie in Jungbluths denkkader (leraargericht versus leerlinggericht) en binnen de beperkingen van het PRIMA-bestand waarin het pedagogisch-didactische proces niet is gemeten. Daarom geef ik hieronder enkele suggesties voor verder onderzoek die deze dichotomie overstijgen door de *interactie tussen leraar en leerlingen* centraal te stellen. Zo hoop ik het dilemma in het onderzoek naar 'self-fulfilling prophecies', namelijk de vraag "Is het proces 'teacher-driven' of 'student-driven'?", op een hoger plan aan de orde te stellen (vgl. Weinstein, 2002).

5 Suggesties voor verder onderzoek: Breed-spectrummethodologie

Bij de studie van sociale ongelijkheid in het onderwijs pleit ik voor een breed-spectrummethodologie. Daarbij is plaats voor (1) experimenteel onderzoek; (2) kwalitatief, descriptief, etnografisch onderzoek; en (3) 'design-experiments'. Ik zie deze drie benaderingen als aanvulling op en als alternatief voor de 'black box'-analyses op het PRIMA-bestand, waarbij de interactieprocessen niet in beeld komen.

5.1 Experimenteel onderzoek

In vervolgonderzoek zou meer aandacht besteed moeten worden aan de observatie van het interactieproces in de klas en de inhoud van het curriculum. Inmiddels zijn er methodologische ontwikkelingen, met name

longitudinale ML-modellen en Structural Equation Modelling (SEM), die het mogelijk maken om herhaalde metingen van leerprestaties en observaties van het proces in de klas op te nemen en zo dieper door te dringen in de relaties tussen interactieprocessen, de factoren die hierbij een rol spelen, en de leerresultaten (vgl. Brekelmans, Van den Eeden, Terwel, & Wubbels, 1997; Holmbeck, 1997; Terwel, Gillies, Van den Eeden, & Hoek, 2001). Door kenmerken van leerlingen en klassen, en interactieprocessen in de analyse te betrekken, kunnen leraarsverwachtingen en de gevolgen hiervan in termen van leerwinst of groei, nog beter in kaart worden gebracht. Zo kan onderzoek ook een bijdrage leveren aan verbetering van het onderwijs in de klas. Jungbluth heeft met het PRIMA-bestand nauwelijks greep gehad op het onderzoeksontwerp, omdat dit werd gedecteerd door de verzamelde gegevens, laat staan dat hij invloed kon uitoefenen op het curriculaire aanbod en de inrichting van de klassituatie. Het gaat om 'post hoc'-analyses die vele alternatieve interpretaties mogelijk maken. Het verdient aanbeveling om met goed gecontroleerde veldexperimenten te werken (van voldoende, maar toch beheersbare omvang van 200-800 leerlingen). Leraren zouden getraind en beloofd moeten worden voor hun medewerking. Uit onderzoek in zowel natuurlijke situaties als uit experimenten met aselechte toewijzing blijkt dat als gemiddelde leerlingen in een hogere 'track' worden geplaatst, zij beter presteren bij wiskunde dan vergelijkbare leerlingen in de lagere track (Mason, Schroeter, Combs, & Washington, 1992). Dergelijke experimentele ingrepen, met aselechte toewijzing van leerlingen aan condities, zou men moeten combineren met leraarsverwachtingen die ook experimenteel - in de klassieke Rosenthal-en-Jacobson-traditie - zouden kunnen worden gevarieerd. In deze kwantitatieve studies zouden ook de interactieprocessen in de klas direct via herhaalde observatie en videoregistratie moeten worden gemeten en in de ML- of SEM-analyses worden meegenomen. Zo'n studie is zeer ambitieus en in de praktijk moeilijk te realiseren, omdat: (1) aselechte toewijzing van leerlingen aan klassen van verschillend niveau op praktische en ethische weerstanden

stuit; (2) systematische, kwantitatieve observaties en effectmetingen op verschillende momenten in een longitudinaal traject hoge kosten meebrengen; (3) het induceren van onjuiste verwachtingen bij leraren ten behoeve van het onderzoek naar self-fulfilling prophecies op praktische en ethische problemen stuit. Toch zou een dergelijke onderzoek echt een doorbraak kunnen betekenen. Eventueel is ook te denken aan computersimulaties om de praktische en ethische problemen te omzeilen.

5.2 Kwalitatief, descriptief, etnografisch onderzoek

Het vraagstuk van de verwachtingen van leraren kan ook nog op een geheel andere wijze worden benaderd. Zou de invloed van leraarsverwachtingen niet ook en vooral via een aaneenschakeling van "incidentele" ontmoetingen, terloopse opmerkingen en inspirerende situaties verlopen? Dergelijke ontmoetingen komen niet in beeld met een puur kwantitatieve benadering. Ik denk daarbij aan de opmerking van de leraar tegen Asuman Bayraktar, maar ook aan die tekenleraar op de Its die tegen een leerling zei: "Weet je wel dat je ook in de avonden een opleiding tot leraar kunt volgen?". Het effect van dergelijke opmerkingen moet niet unicausaal worden opgevat. Het is geen eenrichtingsverkeer van leraar naar leerling. Het gaat ook om de betekenis die de leerling eraan geeft. Sommige opmerkingen van leraren slaan in als een bliksemschicht. Vaak weten leraren niet wat hun opmerking, hun les, hun houding teweeg heeft gebracht. Een respondent in het onderzoek van Jan Brands (1992, pag. 125) zei het volgende over haar leraar "Hij had de opstelling van: wie in de klas dan ook, iedereen heeft een betekenis, wie je ook bent. Ik heb hem daar later op aangesproken en hij herkende het niet, maar hij heeft mijn nieuwsgierigheid aangemoedigd, mijn intuïtieve neiging de dingen niet zo te vertrouwen als ik ze te horen kreeg." In de literatuur over verwachtingen van leraren is hierover maar weinig te vinden. Dit type ervaringen zou ik "onderwijspedagogische ontmoetingen" willen noemen (vgl. Guardini & Bollnow, 1962).

Echte ontmoetingen kun je niet maken of afdwingen. Het goede moment moet er zijn

en er moet openheid zijn van beide kanten. Na zo'n echte ontmoeting, waarin altijd een element van vrijheid aanwezig is, ontstaat - soms pas jaren later - het gevoel "dit heeft zo moeten zijn" of "deze ontmoeting heeft een wending in mijn leven gebracht". Daarbij kan het gaan om een ontmoeting met personen zoals een leraar of een medeleerling, maar ook om de ontmoeting met een inhoud of een inzicht. Echte ontmoetingen zijn niet vrijblijvend; men wordt er tot in de kern door op de proef gesteld (Guardini & Bollnow, 1962). Een andere respondent uit het onderzoek van Brands (1992, pag. 124) zei het volgende over zijn leraar biologie: "Hij stelde vragen als 'Hoe is het leven ontstaan?' Dat vond ik allemaal prachtig. Want dat was voor mij de definitieve afrekening met de godsdienst ... Dat sprak mij erg aan: niks geloven, maar uitzoeken."

Om het gebied van leraarsverwachtingen in volle breedte te onderzoeken, is dus naast kwantitatief onderzoek ook kwalitatief, descriptief, etnografisch onderzoek nodig. Leraren kunnen met hun verwachtingen en opmerkingen vensters openen en nieuwe wegen wijzen. Dat zie ik als een opdracht van de school: ervaringen mogelijk maken die verder reiken dan wat het kind thuis of in de buurt meemaakt. Het zou goed zijn om de invloed van leraren op belangrijke keuzemomenten in het onderwijs, zoals de overgang van basisonderwijs naar voortgezet onderwijs of de plaatsing in bepaalde leerweg in het voortgezet onderwijs, diepgaand te onderzoeken vanuit het perspectief van de direct betrokkenen: de leerling, de leraar, en de ouders. Hoe heeft de leerling het advies ervaren, wat was het dat hem deed "kiezen" voor die leerweg en welke rol heeft de leraar daarin gespeeld met zijn verwachtingen? Zie ook de prachtige casestudy van Weinstein (2002) over de 'power' van leraarsverwachtingen op de ontwikkeling van leerlingen. Zij beschrijft in detail, van geboorte tot in de volwassenheid, het ontwikkelingsproces van Adam, de "helft" van een te vroeg geboren tweeling die met geboorteletsel ter wereld kwam. Zij beschrijft hoe Adam in een systeem van niveaugroepjes bij lezen was geplaatst. Daarbinnen vormde Adam eenzaam en alleen de laagste "leesgroep". De lage verwachtingen

werden van de ene leraar op de andere overgedragen. Adam had het geluk dat zijn ouders weigerden te luisteren naar het oordeel van het hoofd van de basisschool, en dat zij hem verschillende keren overplaatsten; zo kwam hij in contact met nieuwe leraren die hoge verwachtingen van hem koesterden. Steeds opnieuw gingen er deuren voor hem open. Ten slotte voltooide hij met groot succes zijn universitaire studie. Hij weersprak daarmee in persoon de profetie van het hoofd van de basisschool: "not-so-bright, not college material", maar Adam had dan ook een moeder die Rhona Weinstein heet.

5.3 Design experiments

Binnen een breed-spectrummethodologie passen ook en vooral design experiments.⁴ Zie ook onze ontwerpbenadering op het gebied van curriculumonderzoek die wij hebben aangeduid als een 'principle approach' (Terwel & Walker, 2004). Als deze experimenten vanuit een goede theoretische basis zijn opgezet, kunnen zij ook een bijdrage leveren aan het voorkomen van de negatieve effecten van leraarsverwachtingen. Daarbij is onder meer te denken aan onjuiste opvattingen en voorbarige oordelen van leraren over ontwikkelingsmogelijkheden van categorieën van leerlingen. Design experiments maken voor leraren concreet hoe belangrijk het is hoge en adequate verwachtingen van de leerlingen te hebben en hoe men die verwachtingen kan vertalen in de klas. Als uitgangspunt denk ik vooral aan theorieën waarin de *ontwikkeling van alle leerlingen* wordt benadrukt, zoals in het onderzoek van Carol Lee (2001). Als vertrekpunt worden realistische contexten gekozen en van daaruit wordt een brug gebouwd naar nieuwe, meer abstracte, formele inzichten. In dit proces zou men moeten zoeken naar activiteiten die een leerling alleen haast niet kan volbrengen, maar wel met hulp van de leraar, van medeleerlingen of van speciale 'tools' (de zone van de naaste ontwikkeling). Dit proces is aan te duiden met metaforen als *guided reinvention*, *guided co-construction*, *'scaffolding'* en *'cultural modeling'*. Carol Lee (2001) beschrijft hoe dergelijke kernbegrippen in concrete klassituaties kunnen functioneren, bijvoorbeeld bij literatuuronderwijs (vgl. ook Lowyck & Terwel, 2003).

6 Implicaties voor theorie en praktijk

Het lijkt mij van belang dat er een wetenschappelijke discussie plaatsvindt over de theoretische achtergronden en de gebruikte analysemodellen (Opdenakker, 2003). Verschillende modellen genereren verschillende uitkomsten. Illustratief zijn de verschillen in de multilevel-modellen van Westerbeek (1999) en die van Jungbluth (2003). Onze eigen modellen en uitkomsten, vooral op het gebied van het voortgezet onderwijs, liggen veel meer in de lijn van Westerbeek (1999) en Opdenakker (2003). Wij gaan evenals Westerbeek uit van een resource-theoretisch perspectief en kapitaliseren meer op de cognitieve hulpbronnen van individuele leerlingen en die van de klas (Brekelmans, et al., 1997; Terwel, et al., 2001; zie ook noot 3). De uitkomsten van beide typen modellen geven ook aanleiding tot verschillende aanbevelingen voor de praktijk. Jungbluth legt meer de nadruk op het veranderen van de verwachtingen van de leraren, en lijkt minder zwaar te tillen aan de cognitieve compositie van de klas. Westerbeek, daarentegen, pleit voor een meer evenwichtige leerlingpopulatie op school- en klasniveau. Kort samengevat is de vraag: Leraren veranderen, of bevorderen dat er een meer evenwichtige verdeling van leerlingen over scholen en klassen ontstaat?

Jungbluth neemt, mogelijk vanuit zijn preoccupatie met leraarsverwachtingen, in interviews soms de provocerende stelling in dat de kleur van de school er niet toe doet: "Een compleet zwarte school, so what!" (Jungbluth, 2003b). Onbedoeld komt hij daarmee gevaarlijk dicht in de buurt van de neo-conservatieve slogan "separate but equal". Uiteraard doet die kleur er ook niet toe, maar het wordt echt anders als de cognitieve resources (bijvoorbeeld taal en wiskunde) ongelijk zijn verdeeld over witte en zwarte scholen. Die verschillen in resources zijn voor een belangrijk deel herleidbaar tot categoriaal bepaalde ervaringen van leerlingen in hun leefwereld (Tilly, 1998). Daarom lijkt het van belang dat leerlingen niet in hun 'niche' worden opgesloten, maar in aanraking komen met bredere aspecten van de cultuur dan die waarmee zij vertrouwd zijn vanuit hun thuissituatie en buurt (Terwel, 2002).

Het is dan ook bedenkelijk, zowel uit juridisch als uit meritocratisch oogpunt, dat er bijzondere scholen zijn die quota hanteren bij de toelating van leerlingen met een buitenlandse achtergrond of een andere levensbeschouwing. Ook komt het voor dat scholen hun religieuze identiteit accentueren door het introduceren van religieuze symbolen en gebruiken, en daardoor bepaalde categorieën leerlingen buiten de deur houden (NRC Handelsblad, 2003). Ook openbare scholen hanteren strategieën om invloed uit te oefenen op de instroom van hun leerlingen, bijvoorbeeld door het instellen van "profielklassen" met accentueringen op expressievakken of sport. In combinatie met "testlessen" resulteert dat in een proces van "natuurlijke selectie" (Van Kilsdonk, 2003). Meer in het algemeen leidt het marktmechanisme ertoe dat scholen hun leerlingen uitkiezen en dat er voor bepaalde ouders en leerlingen weinig te kiezen valt. Hiermee raken we aan het vraagstuk van segregatie.

Ook al heeft hij dat in zijn eigen onderzoek nog niet overtuigend aangetoond, toch heeft Jungbluth een sterk punt als hij in zijn rapport zegt dat de bestrijding van segregatie nodig is, maar dat kansongelijkheid ook zonder segregatie tot stand kan komen via leraarsverwachtingen (Jungbluth, 2003, pag. 176).

In mijn discussiebijdrage heb ik geprobeerd aan te tonen dat het gaat om leraren *en* leerlingen in hun onderlinge interactie, want leerlingen hebben ook verwachtingen! Ik zie met belangstelling uit naar het aangekondigde vervolgonderzoek van collega Jungbluth. Het betreft een relevant en interessant onderzoeksthema dat ook verwantschap heeft met ons onderzoek.

7 Slotopmerkingen

Mijn motivatie om deze discussiebijdrage te leveren komt voort uit mijn waarneming dat het anno 2003 niet meer zo erg wordt gevonden om veel laagpresterende leerlingen bij elkaar in één klas te zetten. Die opinie wint veld waar het gaat om de basisschool en de eerste jaren van het voortgezet onderwijs. Het maakt toch nauwelijks verschil zegt men. De toverwoorden van het huidige beleid zijn

tracking, ontvlechting, profilering en natuurlijke selectie: 'separate but equal'. Onbedoeld lijkt deze opinie nu ook enige steun uit onderzoek te krijgen. Zo concluderen Peetsma, Van der Veen, Koopman & Van Schooten (2003, p. 354): "Het lijkt er niet op dat veel doelgroep leerlingen van het onderwijsachterstandenbeleid bij elkaar in één klas, de vooruitgang in hun prestaties noch hun sociaal-emotioneel functioneren duidelijk belemmert". Helaas heeft men niet gecontroleerd voor klassengrootte waardoor een vergelijking met klassen met veel hoogpresterende leerlingen niet opgaat omdat die doorgaans aanzienlijk groter zijn. Ook is er niet gekeken naar de cognitieve resources in de klas zoals geïndiceerd door het klasgemiddelde op een kennistoets. En aan de pedagogisch-didactische interactie in de klas is ook in deze studie geen aandacht geschonken. De black-box studies van Jungbluth en Peetsma roepen veel vragen op naar de onderbouwing van hun conclusies. Hebben Peetsma e.a. wel effecten van klascompositie gemeten of zijn het effecten van klasgrootte? Die vragen moeten worden gesteld omdat hun conclusies niet overeenkomen met de uitkomsten van vele andere onderzoeken naar de effecten van cognitieve klascompositie. De algemene conclusie die ik uit vele onderzoeken meen te mogen trekken is deze. De samenstelling van de klas naar gemiddeld niveau op wiskunde en taal heeft invloed op de leerwinst van alle leerlingen in die klas. Dat geldt wellicht in het bijzonder voor de zwakke leerlingen. Daarom lijkt het mij niet goed om alle laagpresterende leerlingen bij elkaar in één klas te zetten. Misschien moeten wij onderzoekers wat vaker in scholen en klassen kijken om met eigen ogen te zien hoe klascompositie doorwerkt in interactieprocessen en leerresultaten om te beginnen in het vmbo.

Noten

- 1 Met dank aan dr. Pieter van den Eeden, dr. Bert van Oers en dr. Wim Wardekker voor hun kritische opmerkingen bij een eerdere versie van deze tekst.
- 2 Vgl. het themanummer uit 2002 van *International Journal of Educational Research. Organization of Schooling and Peer Influences on Learning*, 37(5).

- 3 Geciteerd uit de review van M. Thrupp, H. Lauder, & T. Robinson (2002), *International Journal of Educational Research. Organization of Schooling and Peer Influences on Learning*, 37(5), 490.
- 4 In een reeks opeenvolgende onderzoeken in Nederlandse scholen voor voortgezet onderwijs vonden wij ook deze effecten van klas-samenstelling, alsmede de differentiële effecten voor sterke en zwakke leerlingen. Zie Terwel, J. (2004). Curriculum and Curriculum Differentiation. In J. Terwel & D. Walker (Eds.), *Curriculum as a shaping force. Toward a principled approach in curriculum theory and practice* (pp. 33-50). New York: Nova; en Terwel, J. (2003). Cooperative learning in secondary education: a curriculum perspective. In R. M. Gillies & A. F. Ashman (Eds.), *Cooperative Learning: The social and intellectual outcomes of learning in groups* (pp. 54-68). London and New York: RoutledgeFalmer.
- 5 Zie het themanummer over design experiments: *Educational Researcher*, 32(1).

Literatuur

- Brands, J. (1992). 'Die hoeft nooit meer wat te leren'. *Levensverhalen van academici met laaggeschoolde ouders*. Nijmegen: SUN.
- Brekelmans, M., Van den Eeden, P., Terwel, J., & Wubbels, Th. (1997). Student characteristics and learning environment interactions in mathematics and physics education: A resource perspective. *International Journal of Educational Research*, 27(4), 283-292.
- Broer, T. (2000). Het zwartste gymnasium van Nederland. *Vrij Nederland*, 61(49), 24-27.
- Bryk, A. S., & Raudenbush, S. W. (1992). *Hierarchical linear models. Applications and Data Analysis Methods*. Newbury Park London New Delhi: Sage.
- De Fraine, B. (2003). *Cognitieve en niet-cognitieve effecten van prestatiegericht klimaat, gemeenschapsgericht klimaat en groepssamenstelling in klassen en scholen*. Academisch proefschrift, Katholieke Universiteit Leuven, Faculteit Psychologie en Pedagogische Wetenschappen.

- Guardini, R., & Bollnow, O. F. (1962). *Begegnung und Bildung. Weltbild und Erziehung: Eine Schriftenreihe, 12.* Würzburg: Werkband Verlag.
- Hauser, R. M. (1970). Context and consex: A cautionary tale. *The American Journal of Sociology, 75*, 645-664.
- Holmbeck, G. N. (1997). Toward terminological, conceptual, and statistical clarity in the study of mediators and moderators: Examples from the child clinical and pediatric psychology literatures. *Journal of Consulting and Clinical Psychology, 65*, 599-610.
- Jungbluth, P. (2003a). *De ongelijke basisschool. Etniciteit, sociaal milieu, sekse, verborgen differentiatie, segregatie, onderwijskansen en schooleffectiviteit.* Nijmegen: ITS.
- Jungbluth, P. (2003b). Een compleet zwarte school, so what! *Schooljournaal, 7*(20), 10.
- Kilsdonk, M. van. (2003). De wedergeboorte van een 'zwarte school'. *De Humanist, 57*(5), 12-15.
- Lee, C. D. (2001). Is october brown chinese? A cultural modeling activity system for under-achieving students. *American Educational Research Journal, 38*(1), 97-141.
- Lowyck, J., & Terwel, J. (2003). Ontwerpen van leeromgevingen. In N. Verloop & J. Lowyck (Reds.), *Onderwijskunde. Een kennisbasis voor professionals* (pp. 284-329). Groningen/Houten: Wolters-Noordhoff.
- Mason, D. A., Schroeter, D. D., Combs, R. K., & Washington, K. (1992). Assigning average-achieving eight graders to advanced mathematics classes in urban junior high. *Elementary School Journal, 92*, 307-322.
- Opendakker, M. C. (2003). *Leerling in Wonderland? Een onderzoek naar het effect van leerling-, lesgroep-, leerkracht- en schoolkenmerken op prestaties voor wiskunde in het secundair onderwijs.* Academisch proefschrift, Katholieke Universiteit Leuven, Faculteit Psychologie en Pedagogische Wetenschappen.
- Pedhazur, E. J. (1982). *Multiple Regression in Behavioral Research. Explanation and Prediction.* New York: CBS College Publishing.
- Religieuze symbolen zijn terug op katholieke scholen. (2003, 22 November). *NRC Handelsblad*, p. 1.
- Terwel, J., Gillies, R. M., Eeden, P. van den, & Hoek, D. (2001). Co-operative learning processes of students: a longitudinal multilevel perspective. *British Journal of Educational Psychology, 71*, 619-645.
- Terwel, J. (2002). Overheid en Markt in het onderwijs. Klaarstomen voor het bedrijfsleven of leren denken? *Vernieuwing. Tijdschrift voor Onderwijs en Opvoeding, 61*(10), 8-12.
- Terwel, J. & Walker, D. (2004). *Curriculum as a shaping force. Toward a principal approach in curriculum theory and practice.* New York: Nova.
- Tilly, C. (1998). *Durable Inequality.* Berkeley: University of California Press.
- Rosenthal, R., & Jacobson, L. (1968). *Pygmalion in the classroom: Teacher expectation and pupil's intellectual development.* New York: Holt, Rinehart and Wilston.
- Rosenthal, R. (1987). Pygmalion effects: Existence, magnitude, and social importance. *Educational Researcher, 16*(9), 37-41.
- Weinstein, R. S. (2002). *Reaching higher. The power of expectation in schooling.* Cambridge: Harvard University Press.
- Westerbeek, K. (1999). *The colours of my classroom. A study into the effects of ethnic composition of classrooms on the achievement of pupils from different ethnic backgrounds.* Florence: European University Institute.
- Wineburg, S. S. (1987). The self-fulfillment of the self-fulfilling prophecy. *Educational Researcher, 16*(9), 28-37.

Manuscript aanvaard: 2 december 2003

Auteur

Jan Terwel is hoogleraar Onderwijspedagogiek aan de Faculteit der Psychologie en Pedagogiek van de Vrije Universiteit Amsterdam. Hij is hoofd van de afdeling Onderwijspedagogiek en programmaleider van het onderzoeksprogramma "Strategisch leren in het curriculum".

Correspondentieadres: Prof. dr. J. Terwel, Afdeling Onderwijspedagogiek, Faculteit der Psychologie en Pedagogiek, Vrije Universiteit, Van der Boechorststraat 1, 1081 BT Amsterdam, e-mail: J.Terwel@psy.vu.nl