

### **Tripliek: een onnodig gecompliceerde voorstelling van zaken<sup>1</sup>**

R. J. Bosker

### Samenvatting

In hun dupliek als volgende stap in de discussie over *evidence based* onderwijs, betogen Gravemeijer en Kirschner (2008) dat ik een essentieel onderscheid tussen procesgerichte en resultaatgerichte causaliteit over het hoofd zie, dat ik ze niet heb kunnen overtuigen dat echte experimenten in het onderwijs uitvoerbaar en betaalbaar zijn, en dat ik een confectie- in plaats van een maatwerk-aanpak proclameer. In deze tripliek wordt het standpunt verdedigd dat het onderscheid in typen causaliteit de zaak onnodig compliceert, dat de bewering dat iets moeilijk is, niet wel zeggen dat het niet kan geschieden, en dat de taak van de onderwijskunde als toegepaste discipline inderdaad de levering van confectiewerk betreft waarvan we weten dat het gunstige effecten bij leerlingen teweeg brengt. Voorts wordt betoogd dat niet beweerd is dat evidentie alleen tot stand kan komen door experimenten met aselechte toewijzing van leerlingen, leerkrachten en scholen aan controle en experimentele condities.

### 1 Vooraf

In hun dupliek zoemen Gravemeijer en Kirschner (2008) in op drie elementen in onze discussie: opvattingen over causaliteit, haalbaarheid en betaalbaarheid van het echte experiment, en het belang van de leraar. Ik ga in deze tripliek op hun betoog ter zake in. Voor dat ik dat doe, wil ik eerst even een stap terug zetten. De discussie spitst zich namelijk nu geheel en alleen toe op het gebruik van *cluster randomized controlled trials* om oorzaak-gevolg relaties in het onderwijs op het spoor te komen. In het advies “Naar meer evidence based onderwijs” (Onderwijsraad, 2006), de

aanleiding tot onze gedachtewisseling, is echter een genuanceerder beeld neergezet over kennisaccumulatie. In dat verband het volgende citaat uit dit advies (p. 20):

Een oordeel ‘bewezen werking’ kan op verschillende manieren tot stand komen. De hardste vorm van bewijs met veel bouwstenen bestaat uit gecontroleerde experimenten met aselechte toewijzing van proefpersonen aan condities. Andersoortige onderzoeksopzetten kunnen zachter bewijs opleveren. Te denken valt onder meer aan pilots, cohortstudies, en gevalstudies. Naast deze vormen van (hardere en zachtere) geobjectiveerde kennis kan ook ervaringskennis van professionals een bijdrage leveren aan inzicht in ‘wat werkt’.

Bovendien wordt in het advies uitdrukkelijk gesteld dat het om accumulatie van kennis gaat, en dat reviews en meta-analyses de (voorlopig) laatste stap in deze keten vormen (zie ook: Coalition for Evidence Based Policy, 2003). Overigens kunnen we aan dat geheel ook nog toevoegen het regressie-discontinuïteit-design, dat mits verstandig toegepast en onderzocht op het voldaan zijn van de assumpties, ook oorzaak-gevolgconclusies toelaat (Shadish, Cook, & Campbell, 2002; voor een elegante toepassing zie Luyten, 2006). En voorts heb ik in mijn reactie (Bosker, 2008) duidelijk gesteld dat niet alles zich door experimenten laat onderzoeken, waarbij ik als voorbeeld de werking van geïntegreerde versus gedifferentieerde onderwijssystemen noemde. Toch zal het duidelijk zijn dat kwantitatieve gegevens over de prestaties van beide stelsels onontbeerlijk zijn om daar inzicht in te verkrijgen. Uiteindelijk zit de kern van het verschil tussen de positie van Gravemeijer en Kirschner enerzijds en hetgeen ik voorsta anderzijds hierin, dat ik uiteindelijk zo hard mogelijk bewijs op tafel wil hebben dat het onderwijs werkt voor leerlingen, met andere woorden dat zij beter van allerlei innovaties worden (of, op zijn minst, niet slechter).

## 2 Verwarrende opvattingen over causaliteit

Causaliteit betekent niets anders dan oorzakelijkheid; een betrekking van oorzaak en gevolg. Ik persisteer in mijn opvatting dat het geen enkel nut heeft te weten *hoe* iets werkt als je niet weet *dat* het werkt. Het onderscheid tussen procesgerichte causaliteit en resultaatgerichte causaliteit dat Gravemeijer en Kirschner in navolging van Maxwell maken, doet daar niets aan af; sterker nog het voegt daar helemaal niets aan toe en leidt alleen maar tot spraakverwarring. Belangrijk in onderwijsonderzoek is de notie dat iets pas bewezen causaal werkt, als het beoogde effect bij leerlingen ook daadwerkelijk is aangetoond. Dit standpunt vloeit voort uit de overtuiging dat onderwijskunde een toegepaste discipline is, bedoeld om het onderwijs te optimaliseren, ofwel om leerlingen zich gunstiger (en/of eventueel: doelmatiger) te laten ontwikkelen. Ik kan me voorstellen dat onderwijswetenschappers processen bestuderen (bijvoorbeeld hoe heterogeen groeperen van leerlingen mogelijkwerijs tot prestatieverbetering kan leiden; of hoe een gecontextualiseerde aanpak van het reken- en wiskundeonderwijs in meer begrip kan resulteren), en dat dat leidt tot inzicht in hoe het een het ander beïnvloedt. Maar de *proof of the pudding is in the eating*, namelijk de demonstratie van het effect bij leerlingen.

Interessant is het TIMSS Video Studies voorbeeld dat Gravemeijer en Kirschner aanhalen. Hun bewering luidt dat de bestudering van die video's tot kwaliteitsverbetering heeft geleid. Wat is daarvoor het bewijs? Overigens haalt Raudenbush (2005) in zijn betoog ander onderzoek aan, waaruit blijkt dat men eerst op basis van de kwantitatieve TIMSS-gegevens zag dat Japanse leerlingen een hoog vaardigheidsniveau bereikten, dat men tevens een trend zag in de gegevens dat in landen waar de instructie in het algemeen meer coherent was en meer diepgang vertoonde, leerlingen beter presteerden, en dat men de typering van deze succesvolle instructie bevestigd zag in de video's die waren opgenomen. Eerst zag men in dit geval *dat* er iets werkte, en daarna *hoe* het werkte, en niet omgekeerd.

## 3 Over de uitvoerbaarheid en betaalbaarheid

Bij de zogenaamde *cluster randomized controlled trial* worden niet alleen leerlingen aselekt toegewezen aan de verschillende condities, maar ook scholen of leerkrachten. Alle verschillen tussen de experimentele en controle-groep zijn dan toevallig en verwaarloosbaar (Rubin, 1974), zodat verschillen in de uitkomsten toe te wijzen zijn aan de interventie. Gravemeijer en Kirschner herhalen hun punt dat dit een welhaast onuitvoerbaar en onbetaalbaar onderzoeksontwerp is. Ze vragen zich af wat ik met een 'reguliere setting' bedoel, als ik het heb over de controleconditie. Het antwoord luidt: definiëer helder de populatie (met of zonder inbegrip van vernieuwingsscholen, elitescholen, 'zwarte' scholen) en stel hieruit de controlegroep samen. Grote aantallen zijn zeker nodig, maar zeer grote aantallen alleen als er een gegede vermoeden is dat de interventie onder bepaalde condities (bijvoorbeeld 'witte' scholen) wel en onder andere condities (bijvoorbeeld 'zwarte' scholen) niet werkt. Ter geruststelling, dergelijke interactie-effecten worden in het gros van het onderwijsonderzoek niet gevonden en zo al, dan vallen ze qua betekenis meestal in het niet bij het hoofdeffect van de interventie.

Dan worden enkele bezwaren tegen randomisatie genoemd. Welke leerkracht wil nu twee verschillende methoden hanteren in twee verschillende klassen, zo luidt hun eerste bezwaar. Dat hoeft echter helemaal niet, sterker nog, dat is niet de bedoeling. Bij een echt *crossover design* hanteert de ene groep leerkrachten methode A gedurende een bepaalde periode en dan methode B, terwijl voor een andere groep leerkrachten het omgekeerde geldt. Bij een *crossover design* op het niveau van de school hanteren in een bepaalde jaargroep bij de ene groep scholen de leerkrachten methode A en bij de andere groep scholen hanteren leerkrachten uit die jaargroep methode B, waarna het daaropvolgend schooljaar in de volgende jaargroep de zaak wordt omgewisseld (en dan hebben we het dus in de regel over andere leerkrachten). Het gevaar van een *crossover*-effect, dat vervolgens als bezwaar wordt genoemd, lijkt me

niet echt reëel: zouden de leerkrachten hun methodes gaan wisselen? Tenslotte wordt de vraag opgeworpen of er genoeg bereidwillige leerkrachten gevonden kunnen worden voor het onderzoek. Dat is inderdaad een zorgpunt. Meer in het algemeen maak ik me ernstig zorgen over de bereidheid van scholen en leerkrachten om aan onderwijsonderzoek mee te doen, vermoedelijk overbelast als ze zijn door alle studenten en instellingen die om wat voor legitieme reden dan ook hen benaderen voor deelname aan hun onderzoek. En zijn zij in staat om te beoordelen wat het wetenschappelijk gezien betere onderzoek is? De commissie Parlementair Onderzoek Onderwijsvernieuwingen (in de media: de commissie Dijsselbloem, 2008) beveelt aan dat deelname van scholen aan peilingsonderzoek verplicht moet worden gesteld. Ik ben van mening dat deze aanbeveling ruimer getrokken moet worden: deelname van scholen en leerkrachten aan onderwijsonderzoek moet verplicht gesteld worden in die gevallen waarin overheidsbeleid wordt geëvalueerd, de stand van zaken in het onderwijs op verzoek van de overheid wordt gemonitord of gepeild, onderzoek plaatsvindt dat gefinancierd wordt met van overheidswege beschikbaar gesteld geld voor fundamenteel of toegepast onderwijsonderzoek (in casu gefinancierd door de Programmaraad voor het Onderwijsonderzoek van NWO of de programmacommissie Beleidsgericht Onderzoek Primair Onderwijs van NWO). Die uitgaven voor onderwijsonderzoek worden immers gedaan om onze kennis over het onderwijs en onze kennisbasis over wat er waarom werkt in het onderwijs te vergroten, zodat uiteindelijk de overheid alsmede ook scholen en leerkrachten daar hun voordeel mee kunnen doen.

Een misverstand en enige instemming over betaalbaarheid tenslotte: ik heb niet gezegd dat het geld dat bespaard kan worden op de overheidsuitgaven aan onderwijs, als we eenmaal weten hoe zaken doeltreffend en doelmatig kunnen worden aangepakt, ten goede zou moeten komen aan het onderwijsonderzoek – met Gravemeijer en Kirschner zou ik daar overigens geen bezwaar tegen hebben. Dat er relatief weinig in het onderwijsonderzoek in het algemeen wordt geïn-

vesteerd, ben ik zonder meer met ze eens. Verhoudingsgewijs loopt Nederland allesbehalve voorop als het gaat om de uitgaven aan onderwijsresearch.

#### 4 De leraar als (les)boer?

Het voorbeeld dat Gravemeijer en Kirschner in hun eerste reactie op het advies van de Onderwijsraad noemden, namelijk dat van boeren die kunstig ontworpen kunstmest verkeerd gebruikten, omdat ze de bijgaande handleiding niet goed lazen, verleidde mij tot de bewering dat leraren geen domme boeren zijn. Allicht dat leraren die niet bij de ontwerp- en beproevingsfase van een interventie betrokken zijn geweest, en die geen voorstander van een bepaalde onderwijsverandering zijn, anders met zo'n interventie omgaan. Maar als we hebben kunnen bewijzen dat die interventie werkt, en dat leerlingen er dus echt van profiteren, dan getuigt het niet van professionaliteit als een leerkracht zijn of haar leerlingen deze 'zegen' onthoudt. Het lijkt me ondoenlijk dat we het onderwijs op de maat van elke individuele leerkracht kunnen ontwerpen: dan zijn er net zo veel onderwijskundigen als leerkrachten nodig. We bieden confectiewerk aan in de vorm van methodes, didactieken, inzichten over klasmanagement en dergelijke. Dan is het vervolgens aan de kundige leerkracht om voor de benodigde *finetuning* in zijn of haar onderwijssituatie zorg te dragen.

#### Noot

- 1 In het kader van de discussie rond het rapport "Naar meer evidence based onderwijs" van de Onderwijsraad (2006) zijn hiervoor in *Pedagogische Studiën* achtereenvolgens de discussiebijdragen van Gravemeijer en Kirschner (2007), Bosker (2008) en Gravemeijer en Kirschner (2008) geplaatst. De redactie dankt de auteurs voor hun waardevolle bijdragen en beschouwt met deze laatste bijdrage van Bosker de discussie als gesloten.

## Literatuur

- Bosker, R. J. (2008). Naar meer evidence based onderwijs! *Pedagogische Studiën*, 85, 49 - 51.
- Coalition for Evidence Based Policy. (2003). *Identifying and implementing educational practices supported by rigorous evidence: A user friendly guide*. Washington: United States Department of Education.
- Commissie Parlementair Onderzoek Onderwijsvernieuwingen. (2008). *Tijd voor onderwijs. Eindrapport*. Den Haag, Nederland: SDU.
- Gravemeijer, K. P. E., & Kirschner, P. A. (2007). Naar meer evidence based onderwijs? *Pedagogische Studiën*, 84, 463 - 471.
- Gravemeijer, K. P. E., & Kirschner, P. A. (2008). Dupliek: een te simpele voorstelling van zaken. *Pedagogische Studiën*, 85, 195 - 197
- Luyten, H. (2006). Het effect van een jaar onderwijs op de wiskunde prestaties in groep 5 en 6 en de verschillen tussen scholen. *Pedagogische Studiën*, 83, 432 - 451.
- Onderwijsraad (2006). *Naar meer evidence based onderwijs*. Den Haag, Nederland: Onderwijsraad.
- Raudenbush, S. W. (2005). Learning from attempts to improve schooling: The contribution of methodological diversity. *Educational Researcher*, 34(5), 25 - 31.
- Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, 66, 688 - 701.
- Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston / New York: Houghton Mifflin company.

Manuscript aanvaard: 29 mei 2008

## Auteur

**Roel J. Bosker** is hoogleraar onderwijskunde en directeur van het GION, Gronings Instituut voor Onderzoek van Onderwijs van de Rijksuniversiteit Groningen. Tevens is hij lid van de Onderwijsraad.

*Correspondentieadres:* Roel Bosker, GION, Rijksuniversiteit Groningen, Grote Rozenstraat 3, 9712 TG Groningen, Nederland, e-mail: [R.J.Bosker@rug.nl](mailto:R.J.Bosker@rug.nl)

## Abstract

### **Surrejoinder: An unnecessary complicated representation of things**

In their rejoinder "A too simple representation of things" continuing a debate on evidence-based education, Gravemeijer and Kirschner state that I missed their distinction between the process view and the regularity view of causation, that I could not convince them that real educational experiments are feasible, and that I proclaim a confectio- rather than a custom-made approach. In this surrejoinder the standpoint is taken that the distinction between sorts of causation is complicating things unnecessarily, that hard to conduct experiments are still feasible, and that it is indeed the task of educational science to produce confectio-made knowledge on what works for children in education. Moreover, I reiterate that not only cluster randomized controlled trials produce knowledge.