

Naar meer *evidence-based* onderwijs?

K. P. E. Gravemeijer en P. A. Kirschner

Samenvatting

De Onderwijsraad (2006) benadrukt in haar advies "Naar meer evidence based onderwijs" het belang van gerandomiseerd experimenteel onderzoek, zoals dat in de gezondheidszorg wordt toegepast, als de manier om meer *evidence-based* onderwijsbeleid mogelijk te maken. Kennis over interventie-effecten zou daarmee volgens de Onderwijsraad op de meest sterke evidentie berusten. In deze bijdrage stellen we deze gedachte ter discussie. Er zijn volgens ons tal van problemen verbonden aan de rol van gerandomiseerd experimenteel onderzoek in de context van onderwijsinnovatie. We betogen dat een hierop gebaseerde *evidence-based* benadering in deze context niet uitvoerbaar, niet betaalbaar, en te generaliserend is, wel laat zien wat werkt, maar niet hoe 'het' werkt, en bovendien vertragend en conserverend werkt.

1 Inleiding

De Onderwijsraad constateert in haar advies "Naar meer evidence based onderwijs" (Onderwijsraad, 2006) dat in Nederland op het gebied van onderwijsmethoden en -aanpakken weinig gerandomiseerd effectonderzoek (oftewel RCT's, *randomised controlled trials*), maar wel veel ontwikkelings- en evaluatieonderzoek wordt uitgevoerd. Kennis over interventie-effecten berust daarmee volgens de Onderwijsraad niet op de meest sterke evidentie. Volgens de voorzitter van de Onderwijsraad, Van Wieringen "...kunnen [we] een voorbeeld nemen aan de gezondheidszorg. Daar komen medicijnen pas op de markt als de werkzaamheid wetenschappelijk is aangetoond. Het onderwijs kan daar van leren." (Haaft, 2006). De gedachtegang dat het invoeren van een onderwijsverandering

alleen mag plaatsvinden als de verandering bewezen effectief is, is niet nieuw, maar wel actueel dankzij president Bush's *No child left behind*-act. Deze wet houdt onder meer in dat een onderwijsprogramma niet geïmplementeerd mag worden zonder "evidence, based on rigorous research and evaluation, that [it] can effectively help children to succeed in school". Zoiets klinkt heel logisch en eenvoudig, maar is dat ook zo?

Je kunt je afvragen of je deze strategie die in de gezondheidszorg en in de landbouw volgens velen goed werkt – er zijn ook kritische vragen over wat 'goede gezondheidszorg' inhoudt en of een *evidence-based* benadering daarvoor een antwoord kan geven¹ – zomaar naar het onderwijs kunt overplanten. De vraag stellen is haar beantwoorden. Er zijn enkele voorbeelden van situaties waar *randomised controlled trials* goed lijken te hebben gefunctioneerd. Dat is bijvoorbeeld het geval in het project Scaling Up SimCalc, dat een op RCT's gebaseerd onderzoek uitvoerde, waarin 1.621 leerlingen participeerden en waarin werd gerandomiseerd over variabelen zoals geslacht, docent, schoolsetting (stad, platteland), socio-economische status van de lerenden, burgerlijke status (allochtoon Mexicaan, autochtoon Amerikaan), enzovoorts. Aangetoond werd dat de leerlingen die aan het technologieondersteunde curriculum deelnamen door de bank genomen meer leerden dan die in de controlegroep ($p \leq 0,001$ en effectgrootte gelijk aan 0,84); Roschelle, Tatar, Shechtman, Hegedus, Hopkins, Knudsen, & Stroter, 2007).

Maar er zijn volgens ons – en vele anderen, zoals duidelijk zal worden in de rest van dit artikel – tal van problemen verbonden aan een op gerandomiseerd experimenteel onderzoek gebaseerde benadering van onderwijsinnovatie. In dit artikel staat de idee centraal van de op gerandomiseerd experimenteel onderzoek gebaseerde benadering van de onderwijsverbetering, zoals die zonder voorbehoud door de *No child left behind*-act gepropageerd wordt. We tekenen daarbij aan dat de Onderwijsraad in haar advies een wat genuan-

ceerder standpunt inneemt, maar dat de uitwerking uiteindelijk heel dicht bij die van *No child left behind* ligt. We komen daar aan het eind van dit artikel op terug. De problemen van op gerandomiseerd experimenteel onderzoek gebaseerde benadering van onderwijsverbetering kunnen we samenvatten met:

- niet uitvoerbaar;
- niet betaalbaar;
- te generaliserend;
- laat zien wat werkt, niet hoe ‘het’ werkt, en
- werkt vertragend en conserverend.

Voor we hieronder op deze vijf probleem-punten ingaan, willen wij echter eerst de term *evidence based* ter discussie stellen.

2 Wat is wetenschappelijke *evidence*?

Er lijkt sprake te zijn van een verabsolutering van gerandomiseerd experimenteel onderzoek als *golden standard* voor het verkrijgen van bewijs. In de Verenigde Staten is er nu het What Works Clearinghouse, dat op federaal niveau vaststelt welk onderwijsonderzoek de toets van de wetenschappelijke kritiek kan doorstaan en welk niet. Het criterium hier is eenvoudig, namelijk gerandomiseerd experimenteel onderzoek. Op de AERA 2006 werd een voorbeeld gegeven van een onderzoeksgebied waar van de 11.000 onderzoeken die gevonden werden, slechts 12 waren gebaseerd op *randomised controlled trials*. Het What Works Clearinghouse neemt daarom alleen die 12 onderzoeken in beschouwing. Andere typen onderzoek worden hiermee gediskwalificeerd als onderzoek dat geen wetenschappelijk evidentie oplevert. Op deze manier zou ook onze kennis over de evolutietheorie als niet-wetenschappelijk moeten worden afgedaan, omdat die immers niet is gebaseerd op *randomised trials*.

In het boek, “The hedgehog, the fox, and the magister’s pox”, waarin Gould (2004) de verschillen tussen de natuurwetenschappen en de humaniora bespreekt, betoogt hij dat de methoden die de evolutiebiologie gebruikt om wetenschappelijke uitspraken te funderen niet alleen gelijkwaardig zijn aan de metho-

den van experimenteel onderzoek, maar ook van wezenlijk belang zijn als aanvulling op experimenteel-toetsend onderzoek. Om de methode van de evolutiebiologie te verduidelijken, gebruikt hij de door Whewell (1840) geïntroduceerde term *consilience* – een begrip dat hij omschrijft als *jumping together*. Het gaat daarbij om het plotseling optredende inzicht dat het mogelijk maakt om – tot daarvoor – onsamenhangende feiten met elkaar in verband te brengen². Gould voert Darwins “Origin of species” op om de kracht en effectiviteit van *consilience* aan te tonen:

“Darwin constructed the Origin of Species as a brief for evolution by consilience. In short, he argues: I present you in this book, with thousands of well-attested facts drawn from every subdiscipline of the biological sciences (...). Only one conclusion about the causes and changes of life – the genealogical linkage of all forms by evolution – can possibly coordinate all these maximally various items under a common explanation. And that common explanation must, at least provisionally, be granted the favor of probable truth” (Gould, 2004, pp. 211-212).

De verklaring die Darwin biedt heeft het karakter van een redenering die beschrijft hoe de evolutie in zijn werk is gegaan. Deze redenering steunt op een rijke verzameling empirische gegevens, die samen de evidentie bieden voor deze redenering. Naast het aspect *jumping together*, van puzzelstukjes die op hun plaats vallen, omvat *consilience* ook het aspect van het begrijpen van hoe een en ander werkt. Dit laatste raakt aan wat Maxwell (2004) procesgerichte causaliteit noemt. Naast regelmatigheidsgerichte causaliteit, waarbij causaliteit wordt verbonden met waargenomen regelmatigheden, onderscheidt hij procesgerichte causaliteit. In de procesgerichte benadering wordt causaliteit geconcipieerd als inzicht in de causale mechanismen die ervoor zorgen dat een bepaalde combinatie van omstandigheden tot bepaalde gevolgen leidt. Het gaat dus om het begrijpen hoe iets werkt. Anders gezegd, naast wetenschappelijke evidentie voor causale relaties op basis van waargenomen regelmatigheden, is er ook sprake van wetenschappelijke evidentie voor procesgerichte causaliteit. Merk

op dat het hier, anders dan bij regelmatigheidsgerichte causaliteit, niet gaat om evidentie voor effectiviteit, maar om evidentie voor bepaalde inzichten, zoals antwoorden op 'hoe-vragen'.

Tegenstanders van *evidence-based education* hebben veel vragen gesteld over hoe geschikt deze aanpak is voor het onderwijs. Sommige stellen vragen over de overeenkomsten (zoals wij dat ook eerder deden) tussen onderwijs, en geneeskunde en farmacie, ook wijzend naar het verschil in betekenis van 'bewijs' in deze gebieden (Davies, 1999; Greenhalgh, Toon, Russell, Wong, Plumb, & Macfarlane, 2003; Pirrie, 2001; Simons, 2003). Anderen stellen vragen over de positivistische vooronderstellingen die ten grondslag liggen aan *evidence-based* onderwijs en bekritisieren de enge conceptie van zulk soort onderwijs (Adams St. Pierre, 2002; Atkinson, 2000; Berliner, 2002; Elliot, 2001; Erickson & Gutierrez 2002; Oliver & Conole, 2003).

3 Niet uitvoerbaar

De *evidence-based* benadering komt uit de gezondheidszorg en vooral de farmaceutische industrie. Zo wordt er beslist of, bijvoorbeeld, een geneesmiddel toegelaten wordt op de markt. Maar men moet oppassen met het zomaar toepassen van een methodologie uit één domein in een ander. In de farmaceutische industrie maakt men gebruik van:

- (1) controlegroepen (die vaak een placebo krijgen) waar deelnemers en behandelende artsen en onderzoekers niet weten wie in welke conditie zit (*double-blind*);
- (2) willekeurig toedelen van deelnemers over condities (randomisatie);
- (3) onderlinge vergelijkbaarheid van de groepen;
- (4) zeer grote groepen, en
- (5) de eis van statistisch significante resultaten.

Deze eisen stelt Slavin (1987), pleitbezorger van deze benadering, aan *evidence-based* onderwijs. Maar kan dit in het onderwijs? Kunnen wij leerlingen in een controlegroep onderwijs geven waarvan wij weten dat het geen of zelfs een negatief effect heeft (een

placebo geven) en hoe moeten wij dit *double-blind* doen? Mogen wij leerlingen in een school of klas willekeurig verdelen over verschillende onderzoekscondities? Kunnen wij de vergelijkbaarheid controleren, zoals een farmaceutisch bedrijf dat doet (bijvoorbeeld, alleen mensen met een bepaald ziektebeeld en een bepaalde leeftijd en gewicht, enzovoorts)? Waar vinden wij voldoende grote aantallen leerlingen om deze randomisatie te realiseren? En wat betekent het als iets statistisch significant is? Ter illustratie, het gemiddelde effectgrootte van levensreddende interventies met geneesmiddelen (zoals geneesmiddelen tegen hoge bloeddruk ter voorkoming van beroertes) ligt tussen de 0,02 en 0,08 (Norman, 2003), terwijl de gemiddelde grootte van het effect van onderwijskundige ingrepen – gebaseerd op een review van 302 meta-analyses van meer dan 14.000 studies van onderwijskundige en psychologische interventies – 0,45 bedraagt³ (Lipsey & Wilson, 1993)! Tussen 7 en 20 keer zo hoog!! Volgens Norman: "As it turns out, the apparent life-saving interventions of medicine aren't all that impressive, when the magnitude of the effect is stacked against the educational interventions" (p. 184).

4 Niet betaalbaar

Wat betreft de financiële aspecten het volgende. Het kost een groot farmaceutisch bedrijf gemiddeld \$ 359 miljoen om één nieuw geneesmiddel op de markt te brengen. Slechts 5 van 5000 (0,1%) 'aspirant'-geneesmiddelen die pre-klinisch worden getest worden ooit op mensen uitgetest en alleen één van die vijf wordt goedgekeurd (Wierenga & Eaton, 2005), dus 0,02%. Alleen in de Verenigde Staten werd in 2005 ruim \$ 38 miljard besteed aan onderzoek (Berenson, 2006). Verder zijn de uitgaven voor onderzoek en ontwikkeling (*Research & Development*) circa een kwart van de inkomsten van zo een bedrijf (Network Science, s.d.). Zou een minister van Onderwijs, Cultuur en Wetenschappen bereid zijn om, naar verhouding, vergelijkbare bedragen vrij te maken voor gedegen onderwijsonderzoek met het risico dat maar 0,02% van de onderzochte onderwijs-

veranderingen of -verbeteringen zullen slagen?

5 Te generaliserend

De methodiek van *randomised controlled trials* is niet zomaar over te planten naar het onderwijs. Als een farmaceutisch bedrijf in *evidence-based* onderzoek een medicijn of placebo toedient, doet het er niets toe of de zon schijnt of het regent, of het toegediend wordt in de keuken of in de badkamer, enzovoorts. De dosering is constant en is niet afhankelijk van wie het toedient en waar het toegediend wordt. Maar onderwijs is contextgebonden. Chatterji (2005) constateert ons inziens terecht: What Works Clearinghouse gaat voorbij aan de *critical realities* over sociale, organisatorische en beleidsomgevingen waarin onderwijsprogramma's en -interventies plaatsvinden. Olsen (2004) stelt verder:

“In medical trials, the treatment is assumed to have an invariant or lawlike causal relation with the symptoms of the disease. In education, such simple causal relations do not obtain between teaching and learning; interactions are filtered through the goals, beliefs, and intentions of the teachers and learners [...]. The more simple cause-effect relations so important to the physical and biological sciences are largely inappropriate to the human sciences, which trade on the beliefs, hopes, and reasons of intentional beings. (p. 25)

Oftewel, mensen functioneren niet volgens wetmatigheden, maar geven betekenis aan situaties.

Eén van de kenmerken van gerandomiseerd experimenteel onderzoek is dat het de onderzoeker in staat stelt uitspraken te doen over groepen. Een probleem hiermee is dat statistisch significante verschillen tussen groepen vaak een zeer beperkte betekenis hebben voor individuele gevallen. Omgekeerd bestaat het gevaar dat er wordt gerandomiseerd over relevante verschillen. Neem bijvoorbeeld het volgende geval. Onderwijsinnovaties willen nog wel eens effectief zijn bij *early adopters*. Volgens de strikte RCT-filosofie, zouden *non-adopters* tot deelname gedwongen moeten worden om vast te kun-

nen stellen of de onderwijsaanpak effectief is. Het gevolg is dat je in de populatie twee aanwijsbaar verschillende groepen hebt, voor wie de onderwijsaanpak wel eens verschillend zou kunnen werken. Het zou echter zo kunnen zijn dat een onderwijsaanpak die effectief is voor de *early adopters*, voor de *non-adopters* niet echt goed werkt.

Dit wordt een serieus probleem, wanneer we veronderstellen dat bepaalde onderwijsaanpakken beter bij bepaalde leraren passen dan andere. Leraren de ruimte laten hun eigen 'stiel' te volgen zou dan wel eens effectiever kunnen zijn dan iedereen te verplichten die ene variant toe te passen die als beste naar voren komt in gerandomiseerd experimenteel onderzoek. Wanneer we ervan uitgaan dat er verschillen tussen leraren, leerlingen en onderwijscondities bestaan is het misschien zinvoller daar rekening mee te houden in plaats van te proberen een effectieve aanpak voor iedereen te vinden. Sterker nog, deze aanpak druist juist in tegen de huidige roep om meer maatwerk in het onderwijs!

Zulk rekening houden met verschillen staat echter haaks op het streven om – binnen gerandomiseerd experimenteel onderzoek – de *treatments* zoveel mogelijk te homogeniseren. Immers, hoe minder variatie in de uitvoering van de onderwijsaanpak, des te beter je weet wat je vergelijkt en des te nauwkeuriger ook het advies is dat je op basis van het onderzoek kunt geven. Wanneer dit te ver wordt doorgevoerd, kom je terecht bij de beruchte *teacher-proof*-leerprogramma's en leerboeken van zestiger en zeventiger jaren van de vorige eeuw. De beperkingen van de RDD benadering en de *teacher-proof* programma's zijn sinds die tijd scherp aan het licht gekomen. Het centrale probleem kunnen wij illustreren aan een andere anekdote (Figuur 1).

De moraal van dit verhaal: Als het gebruik van kunstmest al zo afhankelijk is van de gebruiker, hoe belangrijk moet de rol van de leraar dan wel niet zijn als het gaat om onderwijs? Het gaat uiteindelijk om de toepassing van die onderzoeksresultaten, in casu van de kennis die vergaard is in het onderzoek. In het onderwijs zal het dan vooral gaan om het kunnen inspelen op de specifieke kenmerken van de eigen situatie.

Door Dupont werd een nieuw soort kunstmest ontwikkeld. De werking ervan werd uitgebreid getest in gerandomiseerd experimenteel onderzoek. Toen de effectiviteit was vastgesteld bracht men het op de markt. De agrariërs die het gebruikten waren echter niet tevreden, het leidde niet tot verbetering, maar tot verslechtering. Het onderzoek werd herhaald, de nieuwe kunstmest was toch echt effectief. Het werd opnieuw aan agrariërs aangeboden, het bleek opnieuw niet te werken. Toen werd besloten met de gebruikers te gaan praten. Wat bleek, deze mannen lezen geen handleidingen. Ze hadden de kunstmest zo op de akker gestrooid, terwijl het in de grond verwerkt moest worden. Men ging bij Dupont aan de slag om een antwoord op dit probleem te vinden. Het resultaat was het idee de kunstmest in water op te lossen en het over het land te verspreiden. Maar ook hier waren de boeren niet echt enthousiast over, ja het werkte hier en daar, maar op andere plaatsen weer niet. Op sommige plaatsen was het resultaat zelfs slechter. Men ging weer met de betrokkenen praten en wat bleek, de boeren die de kunstmest met een tank aan hun tractor over het land verspreidden, reden niet met een constante snelheid, wat tot een onregelmatige verspreiding leidde.

Figuur 1. Anekdote Dupont.

6 Laat zien wat werkt, niet hoe 'het' werkt

Om goed te kunnen inspelen op specifieke kenmerken van de eigen situatie moet de leraar inzicht hebben in de werking van de aanbevolen onderwijsaanpak. Ook hier dreigt een gevaar van te eenzijdige nadruk op het effectiviteitsonderzoek (lees: wat werkt!). Aantoonbare effectiviteit van wat werkt of wat niet werkt in een specifieke situatie heeft slechts beperkte betekenis wanneer deze niet gepaard gaat met een concrete theorie over waarom het al dan niet heeft gewerkt. De Research Advisory Committee van de NCTM (1996) wijst in dit verband op het gevaar dat de leraar die ontdekt dat een bepaalde aanpak in zijn of haar klas niet werkt, alleen weet dat het zou moeten werken. Terwijl hij of zij zonder een theorie geen idee heeft waarom de bedoelde aanpak niet heeft gewerkt of hoe de bedoelde aanpak aan zijn of haar situatie zou kunnen of zou moeten worden aangepast. Zij pleiten daarom voor andere *norms of justification*, die we kunnen typeren als empirisch gefundeerde theorieën over hoe de interventie werkt. De theorie over de werking wordt dus gezien als een basis voor de wetenschappelijke onderbouwing van een interventie. Dit idee van een theorie over de werking sluit aan bij Maxwells (2004) notie van procesgerichte causaliteit. De twee door hem onderscheidde vormen van causaliteit, procesgerichte causaliteit en regelmatigheidsgerichte causaliteit kunnen wij ook verbinden met afnemers van onderzoek (Cobb, 2004). Voor beleidsmakers is vooral regelmatigheidsgerichte causaliteit van belang, omdat dit type causaliteit goed bruikbaar is

bij het nemen van beslissingen die het niveau van de individuele klas en de individuele school overstijgen. De vraag is dan, “Wat werkt het beste?”; en daar geeft gerandomiseerd experimenteel onderzoek antwoord op. Voor de leraar is het belangrijker antwoord op de vraag “Hoe werkt het?” te krijgen. Voor haar of hem gaat het daarbij niet alleen om een *sense of ownership*, maar de leraar heeft deze kennis ook nodig om de voorgestelde en onderzochte interventie aan te passen aan haar of zijn situatie.

7 Werkt vertragend en conserverend

Een verder probleem is het tijdsbestek. Eerst moet opgemerkt worden dat zulke soorten onderzoek veel tijd kosten in hun uitvoering. Wil men weten of een methode werkt, met vaak verstrekkende doelen als het bereiken van bepaalde competenties (vergelijk dit met een voorgesteld geneesmiddel en/of therapie om een ziekte te genezen), zal een volledige methode uitgeprobeerd moeten worden. Dit eist langlopend, vaak longitudinaal onderzoek. Het duurt gemiddeld 12 tot 20 jaar vanaf het moment van het bedenken of ontdekken van een mogelijke chemische verbinding of geneesmiddel tot een uiteindelijke goedkeuring door de daartoe bevoegde instanties op basis van *evidence-based* onderzoek (Rosen, 2003). Gek genoeg vinden wij een dergelijke *delay* ook in de onderwijspsychologie, bijvoorbeeld bij het onderzoek in de zestiger jaren van de vorige eeuw naar het gebruik van leerdoelen, toegevoegde vragen, *advance organisers*, enzovoorts. De ‘resultaten’ van dat onderzoek zijn nu niet weg te

denken uit leerboeken vanaf de jaren tachtig. Waarschijnlijk is er geen minister van onderwijs of staatssecretaris te vinden die bereid is 10 tot 20 jaar te wachten en dit uitstel te verdedigen in de Tweede Kamer. De vraag is ook of veel onderwijswetenschappers een dergelijk uitstel verantwoord zouden vinden. Wat overigens onverlet laat dat onderwijsinnovatie zoveel mogelijk op onderzoek zou moeten steunen.

Een laatste beperking van gerandomiseerd experimenteel onderzoek betreft ten slotte het weinig innovatieve karakter van dit type onderzoek. Zulk onderzoek werkt eerder conserverend dan innovierend. Zolang gerandomiseerd experimenteel onderzoek gezien wordt als sluitstuk van een proces van onderzoek en ontwikkeling hoeft dit niet zo'n probleem te zijn. Lastiger wordt het, wanneer, zoals in de Verenigde Staten dreigt te gebeuren, gerandomiseerd experimenteel onderzoek een zodanige status als *golden standard* krijgt, dat dit de enige onderzoeksvorm wordt die serieus wordt genomen. Maar ook in andere gevallen kan gerandomiseerd experimenteel onderzoek conserverend werken. Om dit soort onderzoek te kunnen verrichten moet iedere leraar de beoogde onderwijsaanpak goed kunnen uitvoeren. Echter, wanneer het om een ingrijpende vernieuwing gaat, is dit moeilijk te realiseren (immers, het 'goed' uitvoeren van een vernieuwing vereist vaak specifieke vaardigheden van de docent/uitvoerder en leer- of onderwijsmethodes die 'af' zijn). In deze gevallen zal gerandomiseerd experimenteel onderzoek al gauw achterwege blijven.

Verder zit er een probleem bij de te toetsen doelen. Men streeft in gerandomiseerd experimenteel onderzoek – bij voorkeur – naar eenduidige objectief vast te stellen doelen. Meer innovatieve doelen, echter, zijn vaak moeilijk op die manier te operationaliseren. Ten slotte ontstaat er een probleem van vergelijkbaarheid wanneer je onderzoek wilt doen naar doelen die in de praktijk nog niet worden nagestreefd; controlegroepen zijn dan niet erg relevant. Kortom gerandomiseerd experimenteel onderzoek leent zich beter voor onderzoek naar onderwijsaanpakken die dicht bij de bestaande onderwijspraktijk staan, wat het gevaar met zich meebrengt

dat het zich ook op minder vernieuwende onderwijsaanpakken zal richten.

8 Besluit

In het voorgaande hebben we een aantal problemen naar voren gebracht van het in de Verenigde Staten gepropageerde idee van *evidence-based* onderzoek dat zich beperkt tot gerandomiseerd experimenteel onderzoek. Deze problemen betreffen enerzijds technische problemen, die betrekking hebben op uitvoerbaarheid en betaalbaarheid en, anderzijds, inhoudelijke problemen, die er in het kort op neer komen, dat we bij een voorgestelde onderwijsverbetering niet alleen moeten weten, 'wat werkt', maar ook, 'wanneer', 'voor wie' en 'hoe' het werkt. Onze conclusie is dan ook dat, hoewel wij allen, voor dat wij een verandering in het onderwijs in gang zetten, graag *evidence* willen hebben dat die verandering de moeite waard is, dit type *evidence-based* niet de aangewezen oplossing is.

Een aantal van de hier geschetste problemen wordt overigens ook door de Onderwijsraad genoemd, maar die relativeert ze meteen weer. We noemen enkele voorbeelden. De Onderwijsraad constateert dat uitvoeren van experimenteel onderzoek veel geld kost, maar voegt daar direct aan toe dat het een investering is die op termijn geld spaart, omdat bijvoorbeeld onderwijsinstellingen minder hoeven te investeren in hun eigen ontwikkelingswerk. Ook het probleem van beleidsbeslissingen die niet altijd kunnen wachten op resultaten van jarenlang onderzoek wordt door de Onderwijsraad genoemd, maar, zo wordt daar tegenin gebracht, je kunt werken met tussenresultaten.

In het algemeen neemt de Onderwijsraad, zoals gemeld, een genuanceerder standpunt in dan de *No child left behind*-act. In haar rapport onderscheidt zij verschillende vormen van onderzoek en tekent daar bij aan dat deze onderzoeksmethoden verschillen in de mate van hardheid van bewijs. Zij erkent daarbij dat ook 'zachtere' vormen van onderzoek evidentie opleveren, maar geeft ook aan dat *randomised controlled trials* (RCT) de vorm is die het hardste bewijs levert. Het is

deze hardste vorm van onderzoek waar het betoog van de Onderwijsraad zich in de kern op richt. Echter waar de minister de Onderwijsraad tamelijk open vragen heeft gesteld – “Waar en hoe kan een *evidence-based* benadering bijdragen aan de effectiviteit van het onderwijs?” en “Hoe kan de benutting van *evidence-based* onderwijsmethoden en -aanpakken in de onderwijspraktijk worden gestimuleerd?” (Onderwijsraad, 2006, blz. 13) – richt het advies zich op de rol van evidentie bij het meten van de effectiviteit van onderwijsmethoden. Blijkbaar is de gedachte dat je de kwaliteit van het onderwijs alleen kunt verbeteren door bewezen effectieve onderwijsmethoden of -aanpakken in te voeren. Dit blijkt ook uit de toelichting op de term *evidence-based* beleid (ibid, blz. 14): “In dit advies doelen we daarmee op beleid dat gericht is op de toepassing van bewezen effectieve onderwijsmethoden- of aanpakken.” De Onderwijsraad voegt hieraan toe: “Het gaat hier dus om beleid naar aanleiding van *onderwijsonderzoek* (cursivering origineel).” Maar dat laatste omvat veel meer. Er zijn zoals de Onderwijsraad op een andere plaats ook zelf aangeeft, “...verschillende mogelijke bijdragen van onderzoek aan de kwaliteit van het onderwijs. Dat kan bijvoorbeeld zijn *het aandragen van (innovatieve) concepten, het verschaffen van inzicht in hoe iets werkt, en het aangeven van oplossingen voor praktische problemen of het evalueren daarvan.*” (cursivering auteurs, ibid, blz. 16). In al deze gevallen wordt ons inziens gebruik gemaakt van *evidence*, ofwel empirisch bewijs. De Onderwijsraad echter beperkt *evidence* tot bewijs voor effectiviteit. Dit blijkt met name wanneer zij uitlegt, wat zij onder ‘hard’ en ‘zacht’ bewijs verstaat. Daarmee beperkt de Onderwijsraad ook de rol van andere dan op RCT’s gebaseerde onderzoeksmethoden. Andere onderzoeksmethoden kunnen in de ogen van de Onderwijsraad nuttig en misschien zelfs noodzakelijk zijn in een exploratieve fase, uiteindelijk zal toch bewezen moeten worden dat het werkt. Dat het eerder genoemde ‘inzicht in hoe iets werkt’ bijvoorbeeld, van directe betekenis zou kunnen zijn voor een verbetering van de kwaliteit van het onderwijs, blijft buiten beeld. Dat dit wel degelijk kan hebben we hiervoor al betoogd.

We kunnen dit illustreren met een voorbeeld.

Dit voorbeeld betreft een reeks van gevalstudies, door de onderzoeksgroep van Cobb en collega’s, naar de sociale interactie in de reken-wiskundeles vanuit een socioconstructivistisch perspectief (zie bijvoorbeeld Cobb & Yackel, 1996). Dit onderzoek laat zien, dat de in de klas geldende sociale normen rond wat de leraar en de leerlingen van elkaar mogen verwachten, van doorslaggevend belang zijn voor probleemgeoriënteerd reken-wiskundeonderwijs. Het onderzoek van Cobb en collega’s heeft daar de nodige *evidence* voor opgeleverd. De gangbare norm in traditioneel reken-wiskundeonderwijs is dat de leraar verplicht is om alles zo goed mogelijk uit te leggen en dat de leerlingen de taak hebben te proberen deze uitleg te begrijpen en na te volgen. Bij probleemgeoriënteerd onderwijs wordt echter van de leerlingen verwacht dat ze zelf met oplossingen komen en zelf de kwaliteit daarvan beoordelen. In deze klassen heeft de leraar de verplichting om niet alles uit te leggen, maar om het denken zoveel mogelijk bij de leerlingen te leggen en de discussie te sturen. Uit het onderzoek van Cobb en collega’s blijkt dat sociale normen niet zomaar te veranderen zijn door ze aan de leerlingen te vertellen, “We gaan nu probleemgeoriënteerd werken en nu zijn jullie zelf verantwoordelijk voor het vinden van oplossingen” (zie bijvoorbeeld McClain & Cobb, 2001). De gangbare normen zijn door ervaring gevormd en wanneer je deze wilt veranderen, zullen de leerlingen moeten ervaren dat andere zaken gewaardeerd en gehonoreerd worden. De inzichten die de theorie over *social norms* verschaft, maken duidelijk waarom leerlingen zich verzetten tegen de invoering van probleemgeoriënteerd reken-wiskundeonderwijs en hoe deze weerstand kan worden overwonnen. Dergelijke kennis is van direct belang voor leraren die pogingen in het werk stellen om meer probleemgeoriënteerd te gaan werken. Wanneer leraren erin slagen zich deze inzichten eigen te maken en toe te passen, dan is er sprake van een directe bijdrage van ‘inzicht in hoe iets werkt’ aan de kwaliteit van het onderwijs.

Op empirisch bewijs gebaseerde onderwijsverbetering hoeft zich niet te beperken

tot 'bewezen werkzaamheid', zoals de Onderwijsraad doet, je kunt ook evidentie verzamelen om inzichten te ontwikkelen en te onderbouwen. En, zo lieten wij met dit voorbeeld zien, inzichten kunnen op een directe manier tot onderwijsverbetering leiden. *Mutatis mutandis* geldt dit ook voor (innovatieve) concepten, oplossingen voor praktische problemen en het evalueren daarvan. Daarvoor heb je geen in handelingsrichtlijnen en protocollen vastgelegde onderwijsmethoden en -aanpakken nodig, waarvan de werking met behulp van RCT's is aangetoond. Sterker nog, het streven naar onderwijs gebaseerd op, met RCT's aangetoond, effectieve richtlijnen en protocollen kan verbetering juist in de weg staan. Niet alleen omdat wat je in 'aangetonbaar effectieve richtlijnen en protocollen' kunt vangen uiterst beperkt is, maar vooral omdat het verplicht toepassen van dergelijke richtlijnen en protocollen, de rol van de leraar reduceert tot die van uitvoerder en dat volgens ons de kwaliteit van het onderwijs niet ten goede komt.

Noten

- 1 Mol (2006) stelt: "...clinical trials are the gold standard of health care research, employed to prove that the care practices they study are good...we would do better to develop research methods that work toward another goal: to improve care practices. This requires that we no longer foreground the effectiveness but, instead, investigate the various effects of interventions (p. 405) ... however well clinical trials might be able to prove or disprove therapeutic claims, and however strong their credentials when it comes to seeking evidence, they have their limits when it comes to assuring good care" (p. 307).
- 2 Merk op dat Gould's interpretatie van de term *consilience* verschilt van die van Wilson (1998), die *consilience* verbindt met reductionisme.
- 3 De effectgrootte van de gerandomiseerde studies was 0,46; van de niet gerandomiseerde 0,41.

Literatuur

- Adams St. Pierre, E. (2002). 'Science' rejects postmodernism. *Educational Researcher*, 31(8), 25-27;.
- Atkinson, E. (2000). In defence of ideas, or why 'what works' is not enough. *British Journal of Sociology of Education*, 21, 317-330.
- Berenson, A. (2006, 11 januari). Drug development fell in slump in 2005. *International Herald Tribune*. Geraadpleegd op 26 september 2006, op <http://www.iht.com/articles/2006/01/11/business/research.php>
- Berliner, D. C. (2002). Educational research: The hardest Science of all. *Educational Researcher*, 31, 18-20.
- Chatterji, M. (2005). Evidence on "what works": An argument for extended-term mixed method (ETMM) evaluation designs. *Educational Researcher*, 34(5), 14-24.
- Cobb, P. (2004, juni). *Content matters: Towards a symbiosis of general and domain-specific theories of learning and instruction*. Keynote lezing op de jaarlijkse Onderwijs Research Dagen, Utrecht.
- Cobb, P., & Yackel, E. (1996). Constructivist, emergent, and sociocultural perspectives in the context of developmental research. *Educational Psychologist*, 31, 175-190.
- Davies, P. (1999). What is evidence-based education? *British Journal of Educational Studies* 47, 108-121.
- Elliott, J. (2001). Making evidence-based practice educational. *British Educational Research Journal*, 27, 555-574.
- Erickson, F., & Gutierrez, K. (2002). Culture, rigor, and science in educational research. *Educational Researcher*, 31(8), 21-24.
- Gould, S. J. (2004). *The hedgehog, the fox, and the magister's pox*. London: Vintage.
- Greenhalgh, T., Toon, P., Russell, J., Wong, G., Plumb, L., & Macfarlane, F. (2003). Transferability of principles of evidence based medicine to improve educational quality: systematic review and case study of an online course in primary health care. *British Medical Journal*, 326, 142-145.
- Haaft, G. (2006, 20 januari). Onderwijsraad: het nieuwe leren kan nadelig zijn. *Trouw*. Geraadpleegd op 2 oktober 2006, op <http://www.trouw.nl/deverdieping/dossiers/article134145.ece/>.

- Lipsey, M. W., & Wilson, D. B. (1993). The efficacy of psychological, educational and behavioral treatment: Confirmation from meta-analysis. *American Psychologist*, 48, 1181-1209.
- McClain, K., & Cobb, P. (2001). The development of sociomathematical norms in one first-grade classroom. *Journal for Research in Mathematics Education*, 32, 234-266.
- Maxwell, J. A. (2004). Causal explanation, qualitative research, and scientific inquiry in education. *Educational Researcher*, 33(3), 3-11.
- Mol, A. (2006). Proving or improving: on health care research as a form of self-reflection. *Qualitative Health Research*, 16, 405-414.
- NCTM Research Advisory Committee. (1996). Justification and reform. *Journal for Research in Mathematics Education* 27, 516-520.
- Network Science. (s.d.). Return on investments. In *The process of drug development* (hfdst. 9). Geraadpleegd op 26 september 2006, op: http://www.netsci.org/scgi-bin/Courseware/projector.pl?Course_num=course1&Filename=slide09.html.
- Norman, G. (2003). The effectiveness and the effects of effect sizes. *Advances in Health Sciences Education*, 8, 183-187.
- Oliver, M., & Conole, G. (2003). Evidence-based practice and e-learning in higher education: Can we and should we? *Research Papers in Education*, 18(4), 385-397.
- Olsen, D. (2004). The triumph of hope over experience in the search for "what works": A response to Slavin. *Educational Researcher*, 33(1), 24-26.
- Onderwijsraad. (2006). *Naar meer evidence based onderwijs*. Den Haag, Nederland: Uitgave Onderwijsraad.
- Pirrie, A. (2001). Evidence-Based practice in education: The best medicine? *British Journal of Educational Studies*, 49, 124-136.
- Roschelle, J., Tatar, D., Shechtman, N., Hegedus, S., Hopkins, B., Knudsen, J., & Stroter, A. (2007). *Scaling up SimCalc project: Can a technology enhanced curriculum improve student learning of important mathematics?* (Technical Report 01). Menlo Park, CA: SRI International. Geraadpleegd 20 juli 2007 op: http://math.sri.com/research/SimCalc_TechReport01.pdf.
- Rosen, M. (2003). The complexities, cost of mid-west biotech drug development. *Wisconsin Technology Network*. Geraadpleegd op 2 oktober 2006, op: <http://wistechology.com/article.php?id=377>.
- Simons, H. (2003). Evidence-based practice: Panacea or over promise? *Research Papers in Education* 18, 303-311.
- Slavin, R. E. (1987). Ability grouping: A best-evidence synthesis. *Review of Educational Research*, 57, 293-336.
- Whewell, W. (1840): *The philosophy of the inductive sciences* (1967 edition). London: Frank Cass & Co. Ltd.
- Wieringa, D., & Eaton, C. R. (2005). *Pharmaceutical Manufacturers Association*. Geraadpleegd op 2 oktober 2006, op: http://www.allp.com/drug_dev.htm.
- Wilson, E. O. (1998) *Consilience: The unity of knowledge*. New York: Knopf.

Manuscript aanvaard: 9 september 2007

Auteurs

Prof. dr. Koeno Gravemeijer is als hoogleraar Domeinspecifieke Onderwijstheorieën Rekenen-wiskunde werkzaam aan het Freudenthal Instituut en het Langeveld Instituut van de Universiteit Utrecht.

Prof. dr. Paul Kirschner is als hoogleraar Onderwijswetenschappen werkzaam aan het Langeveld Instituut van de Universiteit Utrecht.

Correspondentieadres: Prof. dr. Koeno Gravemeijer, Freudenthal Instituut, Universiteit Utrecht, Postbus 9432, 3506 GK Utrecht, Nederland.
E-mail: koeno.gravemeijer@esoe.nl

Abstract

Towards more evidence based education?

In its recent report "Naar meer evidence based onderwijs" [Towards more evidence-based education] the Education Council (Onderwijsraad, 2006) stresses the value of randomised controlled trials (RTC's) as these are being used in Health Care and Agriculture, as the way to ensure more evidence based educational innovations. In this manner, they reason, our knowledge about effect of an intervention on teaching and learning would be based in the strongest possible evidence. We believe, however, that there are a number of problems in applying this kind of research in the context of innovation in education. In this article, we argue that this kind of evidence-based education is not feasible, is too expensive, is too generalizing, shows what works but not how 'it' works and is conserving and retarding of innovation and change.