

False positive, false negative

Komt het ooit nog goed met het onderwijsonderzoek? Dat heeft door de jaren heen nooit veel potten kunnen breken; niet alleen in Nederland, maar wereldwijd. Docenten laten de uitkomsten veelal hoofdschuddend aan zich voorbijgaan, want van verrassend nieuwe inzichten is zelden sprake. Zij weten zich gesteund door gereputeerde instanties als het Center for Educational Research and Innovation (CERI) van de [OECD](#) dat het onderwijsonderzoek ooit onomwonden kwalificeerde als “irrelevant”, “navelstarend”, “slecht van kwaliteit” en “matig bruikbaar” is. Vernietigender kan het niet. Vrij vertaald: onderzoekers houden zich op de verkeerde manier met de verkeerde dingen bezig. Onderzoek heeft nauwelijks betekenis voor de onderwijspraktijk. Dat is niet alleen slecht voor het onderzoek, maar ook voor het onderwijs, dat zich bij gebrek aan beter slechts kan verlaten op praktische of subjectieve overwegingen.

In 2006 werd het Nederlandse onderwijsland opgeschrikt door het rapport “[Naar meer evidence-based onderwijs](#)” van de Onderwijsraad. De effectiviteit van onderwijsmethoden moet nadrukkelijker worden bewezen met behulp van wetenschappelijk onderzoek, vond de Raad. Zij houdt een stevig pleidooi voor een betere positie van het onderwijsonderzoek. Daar kon natuurlijk niemand tegen zijn, maar het voelde toch een beetje als het belonen van wanprestatie. De bevindingen van het CERI zijn weliswaar onverbloemd in het rapport opgenomen, maar van een bestraffende toon jegens onderzoekers is geen sprake. Onderzoek moet dus meer aandacht krijgen, en dat krijgt het ook. Wie in de recente tender “Onderwijsbewijs” de kleine lettertjes doorneemt, kan constateren dat het vooral moet gaan om *onderzoek*, ...*onderzoek* en ...*onderzoek*, zodat ondubbelzinnig kan worden aangetoond wat werkt, waarom het werkt en hoe.

U voelt het misschien al aankomen: met dit onvoorwaardelijk eerbetoon aan onderwijsonderzoekers is natuurlijk van alles mis. Ik zal dat op drie punten toelichten:

1) De wetenschappelijke methode

De Onderwijsraad spreekt een duidelijke voorkeur uit voor gerandomiseerde gecontroleerde (veld)experimenten, dat wil zeggen een a-selecte opsplitsing in twee groepen onder verschillende condities, waarna de statistiek mag vaststellen of waargenomen verbanden significant zijn. Dat is problematisch, omdat in veel meetsituaties het aantal variabelen erg groot is en vele daarvan buiten de controle van de onderzoekers vallen. Men kan effecten meten, men kan zelfs correlaties vaststellen, maar door het grote aantal ongecontroleerde en oncontroleerbare variabelen zijn er nauwelijks oorzakelijke verbanden te leggen. Shaver (1983) zegt hierover: “Statistische significantie is geen zinvol criterium om te bepalen of onderwijsmethoden werken.” Met andere woorden, effectmetingen onder zogenaamd gecontroleerde omstandigheden kunnen vaak helemaal niet worden toegeschreven aan de interventies.

2) Bevestigend onderzoek

Een ander groot probleem van onderwijsonderzoek is het bevestigende karakter ervan: men probeert eigenlijk common sense wetenschappelijk te bewijzen. Nu wil ik niet beweren dat bevestigend onderzoek helemaal achterwege moet blijven (want evengoed zou common sense op dwalingen, vooroordelen of bijgeloof kunnen berusten), maar over het algemeen schieten we er niet veel mee op. Daar bestaan legio voorbeelden van. Om niet meteen de hele [VOR](#)-familie over me heen te krijgen geef

ik een voorbeeld uit het buitenland. Loken en consorten (2004) bijvoorbeeld voerden een omvangrijke en kostbare studie uit naar het online studiegedrag van 100.000 studenten die zich voorbereidden op het toelatingsexamen voor een van de Amerikaanse colleges. Wat bleek? Studenten roeren maandenlang hun spullen niet aan en beginnen pas een paar dagen voor het examen te werken. Dat wisten we natuurlijk allemaal al, al blijft het de vraag of we het nu zeker weten.

### 3) Positief nieuws

Wil je als onderzoeker serieus worden genomen, dan moet je je onderzoek publiceren in wetenschappelijke tijdschriften. Maar dat kan eigenlijk alleen als je positieve resultaten hebt te melden, lees: significante resultaten. Daarom is het handig om veilige, algemene hypothesen te formuleren, waarvan je bij voorbaat al weet dat die zullen worden bevestigd, bijvoorbeeld: “Studenten beginnen pas een paar dagen voor het examen te werken”. Door de positieve selectiedwang van wetenschappelijke tijdschriften ontstaat er bovendien een vertekend beeld van alle statistische onderzoeksgegevens. Een groot deel wordt immers nooit gepubliceerd, ongeacht hoeveel geld en inspanningen er mee gemoeid waren. In diverse vakgebieden zijn nu tijdschriften verschenen waarin niet-significante en andere negatieve resultaten een positieve plaats kunnen krijgen, zoals het [Journal of Spurious Correlations](#), het [Journal of Negative Results in Biomedicine](#), het [Journal of Negative Observations in Genetic Oncology](#) met het briljante acroniem “NOGO”, het [Journal of Negative Results - Ecology & Evolutionary Biology](#) en het [Journal of Interesting Negative Results in Natural Language Processing and Machine Learning](#). Misschien ook iets voor ons: het *Journal for Negative Outcomes of Neat and Sensible, but Ever-Rejected or Neglected Studies in Education*. Of blijft alles dan toch zoals het was?



En nu maar afwachten welke wetenschappelijke bewijzen de projecten van het Onderwijsbewijs gaan opleveren.

Loken, E., Radlinski, F., Crespi, V.H., Millet, J. & Cushing, L. (2004). Online Study Behavior of 100,000 Students Preparing for the Sat, Act, and Gre. *Journal of Educational Computing Research*, 30-3, 255-262.

Shaver, J.P. (1983). The Verification of Independent Variables in Teaching. *Educational Researcher*, 10, 3-9.