

9

BELEIDSIMPLICATIES BIJ TEGENSTRIJDIGE ONDERZOEKSUITKOMSTEN

*Illustraties van valkuilen bij het gebruik van onderzoek
naar etnische diversiteit in het onderwijs*

Gert-Jan Veerman¹

Wat moet je doen als je beleid wilt maken op basis van wetenschappelijk onderzoek waarvan de resultaten totaal niet eenduidig zijn? Dit essay behandelt vier valkuilen bij het gebruik van tegenstrijdige onderzoeksuitkomsten door beleidsmakers: (1) verwarren van theoretische en empirische argumenten, (2) onderzoekscontext en beleidscontext aan elkaar gelijkstellen, (3) te sterke claims aan onderzoeksresultaten verbinden en (4) achterliggende trends negeren. De valkuilen worden geïllustreerd met eigen ervaringen tijdens onderzoek naar etnische diversiteit in het onderwijs. Het essay sluit af met voorbeeldvragen die

¹ De auteur dankt Jan Kaldeway MSc, MMus, Clemens van de Berg MSc, Maarten Dijkstra MA en prof. dr. Eddie Denessen voor hun waardevolle suggesties bij het tot stand komen van dit essay.

kunnen helpen bij het vertalen van tegenstrijdige onderzoeksuitkomsten naar passend beleid.

1 Inleiding

Wat moet je als politicus of beleidsmaker als onderzoeksresultaten niet eenduidig zijn? Is dan je keuze voor een beleidslijn, gebaseerd op een selectie van studies, niet te willekeurig? En is daardoor het risico op ongewenste beleidseffecten niet te groot? In dit essay behandel ik vier valkuilen rond tegenstrijdige onderzoeksuitkomsten, waarbij ik schets welke vragen je daarbij kunt stellen. Ik illustreer ze met eigen ervaringen op het gebied van onderzoek naar onderwijs en etnische diversiteit.

Onderzoek naar etnische diversiteit vertoont in het laatste decennium tegenstrijdige uitkomsten en is daarom bij uitstek geschikt als illustratie bij de vragen die in dit essay aan de orde komen. Zo blijkt uit het ene onderzoek dat leerlingen met een migratieachtergrond in diverse scholen lagere onderwijsprestaties behalen (Dronkers, 2010), terwijl dat uit ander onderzoek helemaal niet blijkt (Herweijer & Van den Brink, 2011). Op welk van die twee studies moet je je nu als politicus of beleidsmaker baseren?

Na herlezing van onderzoeksartikelen, rapporten en publieke debatten rond etnische diversiteit in het onderwijs uit het laatste decennium, zie ik vier valkuilen waarin je verstrikt kunt raken, wanneer je op tegenstrijdige onderzoeksuitkomsten stuit:

- het verwarren van theoretische en empirische argumenten;
- het vertalen van onderzoeksresultaten naar beleid, terwijl de context van dat onderzoek helemaal niet een-op-een met de beleidscontext te vergelijken valt;
- het verbinden van een sterkere claim aan de onderzoeksresultaten dan de onderzoekers zelf doen;
- het uit het oog verliezen van trends achter een fenomeen, die op een ander beleidsniveau liggen.

Al deze valkuilen wil ik nu één voor één langsgaan.

2 Het verwarren van theoretische en empirische argumenten

Beleidsmakers hebben de beschikking over onderzoeken die meer of minder theorie bevatten. Bij meer op theorievorming gericht onderzoek legt het onderzoek vaak een langere weg af. Zo worden de resultaten van onderzoek naar etnische diversiteit in het onderwijs eerst binnen een vakgroep van collega-onderzoekers en op congressen gepresenteerd. Vervolgens komen de uitkomsten in een artikel te staan. De conceptversie krijgt een beoordeling door onderzoekers van een ander instituut. Wanneer het stuk aan de kwaliteitseisen voldoet, belandt een – vaak aangepaste – versie in een onderzoekstijdschrift. Beleidsmakers kunnen zich daarop abonneren óf de resultaten via expertbijeenkomsten en de media tot zich nemen.

In deze eerste stroom van onderzoek wordt veelal een verbinding met achterliggende theorieën gelegd. Bij onderzoek naar diversiteit in het onderwijs gebeurt dat bijvoorbeeld door vooraf verwachtingen te formuleren over de betekenis van diversiteit in het onderwijs. Na het toetsen van de theorie aan de onderzoeksresultaten volgt een beschrijving van de beleidsimplicaties, die in artikelen of op congressen ter sprake komt.

Een tweede stroom van onderzoeksresultaten belandt in rapporten van onderzoeksinstanties die niet gecontroleerd worden door onderzoekers van andere instituten. Binnen die rapporten kan wel naar theorieën worden verwezen, maar vaak worden de onderzoeksresultaten direct gepresenteerd, waarna de onderzoekers – op basis van deze amper gecontroleerde en nauwelijks aan de theorie gekoppelde resultaten – beleidsimplicaties formuleren.

Om de betekenis van het onderscheid tussen theoretische en empirische argumenten te schetsen, gebruik ik een voorbeeld van het gebruik van theorie uit het onderzoek naar diversiteit in het

onderwijs. Daarbinnen kijken onderzoekers vanuit twee verschillende theoretische perspectieven naar diversiteit: het eerste is uitgewerkt in de *conflicttheorie* (Blalock, 1957) en het tweede in de *contacttheorie* (Allport, 1954). Het eerste perspectief legt de nadruk op conflicten die bij diversiteit ontstaan vanwege het feit dat diversiteit een bedreiging voor de oorspronkelijke meerderheid zou zijn. Het tweede perspectief legt de nadruk op verrijking door diversiteit, omdat die een grotere kans op contact met andersdenkenden zou geven. De contacttheorie stelt daarbij wel voorwaarden aan de context waarbinnen die contacten voorkomen. Is er bijvoorbeeld geen sprake van gelijkwaardigheid, dan zou de positieve uitwerking van diversiteit juist uitblijven. Studies die laten zien dat onderwijsprestaties van leerlingen in een diverse klas lager zijn, bieden op het eerste gezicht een empirische onderbouwing van de conflicttheorie. Toch is daar niet alles mee gezegd. Ze hoeven de contacttheorie namelijk niet tegen te spreken, aangezien het de vraag is of de studie plaatsvond in een context waarbinnen gelijkwaardigheid tussen leerlingen bestond. Hierop ga ik bij de behandeling van de vierde valkuil verder in.

Voor de beleidsmatige context is de kracht van een theoretisch argument afhankelijk van empirisch onderzoek dat de theorie bevestigt. Het bovenstaande voorbeeld illustreert dat het als beleidsmaker moeilijk is om te kunnen wegen hoe sterk een theoretisch argument is. Bij een verwijzing naar onderzoek ter onderbouwing van beleid is het daarom van belang om helder te hebben of de onderzoekers hypothetische uitkomsten formuleren op basis van een theorie of dat zij conclusies trekken op basis van empirische waarnemingen.

3 Onderzoekscontext verschilt van beleidscontext

Bij tegenstrijdige onderzoeksuitkomsten kunnen de contexten waarbinnen de diverse onderzoeken hebben plaatsgevonden, een belangrijke rol spelen. Ook dat blijkt bij studies naar diversiteit in het onderwijs.

Zijn leerlingen die naar een etnisch diverse school gaan in het voordeel of ondervinden ze er juist nadeel van? Empirisch onderzoek naar dit vraagstuk geeft geen eenduidig antwoord. Het was mijn leermeester, Jaap Dronkers, die een steen in de vijver gooide door cijfers te publiceren die aantoonde dat het zitten op een diverse school nadelig was voor de onderwijsprestaties van leerlingen (Dronkers, 2010). Deze gegevens werden in de media meteen opgepakt. De beleidsimplicaties leken immens. Moest in het Nederlandse onderwijsbeleid diversiteit niet gestimuleerd worden? Hadden leerlingen en docenten in diverse klassen meer ondersteuning nodig? Al snel werd Dronkers tijdens een televisiedebat gevraagd of de resultaten van zijn studie ook echt voor Nederland golden. De cijfers waren namelijk gebaseerd op studies in meerdere Europese landen, maar in Nederland was geen onderzoek gedaan. Er zouden dus geen beleidsimplicaties voor Nederland zijn. Maar waarom zouden de uitkomsten voor de Nederlandse context dan anders zijn? vroeg Dronkers zich af.

Uiteindelijk voerde het Sociaal en Cultureel Planbureau (Herweijer & Van den Brink, 2011) studies uit naar diversiteit in het Nederlandse onderwijssysteem. In een onderzoeksrapport stelde de organisatie dat etnische diversiteit geen invloed heeft op de onderwijsprestaties in het voortgezet onderwijs. Alleen leerlingen in etnisch diverse groepen 8 op de basisschool zouden wat lager op rekentoetsen scoren. Een van de auteurs, Lex Herweijer, concludeerde in een opiniebijdrage in het dagblad *Trouw* (14 augustus 2011) dat beleid dat zich richt op het mengen van scholen niet contraproductief is.

Enige jaren later stelde ik in een onderzoeksartikel, samen met Jaap Dronkers en Herman van der Werfhorst (2013), dat leerlingen met een migratieachtergrond in een meer diverse groep 8 lager scoorden op een toets begrijpend lezen. Ook hier bleek het belang van kijken naar de context. Voor andere leerjaren en andere vakken vonden we geen verschillen. Als er dus al beleid zou moeten worden gemaakt op basis van deze gegevens, dan zou dit zich *alleen* op groep 8 moeten richten en op het vak begrijpend lezen. In andere groepen en bij

andere vakken vonden wij geen significant verschil tussen de resultaten van meer en minder diverse klassen. Op basis van dit onderzoek zouden beleidsmakers dus beleid moeten maken voor een specifieke groep (de leerlingen met een migratieachtergrond), in een specifieke leeftijdsfase (groep 8), voor een specifiek vak (begrijpend lezen).

Dit voorbeeld laat zien hoe belangrijk het is om onderzoeksresultaten goed te analyseren en te interpreteren: ga na wat de context van de verschillende onderzoeken is. Richt je vervolgens op onderzoeksgegevens die gebaseerd zijn op een context die aansluit bij je beleidscontext.

4 Te sterke claims aan onderzoeksresultaten verbinden

Het is van belang om de analysemethode binnen een onderzoek bij de beleidsimplicaties te betrekken. Uiteraard is het als beleidsmaker moeilijk om af te wegen welke analysetechniek wenselijk is. Onderzoeksartikelen bevatten veelal disclaimers die aanwijzingen geven over de beperkingen van de analysemethode en de betekenis daarvan voor mogelijke beleidsimplicaties, maar wat moet je daarmee? Ook nu wil ik weer een illustratie uit onderzoeken naar diversiteit in het onderwijs geven.

In studies naar diversiteit en onderwijsprestaties proberen onderzoekers uiteraard om vergelijkbare onderzoeksgroepen samen te stellen. Zo houden ze rekening met zaken als het opleidingsniveau van de ouders. Toch valt nooit helemaal uit te sluiten dat verschillen tussen klassen en scholen te verklaren zijn vanuit negatieve selectie. Zo zou het kunnen dat leerlingen die sowieso al minder kans op hoge cijfers voor begrijpend lezen hebben, eerder in etnisch diverse klassen belanden. Vanwege de gebruikte onderzoeksmethode kan dan niet met zekerheid worden gesteld dat de diversiteit *de* oorzaak van de lagere onderwijsprestaties op het gebied begrijpend lezen is. En omdat het geen experimenteel onderzoek naar beleidswijzigingen betreft, kun je al helemaal niet concluderen dat een beleid dat diversiteit in de

klas bevordert, tot betere onderwijsprestaties leidt. Dit is dan ook de grootste disclaimer bij dit soort diversiteitsonderzoek.

Het verschil in scores tussen meer en minder etnisch diverse klassen wordt dan niet veroorzaakt door de mate van diversiteit, maar door de specifieke leerlingen die voor die scholen kiezen. In de conclusie van mijn artikel wijs ik beleidsmakers er dan ook op dat ze gezien de analysemethode voorzichtig moeten zijn met het vertalen van uitkomsten naar concreet beleid.

5 Achterliggende trends negeren

Wie stelt dat je alleen beleid mag ontwikkelen als je je baseert op studies die plaatsvonden binnen een context die hetzelfde is als de beleidscontext en die gaan over een specifieke beleidswijziging, beperkt beleidsmakers behoorlijk in hun vrijheid. Beleid kan in dat geval pas ontwikkeld worden als er een studie voorhanden is die precies over die context gaat en over de specifieke beleidswijziging. De kans om exact zo'n studie te vinden is klein.

Daar komt bij dat omstandigheden in de loop van de tijd veranderen. Ook dit maakt de vertaling van onderzoeksresultaten naar beleid complex. Als beleidsmaker hoef je je niet alleen te richten op onderzoeksgegevens die gebaseerd zijn op een vergelijkbare context, maar kun je ook kijken naar achterliggende en contextoverstijgende trends. Verschillen tussen contexten hoeven immers niet willekeurig te zijn ontstaan. Achterliggende factoren zoals landelijk beleid kunnen het effect van regionaal beleid sturen. Met andere woorden: achterliggende trends op het ene niveau kunnen aanwijzingen geven voor beleid op een ander niveau.

Wie achterliggende en contextoverstijgende trends wil ontdekken, kan gebruikmaken van overzichtsstudies of landenvergelijkend onderzoek. Bij dergelijk onderzoeken is het van belang om niet alleen beschikbare onderzoeksresultaten van losse studies of afzonderlijke

landen op een rij te zetten. In het geval van onderzoek naar etnische diversiteit binnen het onderwijs wordt dus geen landenlijst opgesteld waarop staat dat leerlingen in diverse scholen wel of geen voordeel van diversiteit hebben als je kijkt naar onderwijsprestaties en sociale uitkomsten, maar zoeken onderzoekers naar achterliggende trends die de verschillen verklaren. Deze trends worden ‘verklarende mechanismen’ genoemd. Dergelijk onderzoek geeft antwoord op de vraag waarom een bepaald fenomeen zich in de ene context wel en in de andere niet voordoet.

Een voorbeeld van een dergelijk diversiteitsonderzoek is een landenvergelijkingsstudie naar diversiteit en wanorde in de klas van mijn hand (Veerman, 2015). Uit een analyse van zestien landen bleek dat leerlingen in meer etnisch diverse scholen – gemiddeld genomen – meer wanorde ervaren. Echter... alleen als we naar alle landen tegelijk kijken. Als we bijvoorbeeld op leerlingen in België inzoomen, blijkt dat hun ervaring van wanorde niet samenhangt met de etnische diversiteit binnen de school. Bij leerlingen in Zwitserland is dit echter wel het geval. Mocht een beleidsmaker zich focussen op de context, dan zou een Zwitserse beleidsmaker dus heel andere keuzes maken dan een Belgische.

We waren echter niet op zoek naar een lijst van landen met verschillende effecten, maar naar een onderliggende trend. En die bleek er inderdaad te zijn. Met het oog op de eerder besproken contacttheorie zou je verwachten dat diversiteit pas tot voordelen leidt als er gelijkwaardigheid tussen alle onderzoeksdeelnemers bestaat. Gelijkwaardigheid tussen de verschillende groepen zou dan een verklarend mechanisme – en dus een achterliggende en overstijgende trend – zijn, waarmee de verschillen tussen landen verklaard kunnen worden. Dit op de contacttheorie gebaseerde mechanisme kon ik zichtbaar maken met behulp van een index, waarmee ik landen kon vergelijken op de mate van de inclusiviteit van het integratiebeleid. Landen die laag scoren laten migranten bijvoorbeeld relatief laat participeren op de arbeidsmarkt en kinderen van migranten pas laat

deelnemen aan het onderwijs. Landen met een inclusief integratiebeleid bieden daarmee ouders én kinderen een grotere kans om gelijkwaardig mee te doen.

Wat bleek? In landen met een inclusief integratiebeleid ervoeren leerlingen niet meer wanorde in meer etnische diverse scholen. In landen met een minder inclusief integratiebeleid ervoeren ze dat juist wel. Ook hier kan overigens selectie een rol spelen en moet er dus voorzichtig worden omgegaan met het formuleren van beleidsimplicaties.

Daar waar de losse landenstudies aanknopingspunten bieden rond het al dan niet beïnvloeden van de etnische compositie, geeft de landenvergelijkende studie handvatten om de mate van inclusiviteit van het integratiebeleid te reguleren. Hiermee wordt het diversiteitsvraagstuk in het onderwijs verplaatst van het beïnvloeden van diversiteit naar het beïnvloeden van hoe gewenste en ongewenste consequenties van diversiteit kunnen worden versterkt of geremd op een ander beleidsniveau. Verschillen in onderzoeksresultaten vallen dus soms te verklaren door beleid op een ander niveau.

6 Tot slot

Op basis van de hiervoor beschreven valkuilen, kom ik tot een overzicht van vragen die kunnen helpen bij het vertalen van tegenstrijdige onderzoeksuitkomsten naar beleid:

- Gaat het in het onderzoek waarnaar verwezen wordt om een theoretische verwachting of om een empirische bevinding?
- In welke specifieke contexten zijn de onderzoeken met tegenstelde uitkomsten uitgevoerd? In welke mate zijn de uitkomsten te vertalen naar de beoogde beleidscontext?
- Kan ik, ondanks de disclaimers wat betreft de onderzoeksmethodiek, de studie in de beoogde beleidscontext als onderbouwing voor mijn beleid gebruiken?

- Is er meer bekend over achterliggende trends die de verschillen in uitkomsten verklaren? Wat zijn de implicaties daarvan voor het beoogde beleid? Op welk beleidsniveau?

Literatuur

- Allport, G.W. (1954). *The Nature of Prejudice*. Cambridge: Perseus Books.
- Blalock, H.M. (1957). Per Cent Non-White and Discrimination in the South. *American Sociological Review*, 22, 677-682.
- Dronkers, J. (2010). *Positieve maar ook negatieve effecten van etnische diversiteit in scholen op onderwijsprestaties? Een empirische toets met internationale PISA-data* (inaugurele rede). Maastricht: Maastricht University.
- Herweijer, L. & Brink, Y. van den (2011). *Gemengd leren. Etnische diversiteit en leerprestaties*. Den Haag: Sociaal en Cultureel Planbureau.
- Veerman, G.-J.M. (2015). The relationship between ethnic diversity and classroom disruption in the context of migration policies. *Educational Studies*, 41 (1-2), 209-225.
- Veerman, G.-J.M., Werfhorst, H.G. van de & Dronkers, J. (2013). The effect of the ethnic composition on educational performance at primary education in the Netherlands. *Educational Research and Evaluation*, 19 (5), 370-401.