

ARTIKEL

Hoe afhankelijk is beleidsonderzoek van onafhankelijkheid?*

*Jos Mevissen & Frans Leeuw***

1 Onafhankelijkheid, een kernwaarde van beleidsonderzoek

Onafhankelijkheid wordt doorgaans als een van de kernwaarden van beleidsonderzoek gezien. Deze kernwaarde moet waarborgen dat objectieve, in de betekenis van ongekleurde, informatie ten behoeve van beleidsvorming, -uitvoering en -evaluatie wordt verzameld. Het is verbonden met het streven wetenschappelijk onderzoek objectief, zo niet waardenvrij te laten zijn. Kwaliteit van onderzoek wordt eveneens als kernwaarde gezien, waarbij de nadruk ligt op validiteit, betrouwbaarheid en relevantie van de verzamelde informatie. Of onderzoek in de praktijk beide kernwaarden realiseert en hoe de nadruk op onafhankelijkheid zich verhoudt tot het streven naar kwaliteit, zijn vragen die niet a priori beantwoord kunnen worden. Immers, omstandigheden in de praktijk van het onderzoek zoals hoe uitvoerders, opdrachtgevers en stakeholders met elkaar omgaan, beschikbaarheid van middelen en methoden en van data kunnen de realisatie van deze kernwaarden bemoeilijken of onmogelijk maken. Ook zijn er situaties waarin de ene kernwaarde wel (of meer) gerealiseerd wordt dan de andere.

Over deze thematiek gaat deze bijdrage. We richten onze aandacht op wat onafhankelijkheid van beleidsonderzoek inhoudt, hoe daar de afgelopen decennia in Nederland mee om is gegaan en welke relaties er zijn tussen onafhankelijkheid en kwaliteit van dergelijk onderzoek. Daarnaast besteden we – beperkt – aandacht aan instrumenten die beogen bij te dragen aan de onafhankelijkheid van onderzoek. In paragraaf 2 geven we een schets van debatten over onafhankelijkheid van (beleids)onderzoek in Nederland en wat we daaruit kunnen leren. In paragraaf 3 gaan we nader in op het begrip onafhankelijkheid in relatie tot kwaliteit van onderzoek. In paragraaf 4 bespreken we welke actoren met elkaar interacteren rond beleidsonderzoek en hoe dit doorwerkt op de onafhankelijkheid van het onderzoek. In paragraaf 5 komt de vraag aan de orde hoe onafhankelijkheid en kwaliteit zich tot elkaar verhouden, terwijl paragraaf 6 ingaat op de vraag hoe geprobeerd wordt de onafhankelijkheid van beleidsonderzoek te borgen en wat over de effecten van

* De auteurs danken Lyn Pleger (ZHAW School of Management and Law, Zurich University) en Jacob van der Wel (Regioplan Beleidsonderzoek te Amsterdam) voor hun bijdrage aan de totstandkoming van dit artikel. Dat geldt ook voor de reacties en suggesties van twee anonieme reviewers.

** Jos Mevissen is zelfstandig adviseur en ondersteuner voor het opzetten, uitvoeren en beoordelen van beleidsonderzoek en daarnaast voorzitter van de redactie van Beleidsonderzoek Online. Frans Leeuw is hoogleraar Recht, openbaar bestuur en sociaalwetenschappelijk onderzoek (emeritus), Maastricht University.

dergelijke maatregelen, óók op het realiseren van de andere kernwaarde (kwaliteit), te zeggen is.

Verantwoording

De aanleiding voor dit artikel was de publicatie van een Zwitsers onderzoek naar de invloed van opdrachtgevers op beleidsonderzoek en beleidsonderzoekers (Pleger & Hadorn, 2018). Het leek ons interessant het te repliceren in Nederland. We werden in dit idee gesterkt door het feit dat een vergelijkbaar onderzoek onder opdrachtgevers ook in de Verenigde Staten werd uitgevoerd. Daarom hebben we in 2020 een enquête¹ over de onafhankelijkheid van beleidsevaluaties, geënt op het Zwitserse onderzoek, uitgezet onder opdrachtgevers van beleidsonderzoek in Nederland. Hoofddoel was om de *Zwitserse 'theorie' van de Grote Boze Wolven* (opdrachtgevers die onderzoek beïnvloeden en dus de onafhankelijkheid ervan aantasten) te onderzoeken voor Nederland.

Het verloop van ons onderzoek en de bestudering van de data die het heeft opgeleverd, hebben ons tot de conclusie gebracht dat de uitkomsten onbruikbaar zijn voor de beantwoording van de vraag of (en waarom) opdrachtgevers beleidsonderzoek beïnvloeden en of daarmee de onafhankelijkheid ervan wordt aangetast. In die zin is het onderzoek mislukt. De belangrijkste redenen waarom wij spreken van een mislukt onderzoek, zijn de zeer lage respons (netto $n = 37$) in combinatie met onduidelijkheid over hoe groot de feitelijke doelgroep van het onderzoek is in combinatie met het ontbreken van een duidelijke afbakening van wat verstaan moet worden onder 'opdrachtgever'. Dé opdrachtgever bestaat niet: zij zijn er in soorten en maten. Het maakt bijvoorbeeld verschil of men één keer opdrachtgever voor een extern uitgevoerd onderzoek is geweest of dat men de functie heeft om dat werk min of meer permanent te doen. Ook de mate van ervaring als opdrachtgever maakt uit. Dat heeft niet alleen invloed op de omgang met onderzoekers, maar ook op het verloop van het onderzoeksproces en op de 'setting' van het onderzoek binnen de organisatie. Onze analyse van de verzamelde data leerde ons ook dat de complexiteit van de opdrachtgever-opdrachtnemer-relatie niet voldoende 'gevangen' kon worden in de gehanteerde vraagstelling en gebruikte antwoordcategorieën.² Hierdoor bleven gradaties in ontevredenheid, de periode waarop een respondent terugkijkt en of het om een eenmalige of een meer frequente ervaring gaat buiten beeld, terwijl dergelijke informatie van belang is voor de interpretatie van antwoorden op vervolgvragen over hoe er is gereageerd op ontevredenheid met het werk van externe onderzoekers.

Maar ook van mislukt onderzoek kan worden geleerd. Het was voor ons in ieder geval aanleiding om uiteenlopende aspecten van het concept onafhankelijkheid

1 Onze enquête (en het rappel) werd uitgezet m.m.v. de Nederlandse Vereniging van Rekenkamers & Rekenkamercommissies (NVRK), de Vereniging voor Beleidsonderzoek (VBO), het Evaluatorennetwerk van de VIDE, enkele kennisinstellingen van de overheid en een honderdtal contacten bij vakbonden, werkgeversorganisaties, A&O-fondsen, sectorale organisaties en uitvoeringsorganisaties waarvan bekend is dat men er opdrachtonderzoek laat uitvoeren of uitvoert (waar relevant met het verzoek om de enquête onder hun opdrachtgevers uit te zetten).

2 Indien lezers meer informatie over het uitgevoerde onderzoek willen hebben, kunnen zij contact opnemen met de auteurs.

met elkaar te bediscussiëren en ons vragen te stellen over de visies op onafhankelijkheid, de rol van onderzoekers én opdrachtgevers, het effect op kwaliteit van een onderzoek, et cetera. Dit artikel vormt de neerslag van onze zoektocht. Omdat we daarbij, in het verlengde van ons onderzoek, in beginsel de Nederlandse situatie voor ogen hadden, hebben we ons verhaal beperkt tot Nederland. Die inperking tot Nederland is in het beknopte historische overzicht vooral ingegeven door de overweging dat er in de literatuur in ruime mate wordt voorzien in studies naar situaties in andere (westerse) landen (waarbij overigens lang niet altijd decennialang wordt teruggekeken). Dat neemt natuurlijk niet weg dat inzichten uit de internationale vakliteratuur nuttig zijn gebleken in onze gedachtevorming.

2 Een kleine geschiedenis van discussies over onafhankelijkheid van beleidsonderzoek in Nederland

Discussies over de onafhankelijkheid van onderzoek vinden al langere tijd plaats en zijn ook vandaag de dag actueel. Denk aan discussies over de rol van het RIVM tijdens de coronacrisis (Sommer, 2020), de onafhankelijkheid van onderzoek door toezichthouders (Van Vollenhoven & Eijlander, 2022) en aan het Indonesië-onderzoek dat begin 2022 verscheen.

De waarschijnlijk eerste keer dat er in Nederland *systematische aandacht* was voor de relatie tussen (sociaalwetenschappelijk) onderzoek en beleid was in 1975, toen een rapport van de Verkenningcommissie Sociaal Onderzoek gepubliceerd werd. Vastgesteld werd dat de communicatie tussen beleid(smakers) en (beleids)onderzoek(ers) slecht functioneerde.³ In vervolg daarop verscheen het meerjarenplan Sociaal Onderzoek en Beleid, waarin werd aangekondigd dat men meer werk ging maken van deze relatie.

Onafhankelijkheid van onderzoek was in beide rapporten *niet* het centrale thema. Integendeel, (het streven naar) interactie en samenwerking was dat wel.⁴ En ook tien jaar na het rapport van de Verkenningcommissie werd in een vergelijkende studie naar de samenwerking tussen onderzoek en beleid niet of nauwelijks aandacht besteed aan onafhankelijkheid (Van Oijen et al., 1984). Toen Van de Vall (1980) zijn boek over het professionele paradigma van beleidsonderzoek publiceerde, dat als een van de centrale thema's had dat beleidsonderzoekers vooral ook adviseur (en soms zelfs agoog) moesten zijn, leidde dat wel tot discussies, die óók op de onafhankelijkheid van onderzoek betrekking had. Schuyt (1984) bracht namelijk in een reactie op Van de Vall (1980) naar voren dat 'de vergrote inspanning om "eindelijk" het sociaalwetenschappelijk onderzoek beleidsrelevant te maken, zal leiden tot een afbraak van de mogelijkheden tot zelfstandige beoefening van de sociale wetenschappen, tot het onmogelijk maken van concurrerende werkelijkheidsinterpretaties en tot *de destructie van een onafhankelijke wetenschapsbeoefe-*

3 NRC Handelsblad besteedde aandacht aan dit rapport: www.nrc.nl/nieuws/1975/09/05/rapport-sociaalwetenschappelijk-onderzoek-deugt-niet-kb_000031948-a3173709.

4 Zie Jolles (1978: 229-230), die in een commentaar op het rapport er op wees dat 'het meerjarenplan een *sturende, begeleidende interdepartementale overheid* kent, die zelf partner en partij is in het krachtenveld dat door de trefwoorden sociaal onderzoek en beleid is aangegeven'.

ning' (Schuyt, 1984: 27). Van de Vall (1980), Van Hoesel (1984) en Coenen (1984) reageerden kritisch ('een betoog vol misverstanden') en beklemtoonden onder andere het belang van interactie tussen onderzoekers, beleid en samenleving. Schuyt (1984) repliceerde dat hij geenszins had beoogd de positieve kanten van beleids-onderzoek te ontkennen, maar bleef tegelijkertijd het belang van onafhankelijk onderzoek benadrukken. Beeldsonderzoek zou immers altijd alleen dát deel van de werkelijkheid onderzoeken waar beleid ('verkokerd en gefragmenteerd tot stand gebracht') voor stond, waardoor het 'halve waarheden' zou produceren. 'Gegeven dit permanente kenmerk van beleid, lijkt mij een herwaardering voor onafhankelijk onderzoek en een relativering van beeldsonderzoek niet eens zo onzinnig' (Schuyt, 1984: 315).

Dat er rond 1975-1985 over onafhankelijkheid gediscussieerd werd, betekent niet dat er niet al veel eerder door landgenoten aandacht aan besteed werd. In de twintigste eeuw zijn er verschillende voorbeelden te noemen waarbij over het belang van 'onafhankelijk onderzoek' geschreven en gediscussieerd werd. Dat speelde onder andere bij het oprichten van het Rijksinstituut voor Oorlogsdocumentatie (RIOD), bij het onderzoek dat De Jong in opdracht van de regering naar de Tweede Wereldoorlog deed⁵ en bij de regeringsopdracht aan het RIOOD om onderzoek te doen naar de Menten-affaire.⁶

Het was ook al eerder, in de jaren twintig en dertig, het geval bij werk dat sociografen deden rondom de droogwinning van de Zuiderzee, annex de migratie daarheen. Van Amersfoort (2007: 138 ff) zette uiteen hoe toentertijd de drijfveer om de sociografie 'sociaal nut' te laten hebben, 'de gedachte onder druk zette van de wetenschap als een bezigheid van de "*freischwebende Intelligenz*". Hij sprak over 'vervlechting' van beleid en onderzoek en over de 'omhelzing door de politiek' van het (toegepaste) onderzoek (p. 142-143). 'De geschiedenis (van het migratieonderzoek) leert ons dat voor zinvol toegepast onderzoek niet alleen maatschappelijke betrokkenheid nodig is, maar ook distantie' (p. 145). Gastelaars (1986) beschreef hoe de Nederlandse sociologie/sociografie in die tijd met het planmatigheidsdenken omging, waarbij zij aan 'de droom van macht en maakbaarheid' refereerde. In de jaren vijftig schreef Ter Hoeven (1956) over de soms problematische relatie tussen de overheidsbureaucratie (opdrachtgever/aanvrager van onderzoek) en de sociaal (beleids)onderzoeker. Hij was van oordeel dat de onderzoeker, als deze in dienst was van een overheidsorganisatie, gedwongen werd te werken in het beleidskader en noemde dat 'zeer verraderlijk' (omdat de onderzoeker dan zijn werk ondergeschikt moest maken aan de doelstellingen die 'de directeur en zijn administratieve staf hebben') (Ter Hoeven, 1956: 189).

Daarnaast vonden er in de periode tussen 1945 en 1960 discussies plaats over de (on)afhankelijkheid van en binnen het ISONEVO (Instituut voor Sociaal Onder-

5 Beekers (2007: 96, 99ff) spreekt over 'Tussen Scylla en Charybdis' als hij een kleine eeuw historisch-wetenschappelijk onderzoek in opdracht van de overheid bespreekt en ingaat op de mate van onafhankelijkheid die L. de Jong had.

6 De zaak Menten was een Nederlandse politieke affaire uit 1976 naar aanleiding van de mislukte arrestatie van de Blaricumse miljonair en veroordeelde oorlogsmisdadiger Pieter Menten. Blom (1984) rapporteerde hierover, waarbij hij ook criteria benoemde waaraan regeringsopdrachten voor historisch onderzoek zijns inziens moesten voldoen, wilden zij 'onafhankelijk' kunnen zijn.

zoek van het Nederlandse Volk), dat in 1940 werd opgericht (en ruim dertig jaar later werd opgeheven). Winkels (1980) zette een aantal discussies en incidenten op een rijtje, zoals de affaire naar aanleiding van het onderzoek van Diels (1951) naar het bevolkingsvraagstuk en voortplanting, dat uit twee enquêtes bestond, waarbij in de eerste meting onder andere gevraagd werd naar wat de ideale gezinsgrootte voor de geïnterviewde vrouwen was en of zij 'hoopten zó spoedig mogelijk kinderen te krijgen' (Winkels, 1982: 111-112). Het rapport veroorzaakte 'opschudding' (in de Kamer en in de media), die er mede debet aan was dat financiële middelen om deel 2 uit te voeren, uitbleven.

In 1989 is het vermoedelijk de eerste keer dat onafhankelijkheid als zelfstandig 'probleemonderwerp' aan de orde komt in een *empirisch onderzoek*. Het betrof een onderzoek van Van Puffelen et al. (1989) over relaties tussen opdrachtgevers van beleidsonderzoek en onderzoekers. Zij constateerden dat er conflictmogelijkheden in het samenspel van beide bestaan, waarbij zij op elementen wezen die gezien kunnen worden als 'geen of beperkte onafhankelijkheid'. Genoemd werden onder andere het tussentijds openbaar maken van resultaten (wat de onderzoeker soms wil maar de opdrachtgever niet (altijd)) en de 'eigenwijsheid' van de onderzoeker ('... regelmatig leidend tot conflicten', p. 2). Een van de oplossingen voor deze problematiek werd gevonden in 'een soepele opstelling van de opdrachtgever naar de onderzoekers toe'.

In 1999 verschijnt van de hand van Köbben en Tromp 'De onwelkome boodschap' met daarin verschillende casus van beïnvloeding van onderzoek door opdrachtgevers. De focus is een wetenschappelijke, in die zin dat het (alleen) ging om onderzoek dat in opdracht is uitgevoerd door universitaire medewerkers. En passant werd gesuggereerd dat commerciële bureaus 'zeer coöperatief meedenken met hun opdrachtgever' en beïnvloedbaar zijn (p. 23), terwijl zij geen onderzoek hadden gedaan naar dergelijke bureaus.

De periode na de millenniumwisseling lijkt te wijzen op een trendbreuk. Vooral na 2010 zijn er tal van voorbeelden waarbij er aandacht is voor de onafhankelijkheid (en kwaliteit) van sociaalwetenschappelijk onderzoek, waaronder beleidsonderzoek. Zo was er in 2016 ophef ten aanzien van de onafhankelijkheid van statistisch onderzoek door het CBS toen de Centrale Commissie voor de Statistiek werd afgeschaft. Onderzoekers van een van de Planbureaus (SCP) repten in 2018 in *de Volkskrant* over vergaande bemoeienis van ministeries. Het Planbureau voor de Leefomgeving kreeg in 2019 van de Nederlandse Melkveehoudersvakbond de Zure Melk Trofee toegekend vanwege volgens zijn suggestieve cijfers over oppervlaktewatervervuiling, kwantificering van de mestfraude en uitwerking van de effecten van het Klimaatakkoord.

Een piek aan aandacht in media en politiek leek de WODC-affaire op te leveren, begonnen rond de jaarwisseling 2017-2018. Het ging vooral over de onafhankelijkheid van het instituut in relatie tot het beleid van het ministerie van Justitie en Veiligheid en de mate waarin het ministerie invloed (probeerde) uit te oefenen op (drie) rapporten over drugs (beleid).⁷ Een primetime tv-uitzending en berichten en

7 Het betrof de rapporten 'Het Besloten club- en het Ingezetenen criterium voor coffeeshops', 'Coffeeshops, toeristen en lokale markt' en 'Internationaal recht en cannabis'.

vragen uit de Tweede Kamer leidde tot drie commissies,⁸ ingesteld door de minister van Justitie en Veiligheid om de vermeende beïnvloeding te onderzoeken. Het rapport van de Commissie Overgaww onderzocht de onafhankelijkheid, betrouwbaarheid en kwaliteit van de drie ter discussie gestelde WODC-rapporten over drugsbeleid en concludeerde dat 'de drie rapporten en de totstandkoming ervan getuigen van eerlijk, zorgvuldig en betrouwbaar onderzoek. De commissie concludeert dat de rapporten inhoudelijk kwalitatief aan de maat zijn. De rapporten zijn evenwichtig en kunnen, als beleidsonderzoeken, de wetenschappelijke toets doorstaan' (p. 117-118). Een jaar later verscheen het rapport van de Commissie Hertogh, die pretendeerde breder te kijken, maar dat deed op een manier die fikse (methodische) kritiek opleverde van onder andere De Ruig (2019) en Derksen (2019).⁹ Dit alles ging niet alleen gepaard met Kamervragen en debatten, persberichten en sociale media-uitingen maar ook met meer aandacht voor procedures, codes, regelingen, escalatieladders en integriteitscommissies bij verschillende rijkskennisinstellingen, waaronder het WODC.

Het RIVM had ook te maken met kritiek op zijn onafhankelijkheid, bijvoorbeeld betreffende onderzoeksgegevens rond de uitstoot van Tata Steel in IJmuiden, waar burgers in de media klaagden over gebrek aan vertrouwen in de data. En ook ten aanzien van de rol van het instituut in de coronacrisis (2020-2022) werd gewag gemaakt van vermenging van functies en rollen tussen 'wetenschap', het Outbreak Management Team en het ministerie van VWS.¹⁰ Ook werd en wordt over de onafhankelijkheid van toezichtsorganen zoals inspecties gediscussieerd. Verschillende jaren hield mr. Pieter van Vollenhoven, ook in zijn rol als voorzitter van de Onderzoeksraad voor Veiligheid, pleidooien voor het onafhankelijk maken van inspecties en andere toezichtsorganen. Want dat zouden ze volgens hem niet zijn. De discussie daarover gaat tot op heden door, inclusief een kritisch rapport van de Auditdienst Rijk (2021) over hoe rapporten binnen de Inspectie Justitie en Veiligheid tot stand kwamen.

De intensivering van de discussie over onafhankelijkheid ging gepaard met een grotere aandacht voor organisatorische en regelgevingsvoorzieningen ('borgingsinstrumenten') om onafhankelijkheid van onderzoek(ers) beter te proberen te waarborgen. In lijn met Van der Knaap et al. (2019) kunnen de meeste van deze voorzieningen tot de categorie 'harde waarborgen' worden gerekend. Daarnaast

8 Twee inhoudelijke commissies, en een derde over de wijze waarop het ministerie en het WODC met een klacht van een klokkenluider waren omgegaan.

9 De Ruig (2019) fileert 'de zwakke plekken (beoordelingskader, methodologie en synthese)' in dit rapport; hij vergeleek het rapport van WODC-commissie Overgaww (dat 'securer te werk is gegaan en haar conclusies en aanbevelingen logisch laat voortvloeien uit de goed onderbouwde bevindingen') met het rapport van WODC-Commissie Hertogh, dat 'analytisch en methodologisch zwakker van onderbouwing' is. Zijn eindconclusie is dat 'het rapport van de commissie Hertogh duidelijk maakt dat een onderzoekscommissie met academische achtergrond geen garanties geeft op een grondige analyse en een valide en betrouwbaar onderzoek. Het rapport voldoet niet aan de kwaliteitseisen die je aan beleidsonderzoek kunt stellen. Dat is spijtig: de WODC-affaire had beter onderzoek verdiend. Met gedegen onderzoek had men de kwestie weer in de juiste proporties kunnen zien' (p. 10).

10 Het eerste OVV-rapport over de aanpak van de coronaproblematiek (OVV, 2022: 17) formuleert op dit punt een aanbeveling.

zijn er ‘zachte maatregelen’, het belang van relaties en responsiviteit. Naast de formele positie, voldoende financiële middelen en vrije inhoudelijke keuze bieden het goed onderhouden van relaties en het organiseren van voldoende responsiviteit ook een aangrijpingspunt om te sturen op onafhankelijkheid in onderzoek. Op (de werking van dergelijke) borgingsinstrumenten komen we in paragraaf 6 terug.

Wat leren we uit deze mini-geschiedenis?

- *Ten eerste* dat het onderwerp zeker niet alleen de afgelopen jaren speelde.
- *Ten tweede* dat waar het in de jaren tachtig vooral ging over het *bevorderen van samenwerking* en interactie tussen beleid en onderzoek en er minder aandacht was voor het thema onafhankelijkheid, dat anno nu eerder omgekeerd lijkt te zijn.
- *Ten derde* dat lang niet altijd via (gedegen) empirisch onderzoek wordt vastgesteld of en in hoeverre de onafhankelijkheid van onderzoek beïnvloed is, met welke beweegredenen en gevolgen, in welke richting en wat de rollen van onderzoekers, opdrachtgevers en anderen daarbij waren.
- *Ten vierde* dat de nadruk in discussies over onafhankelijkheid ligt bij de rol van de onderzoeker en veel minder of niet bij de rol van de opdrachtgever of bij anderen.

In de vierde paragraaf gaan we dieper in op de rol van de opdrachtgever; eerst bespreken we wat ‘onafhankelijkheid van beleidsonderzoek’ eigenlijk is.

3 Onafhankelijkheid en kwaliteit van beleidsonderzoek

Onafhankelijkheid

Onafhankelijkheid is een van de kernwaarden van (beleids)onderzoek. Dat wordt bevestigd als we afgaan op koers- en visie-documenten van onderzoeksinstituten én op wat aan regels, protocollen en duidingen van het belang ervan door dergelijke organisaties, beroepsverenigingen en internationale instituties naar voren is gebracht. De American Evaluation Association¹¹ ziet onafhankelijkheid als een van de manieren om de integriteit van onderzoek te waarborgen. Andere beroepsverenigingen van beleids- en evaluatieonderzoek zitten doorgaans op dezelfde lijn (bijv. Degeval, 2016: 20). De OECD (2021) ziet onafhankelijkheid als een ‘cornerstone with the goal of ensuring quality for the evaluation process and the evaluation product’. Ook de toelichtingen op onafhankelijkheid, bijvoorbeeld van Nederlandse rijkskennisinstellingen zoals de Planbureaus, het CBS, RIVM en WODC, zien onafhankelijkheid als een kernwaarde.

Genoemde bronnen maken daarbij doorgaans niet expliciet wat er, in operationele termen, onder wordt verstaan. De Onderzoeksraad voor Veiligheid (OVV, 2018) probeert dat wel en legt een verbinding tussen onafhankelijkheid en rollen van opdrachtgever en opdrachtnemer van beleidsonderzoek: ‘Onafhankelijke onderzoekers onderkennen dat hun werk plaatsvindt in een politiek-bestuurlijk krachtenveld – maar laten zich niet leiden door de belangen die daar spelen. En zij

11 Zie www.eval.org/About/What-is-Evaluation.

hoeven zich daar ook niet door te laten leiden; de crux van onafhankelijk onderzoek is immers dat belanghebbenden geen grip hebben op de aanpak en uitkomst ervan' (OVV, 2018: 14). De OVV onderscheidt drie dimensies van onafhankelijkheid:

- 1 *Onafhankelijkheid in positie*. Dit betreft de randvoorwaarden voor het verrichten van onafhankelijk onderzoek. Denk aan het afwezig zijn van een gezagsrelatie tussen de opdrachtgever/aanvrager of het object van het onderzoek en de onderzoekers, zeggenschap over de onderzoeksvraag en over het al dan niet starten van een onderzoek, eigen financiering en huisvesting en toegang tot alle gewenste informatiebronnen.
- 2 *Onafhankelijkheid in oordeelsvorming*. Hier gaat het om de vereiste vaardigheden en attitude om te komen tot een eigenstandig oordeel, zoals het voorkomen van persoonsgebonden vertekeningen en biases, om onderzoeksethiek, het organiseren van tegenspraak, het opstellen van de onderzoeksverantwoording en het bieden van weerstand tegen druk van buitenaf.
- 3 *Onafhankelijkheid in beeldvorming*. Dit betreft het uitbrengen van het onderzoek op een manier die de kans vergroot dat de maatschappij het onderzoek ervaart als onafhankelijk. De OVV noemt onder andere als elementen: een weloverwogen besluit nemen over het tussentijds delen van informatie; streven naar een goede balans tussen kwaliteit en tijdigheid; zoveel mogelijk de regie houden over de boodschap van het onderzoek; terughoudend zijn in verdere bemoeienis met het onderwerp van onderzoek.

Alhoewel deze omschrijving van onafhankelijkheid behoorlijk gedetailleerd is, zijn we van oordeel dat naarmate de omschrijving en afbakening van het begrip meer gedetailleerd zijn, het ook aan realiteitszin inboet. In welk beleidsonderzoek is de financiering geheel vrij of ongelimiteerd, worden onderzoeksmethoden alleen door de onderzoekers gekozen (en niet in overleg met bijvoorbeeld (wetenschappelijke) begeleidingscommissies) of is de looptijd van het onderzoek een open einde? Ook doet zich de vraag voor hoe zich wat op papier staat, verhoudt tot sociale processen tussen onderzoekers en met derden zoals opdrachtgevers en 'evaluands'. Absolute onafhankelijkheid bestaat onzes inziens in de praktijk niet, alleen een zekere mate van onafhankelijkheid. Gläser et al. (2022) laten daarnaast zien dat er verschillen tussen vakgebieden zijn, zoals de sociologie, de economie, de rechtswetenschap en de wetenschapsleer, als het over de vraag gaat wat onafhankelijkheid constitueert. Het is een fluïde begrip en dat maakt dat er discussies over mogelijk zijn. Die worden niet alleen tussen opdrachtgever en opdrachtnemer gevoerd, maar ook in de maatschappij, zoals uit de vorige paragraaf is gebleken.

Kwaliteit

Wij zien kwaliteit als het tweede kernbegrip van beleidsonderzoek. Maar wat moet eronder worden verstaan? Het lijkt een multidimensionaal begrip te zijn: naast methodologische dimensies en criteria waaraan voldaan moet worden, zoals de door de onderzoekers bepaalde probleemstelling en scope van een onderzoek, accuraatheid (van bijv. cijfers), diverse vormen van validiteit; betrouwbaarheid, een bij de probleemstellingen passend onderzoekdesign; coherentie (tussen bevin-

dingen en conclusies) en repliceerbaarheid, zijn er andere criteria, zoals relevantie, benutbaarheid, vergelijkbaarheid (transparantie), tijdigheid, toegankelijkheid en aandacht voor ethische en communicatieve aspecten van het onderzoek.

Hoe belangrijk deze *andere dimensies* ook zijn, boven alles gaat het ons om de methodologische kwaliteitsaspecten van beleidsonderzoek. Als daar vraagttekens bij worden gezet, dan is die beoordeling in onze visie belangrijker dan de beoordeling op de andere dimensies. Immers, het gaat om (toegepast) *wetenschappelijk onderzoek*. We opteren daarmee voor een restrictieve duiding van het kwaliteitsbegrip.

Aan deze stellingname liggen verschillende overwegingen ten grondslag. Eén is dat uit tal van systematische reviews van (beleids)onderzoek keer op keer blijkt hoe veel er (nog steeds) mis is met aspecten van de methodologie. Denk aan werk van de Campbell Collaboration, die dergelijke reviews publiceert.¹² Dat geldt ook voor die meta-evaluaties die de methodologische dimensies van het kwaliteitsbegrip centraal stellen. Een recent (internationaal) voorbeeld is de evaluatie van vijf jaar Wereldbank-evaluaties, uitgevoerd door, of in opdracht van, haar Independent Evaluation Group, waarin rapporten werden doorgelicht op basis van criteria zoals probleemstelling, scope en focus van de probleemstellingen, betrouwbaarheid, verschillende typen validiteit en coherentie. Aandacht voor benutting en andere criteria bleef in de meta-studie expliciet buiten beschouwing (World Bank, 2021).

De derde overweging is dat ook in maatschappelijke debatten over onder andere gezondheid en voedselveiligheid aandacht wordt besteed aan methodologische criteria waaraan (evaluatie)onderzoek dient te voldoen. Zie voor Nederland bijvoorbeeld columns van Katan (NRC), Janssens (NRC) en Smeets (Volkskrant).¹³ De vragen die gesteld worden, betreffen regelmatig of de conclusies die getrokken zijn, passen bij de opzet van het onderzoek; hoe robuust de data-verzameling en -analyse is, of bij de data-verzameling rekening is gehouden met artefacten, enzovoort.

Het belang van de methodologie staat in ons kwaliteitsbegrip dus centraal. Wie dat belang niet als uitgangspunt hanteert, vergroot de kans dat een qua methodologie twijfelachtig onderzoek dat wel hoog op enkele niet-methodologische criteria scoort, toch als kwalitatief goed gezien wordt en als zodanig gebruikt wordt. Die benadering is niet de onze.¹⁴

12 Sinds een aantal jaren staat de zogenoemde replicatiecrisis hoog op de agenda. Kern ervan is dat de resultaten die studies laten zien, lang niet altijd te repliceren zijn, onder andere omdat er problemen zijn met de gekozen methodologie.

13 Zie voor meer informatie over deze columns, deze sites: www.nrc.nl/rubriek/martijn-katan/; www.volkskrant.nl/columns/Ionica-Smeets. Zie voor columns van de onlangs overleden epidemioloog prof. C. Janssens: www.nrc.nl/nieuws/2022/09/13/onvermoeibaar-streed-cecile-janssens-tegen-jubelwetenschap-a4141708.

14 Een van de reviewers van een eerdere versie van deze bijdrage stelde de vraag of er een ondergrens is bij het beoordelen van de kwaliteit van beleidsonderzoek. Zogenaamde 'evidence hierarchies' die door onder andere de Campbell- en Cochrane-organisaties gebruikt worden in review-werk, leveren daar materiaal voor aan, maar deze aanpak is niet vrij van kritiek.

4 Opdrachtgevers, onderzoekers en onafhankelijkheid

De volgende twee paragrafen zien op wat het krachtenveld van beleidsonderzoek genoemd kan worden: welke actoren interacteren (met elkaar) en hoe werkt dit door op de onafhankelijkheid van onderzoek?

Voor beleidsonderzoek, of dat nu sociaalwetenschappelijk, historisch of van andere disciplinaire aard is, is de interactie van erbij betrokken partijen van belang. Immers, 'de crux van onafhankelijk onderzoek is dat belanghebbenden geen grip hebben op de aanpak en uitkomst ervan', zoals de OVV (2018: 14) het verwoordt. Maar opdrachtgevers zijn (doorgaans) direct belanghebbende, van wie de vraag of vragen met het onderzoek beantwoord moet(en) worden.

In de Zwitserse en Amerikaanse onderzoeken (Pleger & Hadorn, 2018; Schmidli et al., z.j.) werd primair uitgegaan van de opdrachtgever als persoon en is goeddeels voorbijgegaan aan de opdrachtgever als verlengde van de organisatie waarvoor deze optreedt én helemaal aan de invloed van derde partijen (de politiek, andere stakeholders, etc.). Om de interactie tussen opdrachtgever en opdrachtnemer te begrijpen is het van belang te beseffen dat die interactie zich in een krachtenveld afspeelt. In dat veld zijn niet alleen opdrachtgever en opdrachtnemer van een onderzoek van belang, maar ook andere partijen. Dit is het onderwerp van deze paragraaf.¹⁵

De 'tricky triangle' van beleidsonderzoek

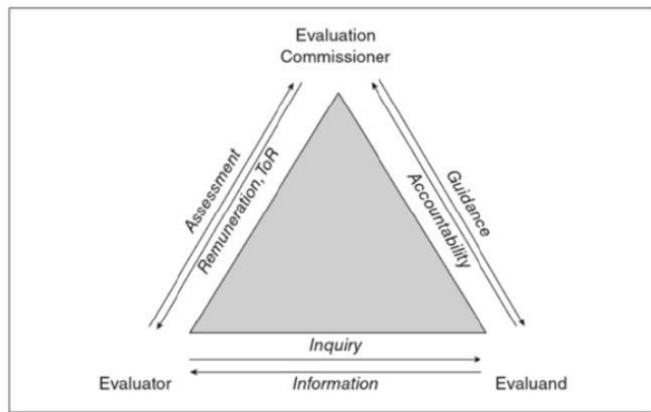
Welke partijen zijn relevant in en voor beleidsonderzoek? Als het in publicaties over de onafhankelijkheid van beleidsonderzoek gaat, staat daarbij doorgaans de relatie tussen opdrachtgever en opdrachtnemer, respectievelijk onderzoeker centraal. Dat is op zich verwonderlijk. Al vanaf de jaren zeventig tot medio jaren tachtig werkte Van de Vall aan het uitwerken van zijn professionele paradigma van sociaal beleidsonderzoek (Van de Vall, 1980). Centraal in dit paradigma stond het principe dat de beleidsonderzoeker, naast voldoen aan epistemologische principes, vooral ook een 'adviseur' moest zijn en moest interacteren met beleidsmakers, stakeholders en potentiële gebruikers van het onderzoek. Hij noemde dat 'implementatie-validiteit'. Beleidsonderzoekers moesten zich daarbij, tot op zekere hoogte, kunnen identificeren met de doelen van de cliënt. De cliënt is niet alleen de opdrachtgever. Van de Vall verwees naar Patton, grondlegger van het 'Utilization-Focused Evaluation'-paradigma, waarbij interactie, samenwerking en co-creatie tussen beleid, stakeholders, *beneficiaries* en onderzoekers de centrale uitgangspun-

- 15 We tekenen hierbij aan dat dit krachtenveld onderdeel is van een bredere context. Er zijn immers tal van andere factoren te benoemen die effect kunnen hebben op de uitvoering van een onderzoek én op de houding, het gedrag en de interactie van bij een onderzoek betrokken partijen, bijvoorbeeld: type opdrachtverlening (bijv. onderhandse aanbesteding versus openbare), wel of geen verplichting tot openbaarmaking van het onderzoeksrapport, het onderzoeksobject, wel of niet onder een maatschappelijk vergrootglas liggen van het onderzoek of de opdrachtgevende organisatie, het wel of niet 'meekijken' van een interne of externe begeleidingscommissie of een klankbordgroep, het wel of niet bestaan van protocollen en codes die iets zeggen over gewenst en ongewenst gedrag, positie van betrokkenen binnen hun organisatie, de organisatiecultuur, enzovoort.

ten zijn.¹⁶ De gedachte was dat het gebruik van de resultaten van het onderzoek dan bevorderd zou worden, alsook de relevantie.

De Laat (2014) heeft de relaties tussen opdrachtgever, opdrachtnemer en evaluand uitgewerkt in wat hij een 'tricky triangle' noemt. Hij benadrukt in die driehoeksrelatie de rol van de opdrachtgever omdat die volgens hem te weinig aandacht krijgt in discussies over onafhankelijk onderzoek. Hij wijst erop dat de opdrachtgever vaak samenvalt met de evaluand en dat daardoor de specifieke rol van de opdrachtgever niet scherp onderscheiden wordt. Denk bijvoorbeeld aan een directie binnen een publieke organisatie die haar beleid laat evalueren. De evaluand en opdrachtgever vallen dan als het ware samen. Zijn tricky triangle ziet er als volgt uit (De Laat, 2014: 23): zie figuur 1.

Figuur 1 De tricky triangle van het beleidsonderzoek



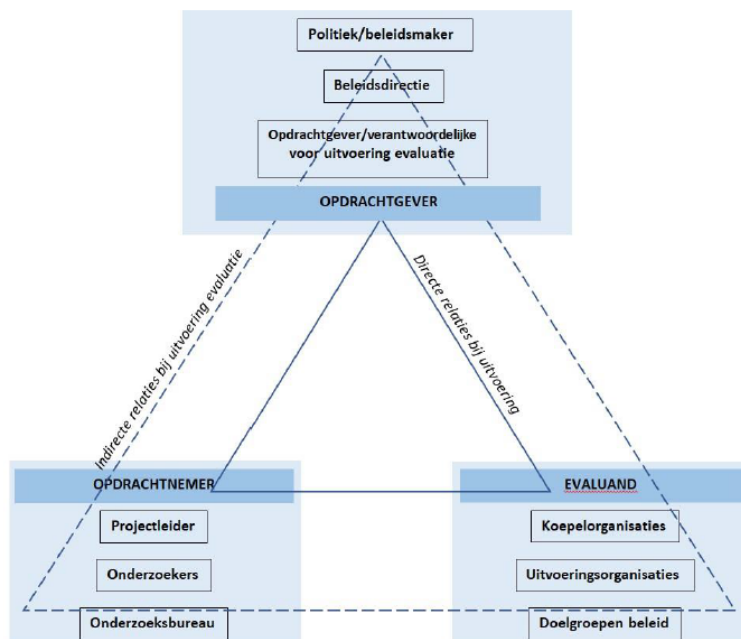
Bron: De Laat, 2014

In deze figuur zijn de door De Laat onderscheiden typen relaties tussen de partijen samengevat. Bij het vaststellen van die relaties gaat het om de rol of hoedanigheid die de direct bij een onderzoek betrokken partijen hebben. In de praktijk kan bijvoorbeeld de opdrachtgever (in de rol van begeleider van het onderzoek vanuit de organisatie) verschillende relaties hebben met de evaluand, de onderdelen van die organisatie die feitelijk geëvalueerd worden. De evaluand op zijn/haar beurt kan bestaan uit in het onderzoek betrokken informanten/respondenten, maar ook uit vertegenwoordigers van de doelgroepen van het onderzoek (bijvoorbeeld cliënten-organisaties, vakbonden, branche- of beroepsorganisaties). Een en ander betekent dat de *triangle* van De Laat meerdimensionaal is en dat daarom niet alleen de rela-

16 In Pattons werk rond 'Blue Marble evaluations' wordt dit verder uitgewerkt met het oog op de uitdagingen waar de samenleving voor staat (zoals de klimaatproblematiek): 'Design, implementation, and evaluation are typically treated as separate functions dealt with sequentially by different people with different roles who don't communicate with each other. At the heart of the Blue Marble perspective you'll find a pattern of breaking down silos, integrating separated functions, connecting people and places, and creating linkages across time' (<https://bluemarbleeval.org/about>).

ties tussen opdrachtgever, opdrachtnemer en evaluand op het netvlies moeten staan als het om de context van beleidsonderzoek gaat, maar ook wie in die rollen optreden en door wie zij weer beïnvloed worden. Daarom zou de hierboven gepresenteerde figuur in algemene zin ook als volgt weergegeven kunnen worden: zie figuur 2.

Figuur 2 De extended tricky triangle van het beleidsonderzoek



Bron: eigen bewerking Mevissen en Leeuw op basis van De Laat, 2014

Elke partij in een dergelijke driehoek kan (maar vermoedelijk: zal) in relatie tot een evaluatie eigen belangen hebben en die proberen na te streven of te beschermen. Wanneer dat het geval is, heeft dat effect op de directe relaties tussen opdrachtgever, opdrachtnemer en evaluand voor, en/of tijdens, en/of na een evaluatie. Ook in dit opzicht is onafhankelijkheid een fluïde begrip, omdat de afbakening/betekenis ervan kan verschillen tussen de verschillende direct of indirect bij een onderzoek betrokken partijen.

Onafhankelijkheid vanuit het perspectief van samenwerking en interactie

Beleidsonderzoek gaat gepaard met interacties tussen organisaties, afdelingen daarbinnen en personen, tussen opdrachtgevers, uitvoerders, stakeholders, evaluands, gebruikers, collega-onderzoekers, enzovoort. Die interactie gebeurt vanuit verschillende intenties¹⁷ en heeft zichtbare, onzichtbare, bewuste en onbewuste,

17 Zie bijvoorbeeld het BUSD-model in de volgende paragraaf.

gewenste en ongewenste invloeden op het onderzoek en het onderzoeksproduct, of kán die hebben.

Onafhankelijkheid is daarom geen absoluut begrip, maar treedt op in gradaties. Moet interactie (altijd of vooral) worden vermeden vanwege het streven naar onafhankelijkheid? Die indruk wordt soms gewekt (Commissie Hertogh, 2019: 99, 154, 159; De Graaf & Hertogh, 2020: 15). Tegelijkertijd zijn er auteurs en stromingen die het belang van dergelijke interacties juist onderstrepen.¹⁸ Immers, dergelijke interacties, resp. vormen van beïnvloeding, hoeven niet alleen negatieve connotaties en uitkomsten te hebben, maar kunnen ook dienstbaar zijn aan de kwaliteit en/of de bruikbaarheid van onderzoek, zonder dat de uitkomsten ter discussie gesteld worden. Gepaard gaand met de aandacht voor onafhankelijkheid en kwaliteit van beleidsonderzoek zijn er steeds meer ‘instrumenten’ ontwikkeld en ingevoerd (in de vorm van onder andere protocollen, regels, communicatiepraktijken en ‘early warnings’) om meer grip te krijgen op de onafhankelijkheid en kwaliteit van wetenschappelijk onderzoek en beleidsonderzoek. Soms wordt daarbij aan ‘borgingsinstrumenten’ gerefereerd. In de volgende paragraaf gaan we in op deze instrumenten.

5 Borging van onafhankelijkheid en kwaliteit van beleidsonderzoek

De kernwaarde ‘kwaliteit’ van onderzoek wordt vooral nagestreefd door (methodologische) eisen te stellen aan het onderzoek; bij ‘onafhankelijkheid’ gebeurt dat vooral door maatregelen in te voeren om dat te borgen. In deze paragraaf komen twee vragen aan de orde. De eerste is: hoe verhouden beide zich tot elkaar? De tweede is hoe effectief borging van onafhankelijkheid is en wat borgen betekent voor de kwaliteit van onderzoek.

Beleidsonderzoek moet wetenschappelijk verantwoord zijn, maar wat betekent dat? Zoals we eerder al aangaven, en zoals ook terug te vinden is in bronnen over wat kwaliteit in beleids-/evaluatieonderzoek is (Wereldbank, UN, OECD, nationale beroepsverenigingen voor evaluatieonderzoek¹⁹ en wetenschappelijke literatuur²⁰) kan de kwaliteit van dat onderzoek worden afgemeten aan hoe ‘credible’ het onderzoek is, d.w.z. of en in hoeverre het is gebaseerd op expertise, objectiviteit, transparantie en degelijke methodologie (World Bank, 2019: 5; zie voetnoot 19). Robuuste methodologie houdt onder andere in: repliceerbaarheid en validiteit, uiteengelegd in categorieën zoals constructvaliditeit, in- en externe validiteit, ‘data

- 18 Dat werd al vastgesteld in Van Oijen et al. (1984). Maar ook recent wordt gewezen op voordelen van interactie, bijvoorbeeld bij responsive evaluations, blue-marble evaluations, bij onderzoek naar de (re)constructie en toetsing van beleidstheorieën. De redenering is dat uit dergelijke interacties (door Commissie Hertogh als ‘spanningen’ geduid) belangwekkende inzichten kunnen voortkomen.
- 19 UN Evaluation Group, Norms and Standards for Evaluation (2016), www.uneval.org/document/guidance-documents; OECD, Applying Evaluation Criteria Thoughtfully (2021), <https://doi.org/10.1787/543e84ed-en>; World Bank, World Bank Group Evaluation Principles (2019), <https://ieg.worldbankgroup.org/sites/default/files/Data/reports/WorldBankEvaluationPrinciples.pdf>; Canadian Evaluation Society, Standards (2012), en diverse andere bronnen.
- 20 Onder andere: Scriven (1991), Picciotto (2013) en diverse handboeken.

analysis validity', probleem- en theorie-gerichtheid, enzovoort. Transparantie over werkwijzen moet de repliceerbaarheid vergroten, maar ook zicht geven op de mate waarin die werkwijzen onafhankelijk zijn uitgevoerd.

Onafhankelijkheid is een noodzakelijke, maar geen uitsluitende voorwaarde voor kwaliteit van onderzoek. Scriven merkt er dit over op:

'Independent evaluation can decrease certain types of bias (including) (...) extreme conflicts of interest where the evaluator is "in bed with" the program being evaluated (...) typical of much program monitoring by agencies and foundations where the monitor is usually the godfather of the program, sometimes its inventor, and nearly always its advocate at the agency. This said, independence *on its own does not guarantee* evaluation quality: relevant skills, sound methods, adequate resources and transparency are also required. Independent but incompetent evaluations can be misleading, disruptive and costly. On the other hand, evaluation quality without independence lacks credibility. This is why persons and entities that have reason to fear the outcome of an evaluation will frequently throw doubt on its independence.' (Scriven, 1991: 192-193)

Wat zijn redenen waarom onafhankelijkheid een noodzakelijke, maar geen uitsluitende voorwaarde voor de kwaliteit van beleidsonderzoek is?

- Onafhankelijkheid kan '*splendid isolation*' worden, een situatie die wellicht voor genieën tot prachtproducten leidt, maar dat geldt niet voor (elke) beleidsonderzoeker.
- Onafhankelijkheid kan een dogma worden. In de woorden van de OVV: 'Wanneer onafhankelijkheid als dogma wordt gehanteerd, zal het de kwaliteit van het onderzoek eerder in de weg staan dan eraan bijdragen' (OVV, 2018: 5).
- In het verlengde hiervan maakt Perrin (2018: 2) het punt dat voor elke auteur feedback en kritiek belangrijk zijn: 'Just about any author, be it a novelist, essayist or journalist, academic or researcher, is subject to review and feedback, no matter what it is they write. (...) Why should evaluators be treated any differently? It is unreasonable to suggest that evaluators should be immune from feedback and criticism.'
- Diverse bronnen (Diercks et al., 2019; Verwest et al., 2019; Klein Haarhuis, 2016) laten zien dat het onderhouden van contacten met 'stakeholders', beleidsmakers en doelgroepen, teneinde *au courant* te blijven waar het beleidsrelevantie en bruikbaarheid van het onderzoek betreft, niet veronachtzaamd dient te worden.

Borging van onafhankelijkheid

Gezien de samenhang tussen onafhankelijkheid en kwaliteit en de relaties die tussen partijen bestaan in een beleidsonderzoek, is het de vraag wat nodig is om onafhankelijkheid te realiseren en tegelijkertijd kwalitatief hoogwaardig onderzoek af te leveren.

Het antwoord is in de (Nederlandse) praktijk onder andere het *invoeren van waarborgen of borgingsinstrumenten ten behoeve van onafhankelijkheid*. Het Rathenau

Instituut (Diercks et al., 2019) heeft voor de Nederlandse rijkskennisinstellingen die beleidsonderzoek uitvoeren, een overzicht gemaakt van deze waarborgen (zie tabel 1).

Tabel 1 *Borgingsinstrumenten bij rijkskennisinstellingen*

Borgingsinstrument	Tegengekomen bij:
Gedragscode Integriteit Rijk	Allemaal behalve RDK
Vertrouwenspersonen/Klokkenluidersregeling	Allemaal
Gedragscode Wetenschapsbeoefening	KNMI, RIVM
Andere sectorspecifieke gedragscodes	NFI, RKD, RCE
Raad van Advies of Raad van Toezicht	KNMI, RIVM, NVWA, RKD, RWS, WODC
Periodieke externe visitatie	KNMI, RIVM, NVWA, WODC
Klachtenregeling wetenschappelijke integriteit (LOWI)	KNMI, RIVM
Aangeven nevenfuncties/belangenverstrengeling	RIVM, NVWA, NFI, RKD, KiM, WODC
Handreiking publiek-private samenwerking	RIVM, KNMI
Begeleidingscommissies	SWODC, KNMI, RCE, RWS, RIVM, RKD
Externe peerreview	Allemaal
Interne peerreview	KiM, NFI, RIVM, RCE
Formele projectplannen	WODC, KiM, RCE
Trainingen	RWS, RIVM
Intervisie / moreel beraad	WODC, KiM

Bron: Diercks et al., 2019

Ruim drie jaar na deze bestandsopname zijn er enkele uitbreidingen te melden, zoals de inrichting van de gezamenlijke Commissie Wetenschappelijke Integriteit Planbureaus: zij onderzoekt klachten, beoordeelt of de wetenschappelijke integriteit is geschonden en brengt daarover advies uit aan de directie van de organisatie waarop de klacht betrekking heeft. Ook wordt de Nederlandse gedragscode voor wetenschappelijke integriteit (KNAW et al., 2018) inmiddels door de planbureaus en (andere) rijkskennisinstellingen, zoals het RIVM, PBL, WODC en KNMI, onderschreven (Wittebrood & Braak, 2022: 41).

Van der Knaap et al. (2019: 6) maken, in lijn met Diercks et al. (2019) en Verwest et al. (2019), een onderscheid in twee soorten borging: harde maatregelen omvatten de formele juridische positie ten opzichte van de overheid, wettelijke bepalingen en geformaliseerde institutionele arrangementen (zoals gedragscodes, protocollen en adviesraden). Omdat 'papier altijd geduldig is en procedures zelden dwingend, [is de effectiviteit van deze manieren van borging] in belangrijke mate afhankelijk van de wijze waarop deze geïnternaliseerd zijn in de professionaliteit, de arbeidsroutines en de organisatiecultuur van een instituut. Dat is het terrein van de zachte borgingspraktijken. Dit zijn de praktijken die het eerbiedigen van gedragscodes bevorderen, het zich bewust zijn van de eigen rol en die van de ander, het nemen van de verantwoordelijkheid die daarmee gepaard gaat, en het elkaar

aanspreken op rolvastheid' (Verwest et al., 2019: 21). Wat Kunseler en Verwoerd (2019: 11; 49 ff) 'reflexief werken' noemen, kan daar ook onder worden gerekend: 'de wisselwerking en interactie tussen de ontwikkeling van kennis en de omgeving waarin de kennis gestalte krijgt en wordt toegepast'. De genoemde auteurs poneeren de stelling dat een *effectieve borging* van onafhankelijkheid om een *combinatie* van 'harde' en 'zachte' maatregelen vraagt.

Inmiddels is er een (internationale) praktijk ontstaan rondom borgingsinstrumenten: protocollen, (raden van) toezicht, (gedrags)codes, regelgeving inzake wetenschappelijke onafhankelijkheid; kaders voor de omgang tussen onderzoekers en 'beleid', peer reviews en visitaties; klachtenregelingen, vertrouwenspersonen, commissies voor integriteit, en websites met 'tools' hoe om te gaan met QRP's ('Questionable Research Practices') (Bouter, 2020). Het is niet verwonderlijk dat er ook onderzoek op dat gebied verricht wordt, zoals naar de rol van gedragscodes in evaluatieonderzoek (Neale et al., 2003), naar zorgplichten van onderzoeksinstituten 'to empower their research staff to steer away from QRP's' (Bouter, 2020; Gopalakrishna et al., 2022), naar gedragscodes in het onderwijs (Talankar, 2018) en de gezondheidszorg (Bijani et al., 2017; Borysowski et al., 2021), enzovoort. In Nederland voert het Rathenau Instituut onderzoek uit hiernaar en zijn er case-studies over het Planbureau voor de Leefomgeving (PBL) (Verwest et al., 2019; Kunseler & Verwoerd, 2019), het WODC ('t Hart, 2021) en over drie andere rijks-kennisinstellingen in de wereld van de gezondheidszorg (Bressers, 2022²¹). Bressers (2022: 23) maakte het thema hoe onderzoek- en adviesorganen omgaan met de spanning tussen hun streven enerzijds 'gepaste afstand te houden om kritisch en onafhankelijk te werk te kunnen gaan' (OVV, 2018) en anderzijds 'de noodzakelijkheid voldoende aangesloten te zijn', tot haar centrale vraagstelling.²² Uit haar case-studies (Nederland) en uit de OECD-landen-vergelijkende analyse blijkt onder andere dat 'onderzoekers/adviseurs benadrukken dat "onafhankelijkheid" en "zelfstandig keuzes kunnen maken" in het onderzoeks- en adviestraject belangrijke waarden zijn, die zij beschermen in de interacties met ambtenaren en politici. Ambtenaren en politici benadrukken sterk de aansluiting bij beleid en de kaders waarbinnen de organen opereren. Het is belangrijk dat onderzoek/advies relevant is en past bij het politieke discours' (Bressers, 2022: 318).

Borging van onafhankelijkheid maar ook van kwaliteit ... quod erat demonstrandum?

De centrale veronderstelling bij deze borgingsinstrumenten lijkt te zijn dat de invoering ervan niet alleen (rechtstreeks) gericht is op het realiseren van onafhankelijkheid van onderzoek maar ook, rechtstreeks dan wel via tussenstappen, een positieve invloed heeft op de verschillende dimensies van kwaliteit, de methodologische voorop. In die zin kunnen de borgingsinstrumenten als *gelaagd* worden gezien:

- 21 Bressers (2022) heeft een vergelijkend onderzoek gedaan naar de situatie in diverse OECD-landen.
- 22 Deze luidt: 'Hoe verlopen de dynamieken tussen "autonomie" en "controle" in het secundaire proces en "distantie" en "betrokkenheid" in het primaire proces in de praktijk van kennis- en adviesorganen?'

borgingsinstrument A → draagt bij aan: onafhankelijkheid en → draagt vervolgens bij aan → (bewaking of verhoging) van de kwaliteit van beleidsonderzoek

Wie verwacht dat de belangstelling voor harde en zachte borging van onafhankelijkheid gemaakt zou hebben dat de effecten ervan op *zowel onafhankelijkheid als kwaliteit van onderzoek* grondig geëvalueerd zijn, komt bedrogen uit. We hebben (onder de aantekening dat we géén systematic review hebben uitgevoerd)²³ niet veel recent onderzoek naar de effecten en neveneffecten van het invoeren van de harde/zachte maatregelen (al dan niet gecombineerd) op onafhankelijkheid en kwaliteit aangetroffen. Ons vermoeden is dat er méér aandacht is voor het *maken en invoeren* van borgingsinstrumenten dan voor (evaluatie)onderzoek naar *effecten en neveneffecten* ervan. Bij de uitspraak dat het combineren van harde en zachte instrumenten (positieve) effecten op onafhankelijkheid én kwaliteit heeft, zijn vraagtekens te plaatsen.

Op de eerste plaats heeft deze stelling vaak betrekking op kennisinstellingen van overheden. Die staan als uitvoerders van onderzoek in een andere relatie tot met name de opdrachtgevers en de politiek dan andere typen onderzoeksinstellingen. Daardoor kunnen zij zich waarschijnlijk eerder en meer beroepen op harde borgingsinstrumenten.

Op de tweede plaats zal de kennis van de beschikbare harde borging en de aandacht voor zachte borging binnen organisaties afnemen in periodes dat de vrees voor en confrontatie met beïnvloeding minder aanwezig is (bijvoorbeeld omdat er geen gevoelige onderwerpen geëvalueerd worden).

Een derde punt is dat borgingsinstrumenten zich (kunnen) richten op specifieke aspecten van onafhankelijkheid zonder aandacht te hebben voor het doel van het betreffende beleidsonderzoek. Derksen (2013) onderscheidt twee typen onderzoek: (a) gericht op kennisgebruik (bijvoorbeeld een verkennend ex ante onderzoek of een ex post beleidsevaluatie) en (b) gericht op coproductie (het samen met doelgroepen ontwikkelen en/of uitvoeren van beleid). Borging die bij type a passend is om ongewenste beïnvloeding tegen te gaan, zal vermoedelijk averechts werken bij type b (Korsten & Mevissen, 2015).

Wie empirisch onderbouwde uitspraken wil doen over de effecten op onafhankelijkheid en kwaliteit van deze (en soortgelijke) borgingsinstrumenten, zal evaluatieonderzoek moeten doen, waarin de volgende zaken aan de orde komen:

- Ten eerste moet de theorie die eraan ten grondslag ligt, worden gearticuleerd: welke (gedrags- en andere) mechanismen kunnen maken dat er van gelaagde effectiviteit sprake is zonder neveneffecten.

23 Een voorbeeld van een dergelijke studie op een verwant gebied is die van Recio-Saucedo et al. (2022) naar de vraag 'what works for peer review and decision-making in research funding', waarin deze vraag vanuit een *realist evaluation* gezichtspunt wordt beantwoord.

Jos Mevissen & Frans Leeuw

- Ten tweede zal moeten worden vastgesteld of en in hoeverre de maatregelen zoals ze zijn afgekondigd en ingevoerd, de beoogde invloed op *onafhankelijkheid* hebben.
- Ten derde moet dat gebeuren voor de verschillende dimensies van kwaliteit.
- Ten vierde zullen de *professionele competenties van onderzoekers en hun leidinggevenden en hun instituten* in ogenschouw genomen moeten worden. Hoe veerkrachtig zijn zij in het tegenspreken en afweren van op verstoring of ondermijning gerichte interacties? Hoe steviger een onderzoeker of leidinggevende in haar of zijn schoenen staat, des te groter de weerbaarheid is en des te kleiner de noodzaak om op protocollen, regels en andere borging terug te vallen. Mark (2001) benadrukt professionele capaciteiten van onderzoekers en leidinggevenden die vooral in staat (zouden moeten) zijn te voorzien in wat voor dynamiek of spel zij zich bevinden, bijvoorbeeld: zien aanvragers/beleidsmakers het onderzoek als een manier om beleid te legitimeren of zijn ze daadwerkelijk gericht op kwaliteitsbevordering, resp. leren?
- En ten vijfde moet worden vastgesteld *welke rol interacties* tussen onderzoeker en andere genoemde actoren spelen met het oog op de kwaliteit. Daarbij zal de *aard van de interacties* tussen de betrokken partijen gezien moeten worden. Regelmatig wordt over ‘interacties’ en ‘beïnvloeding’ in algemene zin gesproken. We wezen al eerder op benaderingen die juist het belang en het positieve van interacties benadrukken. Echter, ook op het omgekeerde is gewezen (negatieve beïnvloeding). Dat betekent dat er risico’s verbonden zijn aan interacties tussen beleid, stakeholders en onderzoek, maar het is niet zo dat elke vorm van interactie problematisch is of verkeerd. De beoordeling van het gewenst of ongewenst zijn van interactie (en beïnvloeding) dient in de ogen van Fynn et al. een relatie te hebben met het doel van de interactie (Fynn et al., 2021: 21). Juist op het doel van interactie, resp. beïnvloeding, grijpt het BUSD-model van Pleger en Sager (2016) aan: zie figuur 3.

Figuur 3 Het BUSD-heuristisch model van beïnvloedingstypen

		Explicitness of Pressure	
		Explicit, direct	Implicit, indirect
Direction of Influence	Destructive	Distortion <i>Negative Pressure</i>	Undermining <i>Negative Persuasion/Counsel</i>
	Constructive	Betterment <i>Positive Pressure</i>	Support <i>Positive Persuasion/Counsel</i>

Bron: Pleger & Sager, 2016

De eerste dimensie die zij noemen, ‘direction of influence’, maakt een onderscheid tussen positieve (‘constructive’) en negatieve (‘destructive’) beïnvloeding. Daarmee wordt niet bedoeld op het gedrag zelf, maar op het doel ervan en op het effect op de kwaliteit van een evaluatie. De tweede dimensie, ‘explicitness of pressure’, maakt een onderscheid tussen impliciete (‘subtle’), indirecte en expliciete

('obvious'), directe invloed uitoefenen en betreft dus de nadruk waarmee invloed wordt uitgeoefend. Deze twee dimensies leiden tot vier gedragsvormen:

- 1 *Support*: een poging tot beïnvloeding kan 'een positief, opbouwend karakter dragen: de opdrachtgever kan lopende het onderzoek bijvoorbeeld zijn steun uitspreken voor tussentijdse resultaten en/of conceptversies van een rapport'.
- 2 *Betterment*: opmerkingen en opbouwende kritiek kunnen leiden tot verbetering van het onderzoek en/of de rapportage. Bij deze vormen van invloed blijft de rolverdeling duidelijk en blijven de inhoudelijke afwegingen bij de onderzoekers rusten; de onafhankelijkheid van het onderzoek is hierbij meestal niet in het geding.
- 3 *Distortion*: de opdrachtgever kan proberen de onderzoeksresultaten bij te sturen in een door hem gewenste richting.
- 4 *Undermining*: een opdrachtgever oefent sterke druk uit op de onderzoekers en dwingt hen om de resultaten of rapportage aan te passen zonder dat daarvoor een basis in het onderzoek gevonden kan worden (Commissie Overgaw, 2018: 28 ff).

Het is de kunst van competente onderzoekers, hun organisaties en leidinggeven- den om te beoordelen met welke intentie een opdrachtgever of andere belanghebbende een poging tot beïnvloeding doet en hoe daarmee om te gaan. In geval van support en betterment is een halsstarrige houding ongunstig voor de kwaliteit van het onderzoek of het onderzoeksproduct. In geval van distortion en undermining is diezelfde vasthoudendheid juist heel zinvol, zo niet cruciaal.

Bij welke van de vier gedragsvormen uit het model kunnen harde en zachte borgingsmaatregelen van onafhankelijkheid een effect hebben?

Bij *Support* is de onafhankelijkheid niet in het geding. Support kan echter wel een extra aanmoediging voor de kwaliteit van een onderzoek betekenen. In geval van *Betterment* kunnen harde én zachte vormen van borging ervoor zorgen dat de opdrachtgever binnen de lijntjes van het toelaatbare blijft. Ook hier zal waarschijnlijk een positief effect optreden op de kwaliteit van het onderzoek.

Bij *Distortion* werken zachte borgingsmechanismen onvoldoende en moeten de harde maatregelen hun werk doen. Dat kan echter alleen als de opdrachtnemer of een van de andere partijen er een beroep op kan doen. Of dat het geval zal zijn en of dat effect zal hebben, hangt af van de bekendheid met de mogelijkheden die de borgingsmaatregelen bieden, het al dan niet overeenkomen van belangen, de (machts)verhoudingen tussen de betrokken partijen en de mate van handhaving van de maatregelen. Hetzelfde geldt in geval van *Undermining* van een onderzoek.

Op basis van het bovenstaande kunnen we het volgende concluderen:

- Veel is niet bekend over de effecten en neveneffecten van het invoeren van (combinaties van) harde en zachte maatregelen, gericht op het realiseren van onafhankelijkheid.²⁴
- Dat lijkt ook te gelden voor de vraag welke doorwerking de borging van onafhankelijkheid heeft op de kwaliteit van het onderzoek.
- Tegelijkertijd worden diverse harde en zachte maatregelen uitgerold, doorgaans vanuit de gedachte dat het (altijd) dienstig en goed is dat te doen.
- Hoe harde en zachte maatregelen (al dan niet gecombineerd) uitpakken voor de kwaliteit van het onderzoek, zou vaker onderwerp van onderzoek moeten zijn. Het reconstrueren en toetsen van de theorie achter de claim dat gecombineerde borgingsinstrumenten positief inwerken op onafhankelijkheid én kwaliteit, zonder schade te berokkenen, is daarbij slechts een eerste stap.
- In discussies over onafhankelijkheid en kwaliteit wordt te weinig geëxpliciteerd welke kwaliteitsaspecten in het geding zijn.

6 Tot besluit

We stellen vast dat er veel aandacht is voor de onafhankelijkheid van beleidsonderzoek, maar veel minder voor wat dat eigenlijk is, hoe het te bewerkstelligen is en wat het voor effecten heeft op de verschillende dimensies van de kwaliteit van beleidsonderzoek. Hierdoor krijgt het begrip iets Babylonisch. Iedereen vindt het een kernwaarde van beleidsonderzoek, maar men geeft er een eigen betekenis aan, die varieert van (zeer) breed tot behoorlijk restrictief. Ook lijkt het vanzelfsprekend te zijn dat andere partijen die betekenis(sen) kennen en delen. De gewenste gelaagdheid in effecten (van instrument naar ‘onafhankelijkheid’ en van daar naar ‘kwaliteit’) lijkt ook een zekere vanzelfsprekendheid te hebben, terwijl ook daar vraagtekens bij te plaatsen zijn.

De vraag wat onafhankelijkheid doet met een onderzoek, wordt doorgaans beantwoord vanuit de (impliciete) veronderstelling dat als een opdrachtgever of anderen zich bemoeien met een onderzoek, dat nadelig is. De uitkomsten zouden niet meer te vertrouwen zijn. Op ongewenste effecten van onafhankelijkheid als ‘dogma’ wordt veel minder gewezen door voorstanders van ‘de’ onafhankelijkheid.²⁵

Met borgingsinstrumenten (doorgaans gedragscodes, procedures, voorschriften en zachte praktijken) wordt onafhankelijkheid van onderzoekers nagestreefd. Tegelijkertijd zijn er anderen die juist het belang van interactie tussen onderzoekers en derden toejuichen.

24 Dahler-Larsen en Raimondo presenteerden tijdens het European Evaluation Society-congres (juni 2022) hun ‘Copenhagen Framework for Sound Evaluation Systems’, met als een van de tien principes dit: ‘While the independence of evaluations might be an important principle, it should be balanced against the ability of the evaluation system to understand and respect how problems and solutions appear in the local situations in which the evaluations are embedded.’

25 De eindrapportage van ‘t Hart (2021) over het WODC, enkele jaren na invoering van verschillende borgingsmaatregelen, heeft tegen die achtergrond een wat pikante titel: ‘Terug Naar Gezond Verstand en het Goede Gesprek’.

De vraag is relevant of en hoe borgingsinstrumenten werken en welke effecten ze hebben. Die vraag krijgt in onderzoek en beleid nog weinig aandacht. Dat is te betreuren.

We hebben in dit artikel al diverse factoren en mechanismen de revue laten passeren die van invloed zijn op de mate van onafhankelijkheid en typen kwaliteit. Op dergelijke factoren en mechanismen moeten borgingsinstrumenten aangrijpen. Recent besteden Wittebrood en Braak (2022) aandacht aan de vraag hoe de onderzoeker om kan of moet gaan met wat zij ‘onderzoeksgedoe’ noemen: verschijnselen die lang niet altijd direct waarneembaar zijn, maar het integer doen van wetenschappelijk onderzoek lastig maken (p. 17).²⁶ Hun oplossingen liggen niet zozeer in (harde) borgingsinstrumenten, maar in gedrag van de onderzoeker (en diens omgeving).

Dit sluit aan bij onze bevinding in de tweede paragraaf dat de nadruk in discussies over onafhankelijkheid ligt bij de rol van de onderzoeker. In paragraaf 5 hebben we vastgesteld dat er te weinig bekend is over de impact van harde en zachte borgingsinstrumenten. In ieder geval is het inventariseren van maatregelen en het bedenken van nieuwe niet voldoende.²⁷ Daarom is het naar onze mening belangrijk om meer aandacht te gaan besteden aan de relatie tussen onafhankelijkheid van beleidsonderzoek en de kwaliteit ervan. Juist omdat kwesties over de onafhankelijkheid en kwaliteit van onderzoek spelen in complexe situaties met veel verschillende belangen, is het daarnaast en daarenboven van belang om niet meer alleen naar de rol van de onderzoeker te kijken, maar naar de rol van alle direct en indirect betrokken partijen.

Literatuur

- Auditdienst Rijk. (2021). *Onderzoek naar de totstandkoming van rapporten van de Inspectie Justitie en Veiligheid*. Den Haag: Auditdienst Rijk, Ministerie van Financiën.
- Beekers, W. (2007). Tussen Scylla en Charybdis: Over een kleine eeuw historisch-wetenschappelijk onderzoek in opdracht. In L. Dorsman & P. Knegtmans (red.), *Onderzoek in opdracht: De publieke functie van het universitaire onderzoek in Nederland sedert 1876*. Hilversum: Verloren.
- Bijani, M. et al. (2017). An evaluation of adherence to ethical codes among nurses and nursing students. *Journal of Medical Ethics and History of Medicine*, 10(6).
- Blom, J. (1984). Historical research as an answer to critical political questions: The example of the Menten case. *The Public Historian*, 6(4), 37-48.

- 26 Denk aan onduidelijkheid over de professionele rol en de daarbij passende verantwoordelijkheden van onderzoekers, de rol van het morele kompas van onderzoekers in relatie tot wat zij ervaren (en ‘voelen’) in de omgang met anderen of aan de rol en het morele kompas van het management van een instelling. Daarnaast kan er onderzoeksgedoe ontstaan omdat onderzoekers ook gekenmerkt kunnen worden door biases in hun handelen en oordelen, waarop zij door anderen gewezen worden. Coffeng (2022) constateerde bijvoorbeeld dat er sprake is van ‘blind spot bias’ (= blinde vlek voor eigen denkfouten) bij onderzoekers van Nederlandse inspecties en toezichhoudende autoriteiten.
- 27 In mei 2022 kondigde de minister van OCW aan de gedragscode wetenschappelijke integriteit te gaan evalueren. Zie www.scienceguide.nl/2022/05/dijkgraaf-laait-na-incidenten-gedragscode-wetenschappelijke-integriteit-evalueren.

- Borysowski, J., Ehni, H.-J., & Górski, A. (2021). Ethics codes and medical decision making. *Patient Education and Counseling*, 104(6), 1312-1316. doi:10.1016/j.pec.2020.10.034
- Bouter, L. (2020). What research institutions can do to foster research integrity. *Science and Engineering Ethics*, 26, 2363-2369. <https://doi.org/10.1007/s11948-020-00178-5>
- Bressers, D. (2022). *Adviseren of afstand: Een onderzoek naar de verbinding tussen de rijksoverheid en kennis en adviesorganen*. Dissertatie Rotterdam.
- Bundi, P., Frey, K., & Widmer, T. (2021). Does evaluation quality enhance evaluation use? *Evidence & Policy: A Journal of Research, Debate and Practice*, 17(4), 661-687.
- Coenen, A.W.M. (1984). Een hoogleraar schrijft over beleidsonderzoek? *Beleid en Maatschappij*, 11(10), 313-315.
- Coffeng, T. (2022). *Bias in supervision: A social psychological perspective on regulatory decision-making*. DOI: 10.33540/1062. PhD Utrecht.
- Commissie Hertogh. (2019). *Ongemakkelijk onderzoek: Naar een betere balans in de relatie tussen WODC en beleid*. Onderzoekscommissie WODC II. Den Haag.
- Commissie Overgaw. (2018). *Betrouwbaar onderzoek*. Onderzoekscommissie I inzake deugdelijkheid drugsonderzoeken. Den Haag.
- Council of Europe. (2019). *Evaluation policy*. Strasbourg.
- Dahler-Larsen, P. (2018). The skeptical turn in evaluation. In J.E. Furubo & N. Stame (Eds.), *The evaluation enterprise* (pp. 58-80). New York: Routledge.
- Degeval Gesellschaft für Evaluation e.V. (2016). *Standards für Evaluation*. Mainz.
- De Laat, B. (2014). Evaluator, evaluand, evaluation commissioner: A tricky triangle. In M. Läubli Loud & J. Mayne (Eds.), *Evaluation use: Insights from internal evaluation units* (pp. 15-37). San Francisco: Sage.
- Derksen, W. (2013). *Kennis en beleid verbinden: Praktijkboek voor beleidsmakers*. Den Haag: Boom Lemma uitgevers.
- Derksen, W. (2018). *Over betrouwbare onderzoekers en slinkse ambtenaren* (blog). <https://www.wimderksen.com/>
- Derksen, W. (2019). *We verbannen het WODC naar een eiland* (blog). <https://www.wimderksen.com/>
- De Ruij, L. (2019). De WODC-affaire: een kritische blik op het werk van de onderzoekscommissie-Hertogh. *Beleidsonderzoek Online*, juni. DOI:10.5553/BO/221335502019000004001
- Diels, A.E. (1951). *Opvattingen van ondertrouwde vrouwen omtrent de grootte van haar toekomstig gezin. Deel 1*. Commissie Geboorteonderzoek, publicatie no. 3. Amsterdam: Instituut voor Sociaal Onderzoek van het Nederlandse Volk.
- Diercks, G., Faasse, P., Van der Meulen, B., & Diederens, P. (2019). *Met gepaste afstand: Onafhankelijkheid en integriteit bij onderzoek door rijkskennisinstituten*. Den Haag: Rathenau Instituut.
- Dorsman, L., & Knegtmans, P. (red.). (2007). *Onderzoek in opdracht: De publieke functie van het universitaire onderzoek in Nederland sedert 1876*. Hilversum: Verloren.
- Fynn, J.F., Hardeman, W., Milton, K., & Jones, A. (2021). Exploring influences on evaluation practice: A case study of a national physical activity programme. *International Journal of Behavioral Nutrition and Physical Activity*. doi.org/10.1186/s12966-021-01098-8
- Gastelaars, M. (1986). De Nederlandse sociologie en haar planmatigheidsidealen: De droom van macht en maakbaarheid. *Mens en Maatschappij*, 61(1), 5-20.
- Gläser, J. et al. (2022). The independence of research: A review of disciplinary perspectives and outline of interdisciplinary prospects. *Minerva*, 60, 105-138.

- Gopalakrishna, G., Wicherts, J.M., Vink, G., Stoop, I., Van den Akker, O.R., Ter Riet, G., & Bouter, L.M. (2022). Prevalence of responsible research practices among academics in The Netherlands. *F1000Research*, 11, 471.
- Houppermans, M., Van Hoesel, P., & Van Nispen, F. (2015). De geschiedenis van beleidsonderzoek in Nederland. In P. van Hoesel, J. Mevissen, & B. Dekker (red.), *Kennis voor beleid: Beleidsonderzoek in Nederland* (pp. 23-55). Assen: Van Gorcum.
- Jolles, H. (1978). Beleid en sociaal-wetenschappelijk onderzoek: Kanttekeningen bij een meerjarenplan, *Beleid & Maatschappij*, 9(2), 227-235.
- Klein Haarhuis, C.M., m.m.v. A. Parapuf. (2016). *Evaluatievermogen bij beleidsdepartementen: Praktijken rond uitvoering en gebruik van ex post beleids- en wetsevaluaties*. Cahier 2016-5. Den Haag: WODC.
- KNAW et al. (2018). *Nederlandse gedragscode wetenschappelijke integriteit*. <https://doi.org/10.17026/dans-2cj-nvwu>
- Köbben, A., & Tromp, H. (1999). *De onwelkome boodschap, of hoe de vrijheid van wetenschap bedreigd wordt*. Amsterdam: Mets.
- Korsten, A., & Mevissen, J. (2015). Risico's en remedies. In P. van Hoesel, J. Mevissen, & B. Dekker (red.), *Kennis voor beleid; Beleidsonderzoek in Nederland* (p. 247-276). Assen: Van Gorcum.
- Kunseler, E., & Verwoerd, L. (2019). *Kennis met impact: Reflexief werken bij publieke kennisorganisaties*. Den Haag: PBL.
- Mark, M. (2001). Evaluation's future: Furor, futile, or fertile? *American Journal of Evaluation*, 22(3), 457-479.
- Matthijssen, N. (2022, 10 februari). Een overheid die schraapt, herschrijft, stuurt. *De Groene Amsterdammer*. (<https://www.groene.nl/artikel/een-overheid-die-schraapt-herschrijft-stuurt>)
- Neale, J., Owen, J., & Small, D. (2003). Encouraging the use of codes of behavior in evaluation practice. *Evaluation and Program Planning*, 26, 29-36.
- OECD. (2020). *How can governments leverage policy evaluation to improve evidence informed policy making? Highlights from an OECD comparative study*. Paris: OECD.
- OECD. (2021). *Applying evaluation criteria thoughtfully*. Paris: OECD.
- Onderzoeksraad voor Veiligheid (OVV). (2018). *Onafhankelijk onderzoek in het publiek belang*. Den Haag: OVV.
- Onderzoeksraad voor Veiligheid (OVV). (2022). *Aanpak coronacrisis. Deel 1: tot september 2020*. Den Haag: OVV.
- Patton, M. (2019). *Blue marble evaluation: Premises and principles*. New York: Guilford.
- Perrin, B. (2018). How to manage pressure to change reports: Should evaluators be above criticism? *American Journal of Evaluation*, 23, 1-22 .
- Picciotto, R. (2013). Evaluation independence in organizations. *Journal of MultiDisciplinary Evaluation*, 9(20), 18-32.
- Pleger, L.E., & Hadorn, S. (2018). The big bad wolf's view: The evaluation clients' perspectives on independence of evaluations. *Evaluation*, 24(4), 456-474.
- Pleger, L., & Sager, F. (2016). Die Beeinflussung in der Evaluationstätigkeit in der Schweiz und was die Seval dagegen tun kann. *Leges*, (5), 33-49.
- Pleger, L.E., & Sager, F. (2018). Betterment, undermining, support and distortion: A heuristic model for the analysis of pressure on evaluators. *Evaluation and Program Planning*, 69, 166-172.
- Pleger, L., Sager, F., Morris, M., Meyer, W., & Stockmann, W. (2017). Are some countries more prone to pressure evaluators than others? Comparing findings from the United States, United Kingdom, Germany, and Switzerland. *American Journal of Evaluation*, 38(3), 315-328.

- Raimondo, E., & Leeuw, F.L. (2021). Evaluation systems and bureaucratic capture: Locked in the system and potential avenues for change. In B. Perrin & T. Tyrrell (Eds.), *Changing bureaucracies adapting to uncertainty, and how evaluation can help* (chapter 9). London: Routledge.
- Recio-Saucedo, A., Crane, K., Meadmore, K., Fackrell, K., Church, H., Fraser, S., & Blatch-Jones, A. (2022). What works for peer review and decision-making in research funding: A realist synthesis. *Research Integrity and Peer Review*, 7, 2. <https://doi.org/10.1186/s41073-022-00120-2>
- Schmidli, F.H., Pleger, L.E., & Hadorn, S. (z.j.). *Don't you forget about me: Independence of evaluations from the perspective of U.S. evaluation clients*. Draft.
- Schuyt, C.J.M. (1984). De kloof tussen weten en willen. *Beleid en Maatschappij*, 11(1-2), 21-28.
- Scriven, M. (1991). *Evaluation thesaurus*. Newbury Park: Sage.
- Sommer, M. (2020, 19 juni). Corona maakt een eind aan illusies over eigen politieke koers (column). *de Volkskrant* (<https://www.volkskrant.nl/columns-opinie/corona-maakt-een-eind-aan-illusies-over-eigen-politieke-koers~b9d9de68>)
- Stern, E. et al. (2012). *Broadening the range of designs and methods for impact evaluations*. London: Department for International Development.
- Storeng, K.T. et al. (2019). Action to protect the independence and integrity of global health research. *BMJ Global Health*, doi:10.1136/bmjgh-2019-001746
- Talanker, S. (2018). Teaching evaluation tools as robust ethical codes. *Ethics and Education*, 13(2), 221-233. doi:10.1080/17449642.2017.1412165
- Ter Hoeven, A. (1956). De politieke rol van het sociale onderzoek. *Sociologische Gids*, 3(10), 185-194.
- 't Hart, P. (2021). *Terug Naar Gezond Verstand en het Goede Gesprek: Eindrapportage transitie-adviseur relatie onderzoek en beleid op het ministerie van Justitie en Veiligheid*. <https://www.rijksoverheid.nl/documenten/rapporten/2021/12/03/tk-bijlage-eindrapport-transitieadviseur-relatie-wodc-beleid>
- Van Amersfoort, J. (2007). De Amsterdamse sociale geografie tussen autonomie en nuttigheid. In L. Dorsman & P. Knegtmans (red.), *Onderzoek in opdracht: De publieke functie van het universitaire onderzoek in Nederland sedert 1876* (pp. 129-148). Hilversum: Verloren.
- Van den Berg, R., Hawkins, P., & Stame, N. (Eds.). (2021). *Ethics for evaluation: Beyond 'doing no harm' to 'tackling bad' and 'doing good'*. New York: Routledge.
- Van de Vall, M. (1980). *Sociaal beleidsonderzoek: Een professioneel paradigma*. Alphen aan den Rijn: Samson.
- Van de Vall, M. (1984). Beleidsonderzoek in discussie: Commentaar bij: C.J.M. Schuyt 'De kloof tussen weten en willen'. *Beleid en Maatschappij*, 11(10), 302-305.
- Van der Knaap, P., Bokhorst, M., & Kleistra, Y. (2019). Onafhankelijk onderzoek in positie, oordeelsvorming en beeldvorming: Institutionele en professionele dimensies van 'speaking truth to power'. *Bestuurskunde*, 28(4), 3-11. doi:10.5553/Bk/0927338720190290040012021
- Van Hoesel, P. (1984). De waarde van beleidsonderzoek: Reactie op een betoog vol misvattingen. *Beleid en Maatschappij*, 11(10), 305-311.
- Van Hoesel, P. (2017). Beleidsevaluatie als ritueel. *Beleidsonderzoek Online*, februari. DOI:10.5553/BO/221335502017000001001
- Van Oijen, P. (1987). *Beleid voor kennis: Bijdrage tot een methodologie van de vraag*. Dissertatie Utrecht.

- Van Oijen, P., Snellen, I., & Van Westerlaak, J. (1984). *Ambtenaren en onderzoekers: Een vergelijkende studie naar de samenwerking tussen onderzoek en beleid bij 45 onderzoeksprojecten in opdracht van 10 departementen*. Den Haag: Staatsuitgeverij.
- Van Puffelen, F. et al. (1989). *Opdrachtgevers onderzocht*. Amsterdam: SEO.
- Van Vollenhoven, P., & Eijlander, P. (2022, 24 april). Nog steeds zijn toezichthouders niet onafhankelijk van de ministers. *de Volkskrant* (<https://www.volkskrant.nl/columns-opinie/opinie-nog-steeds-zijn-toezichthouders-niet-onafhankelijk-van-de-ministers~bb67df21>)
- Verwest, F., Kunseler, E., Dieren, P., & Faasse, P. (2019). Van procedure naar praktijk: Inzet op effectieve onafhankelijkheidsborging bij het Planbureau voor de Leefomgeving. *Bestuurskunde*, 29(4), 19-31.
- Winkels, J. (1982). Het Instituut voor Sociaal Onderzoek van het Nederlandse Volk: Een bijdrage aan de geschiedschrijving van het sociaal-wetenschappelijk onderzoek in Nederland. *Sociologische Gids*, 29, 98-128.
- Wittebrood, K., & Braak, R. (2022). *Onderzoeksgedoe*. Den Haag: Boom bestuurskunde 2022.
- World Bank. (2021). *A meta-evaluation of independent evaluation group evaluations (fiscal years 2015-19)*. Independent Evaluation Group. Washington, DC: World Bank.
- Zouridis, S. (2020). De elektronegativiteit van justitie: Waarom het ministerie van Justitie (en Veiligheid) zo incident-gevoelig is. In C. van Baalen (red.), *Het geplaagde ministerie van Justitie: Jaarboek parlementaire geschiedenis 2020* (p. 11-31). Amsterdam: Boom uitgevers.