

Wetenschap in crisis

Citation for published version (APA):

Martens, R. (2012). Wetenschap in crisis. *OnderwijsInnovatie*, september 2012(3), 9-11.

Document status and date:

Published: 01/09/2012

Document Version:

Peer reviewed version

Document license:

CC BY

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

<https://www.ou.nl/taverne-agreement>

Take down policy

If you believe that this document breaches copyright please contact us at:

pure-support@ou.nl

providing details and we will investigate your claim.

Downloaded from <https://research.ou.nl/> on date: 02 Jul. 2022

Open Universiteit
www.ou.nl



Wetenschap in crisis

Demissionair staatssecretaris Zijlstra laat naar aanleiding van de incidenten van de laatste maanden onderzoeken hoe de academische wereld kan worden beschermd tegen wetenschapsfraude. Een overdreven reactie, of is er echt iets aan de hand? Martens vraagt zich af of we überhaupt nog wel kunnen vertrouwen op onderwijs-onderzoek of op wetenschap in het algemeen. Volgens hem is er een flink probleem.

Er woedt al enkele jaren een veenbrand onder wetenschappelijk onderzoek, met als maatschappelijke tendens om het minder serieus te nemen. Dat komt deels door de stijging van het opleidingsniveau waardoor mensen minder geïntimideerd zijn door wetenschappelijke claims. Dit gezonde wantrouwen blijkt zich zelfs uit te betalen: hoog opgeleide kankerpatiënten zijn kritischer over en tijdens hun behandeling en hebben daardoor hogere overlevingskansen. En door internet is informatie veel makkelijker vindbaar. Dat alles maakt mensen kritischer tegenover wetenschap.

Verzonnen

Affaires die het wetenschapsimago aantasten volgen elkaar snel op, zoals de fraude door het klimaatpanel, het overdrijven van griepisico's en misleidend onderzoek over pesticiden. De VSNU heeft haar integriteitsbeleid al aangescherpt, maar de stroom aan negatieve berichten gaat door en raakt veel wetenschapsgebieden. Eén op de zeven artsen heeft met eigen ogen gezien dat resultaten werden verzonnen, zo meldt het NRC Handelsblad. Bijna een kwart heeft weleens meegemaakt dat resultaten werden bewerkt om het gewenste resultaat te krijgen. Eenzelfde aantal artsen zag collega's alleen de gegevens gebruiken die goed uitkwamen, aldus artsenblad Medisch Contact na een enquête onder achthonderd geneeskundigen. Veel medisch hoogleraren wantrouwen wetenschappelijke publicaties en meer dan de helft vindt de publicatiedruk te hoog, aldus hoogleraar Smulders. Sinds vorig jaar september is deze veenbrand een uitslaande brand geworden. De mediagenieke hoogleraar/decaan Stapel bleek vrijwel al zijn onderzoeksgegevens verzonnen te hebben. 'Leuke', positieve resultaten publiceren is blijkbaar makkelijker dan onduidelijk of negatieve resultaten. Publicatiedruk leidt tot publicatiedrift.

Fraudejager

Een nieuwe fraudeaffaire die veel aandacht trok betreft de op non-actief gestelde hoogleraar consumentengedrag Smeesters. Zijn verweer tegen de beschuldiging van de Amerikaanse fraudejager Simonsohn dat resultaten van Smeesters te positief zijn, is echter opmerkelijk: hij heeft niets

verzonnen, alleen dingen weggelaten. "Smeesters heeft herhaaldelijk aangegeven dat de cultuur binnen zijn werkveld en afdeling zodanig is dat hij zich als persoon niet schuldig voelt. Hij is ervan overtuigd dat velen (...) gebruikmaken van bewust weglaten van data om significantie te bereiken zonder dit te vermelden", zo concludeerde de onderzoekscommissie Wetenschappelijke Integriteit. Smeesters vindt dat hij gezien de publicatiedruk en de wetenschappelijke mores eigenlijk niets verkeerd heeft gedaan en dat de regels omtrent selectief datagebruik vaag zijn. En inderdaad, in vrijwel alle vakgebieden wordt gemord over die hoge publicatiedruk. Zo brak onlangs bij de VU een kleine opstand uit, deels ook tegen de grote nadruk op de impactscores. En de ergernis van veel wetenschappers over de almacht van wetenschappelijke uitgevers heeft inmiddels tot de 'Elsevier revolutie' geleid.

Klopt het verweer van Smeesters dus, of zijn hij en Stapel grote uitzonderingen? Even een heel korte inleiding op de wetenschappelijke methode.

Aanzien

In de natuurwetenschappen is het vaak mogelijk verschijnselen te zien of op te wekken. Wie theoretisch begrip heeft van verbrandingsprocessen, uitzetting en de eigenschappen van metalen kan een motor bouwen. Sterrenkundigen of biologen kunnen processen vaak observeren. Als een astronoom een astroïde ontdekt die op de aarde afsuist is dat een aantoonbare, zichtbare feitelijkheid. Sommigen menen dat breinwetenschappen kunnen zorgen dat psychologie ook zo'n soort natuurwetenschap wordt of tenminste een bètawetenschap. Psychologische processen zouden immers te 'zien' zijn. Uit scans zou bijvoorbeeld blijken dat het puberbrein niet kan plannen. Het verantwoordelijke deel, de frontaal kwab, zou zich wat minder actief tonen op scans, luidt het gewichtig. Pubers hebben permanent een soort frontaal syndroom. Daarom heeft breinonderzoek tegenwoordig aanzien. Veel psychologen betwijfelen echter of dergelijke rechtlijnige conclusies wel kloppen. Een eekhoorn die eikels verstopt heeft een frontaalkwabje zo groot als een speldenknop, maar kan toch uitstekend plannen. Heeft u wel eens het onder pubers populaire Minecraft gespeeld, een computerspel dat grotendeels om plannen draait?

Rob Martens

Reacties op dit artikel kunnen worden gemaild naar: rob.martens@ou.nl

Pubers kunnen uitstekend plannen, maar alleen als het hen uitkomt. Ze hebben gewoon geen zin om de saaie dingen te plannen die volwassenen belangrijk vinden, zoals huiswerk. Waarschijnlijk zal het nog lang lastig blijven sociaal wetenschappelijk onderzoek om te vormen tot een 'echte' bètawetenschap. Veel wetenschapsfilosofen (Huisman, 1996) wijzen er dan ook op dat in sociaal wetenschappelijk onderzoek rechtlijnige oorzakelijkheden zelden zijn vast te stellen. Sociale processen zijn zeer complex, zijn niet direct te zien of te isoleren en spelen zich fysiek af in het brein waar we welbeschouwd nog bijna niets van begrijpen.

Toeval?

Omdat processen niet direct observeerbaar zijn draait alles in de sociale wetenschap om de bijna heilige vraag: is iets toeval? Iedere statistische analyse die een sociaal wetenschapper hanteert, van een eenvoudige correlatieberekening tot multi-level analyse, draait om deze vraag. Vrijwel alle waarnemingen kennen een zogeheten normaalverdeling of een variant daarop. Alle statistische technieken die in de sociale wetenschap gebruikt worden toetsen hoe groot de kans is dat iets nog toeval is of dat er bijvoorbeeld 95 procent zekerheid is dat een gevonden verschijnsel geen toeval meer is. Nu wil het toeval dat wanneer ik uitga van een betrouwbaarheidsinterval van bijvoorbeeld 95 procent (p -waarde kleiner dan 0.05) ongeveer één op de twintig waarnemingen die ik doe 'significant' lijkt. Dat wil niet zeggen dat er een causaal mechanisme onder schuilt; het wordt volledig veroorzaakt door toeval. Wanneer ik twintig correlaties bereken is er dus een bepaalde kans dat één van die correlaties 'significant' is, terwijl in werkelijkheid de 'aange-toonde' relatie niet bestaat. Voor dit verschijnsel, veroorzaakt door herhaalde metingen, zal iedere goede onderzoeker corrigeren. Een voorbeeldje: het is onmogelijk om mensen te trainen om met een dobbelsteen altijd een zes te gooien. Wanneer ik honderd mensen toch een werptraining geef en honderd andere mensen niet, dan zullen er bij de honderd getrainde mensen altijd een paar zijn die in een trial van bijvoorbeeld tien werpbeurten opmerkelijk veel zessen gooien. Vrijwel zeker zal echter, indien ik de statistische technieken goed toepas, blijken dat in de controlegroep ook mensen voorkomen die veel zes-

sen gegooid hebben. En wanneer ik de gemiddelde scores van beide groepen vergelijk (rekening houdend met de omvang van de groepen, de spreiding en nog wat andere correcties) zal de uitkomst zijn dat de werptraining hoogstwaarschijnlijk niet tot significante verschillen geleid heeft. Iets ingewikkelder: als ik vijftien kleine groepjes van mensen train in dobbelsteenwerpen is er best een kans dat een van die groepen het gemiddeld beter doet dan de 'controlegroep'.

En nu komt het: hoogleraren als Smeesters zeggen dat het heel normaal is om alleen die resultaten te gebruiken die positief zijn, die jouw hypothese bevestigen. Ik reken dan alleen verder met een experimentele groep die toevallig veel zessen gooid en laat de rest buiten beschouwing. Of ik publiceer alleen het experiment met een positieve uitkomst en laat de rest onderin de la liggen: het 'file drawer fenomeen'. Zo kan ik een hype van dobbelsteenwerptrainingen aanrichten, ook al is het onderliggende bewijs niet waar. Ik heb dan niets zelf verzonnen, maar wel onwelgevallige resultaten weggelaten. De Volkskrant gaf onlangs een inkijkje in de 'datamassagesalon'. Bijvoorbeeld: meerdere experimenten doen en alleen rapporteren over die delen 'waar iets uitkomt' (alias de sleepnetmethode), redenen verzinnen om onwelgevallige cijfers onbetrouwbaar te verklaren, husselen met schalen bij vragenlijsten (bijvoorbeeld twee antwoordcategorieën samenvoegen), net zo lang statistische technieken op subsets van data loslaten tot er iets uitkomt, of populatiecorrecties selectief toepassen.

In mijn oratie (Martens, 2010) gaf ik nog andere voorbeelden, zoals harken (to hark: Hypothesizing After the Results are Known). Het happy ending van deze datamassage is de wetenschappelijke publicatie. Maar het aldus weglaten van onwelgevallige gegevens om zo naar een positieve uitkomst toe te werken is statistisch gezien net zo erg als het zelf verzinnen van scores: het maakt de resultaten volkomen onbetrouwbaar en als het regelmatig gebeurt ontstaat pseudowetenschap. Voor iedere wetenschap die zich tot in de haarvaten baseert op de (statistische) vraag of iets toeval is, is positieve bias of datamassage als betonrot in de fundering; het zal vroeg of laat het kennisgebouw doen instorten.

Permanente druk

Hebben wetenschappelijk tijdschriften inderdaad zo'n voorkeur voor positieve, duidelijke resultaten? Wie tot hier gevorderd is weet nu hoe huiveringwekkend de volgende citaten zijn uit publicatie-afwijzingsbrieven die promovendi van mij de afgelopen jaren van 'toptijdschriften' ontvingen, omdat hun resultaten negatief of ambigu waren: "In general, I was disappointed about the result pattern. Except for (...), no treatment effects were found. I wonder why.", "The overall effect is very small and they were not able to replicate it in the other studies. This does not give a lot of confidence to these findings.", "All in all the results are very inconsistent with theoretical and empirical considerations."

Smeesters heeft dus, vrees ik, gelijk als hij zegt dat dit veel gebeurt. En dat onderzoekers daardoor onder permanente druk staan 'selectief' te zijn, om het eufemistisch te omschrijven, omdat wetenschappelijke tijdschriften inderdaad een notoire voorkeur hebben voor positieve, duidelijke resultaten. Dat wil niet zeggen dat het daarom niet erg is. Sterker nog: het is rampzalig. Als we niet uitkijken met de welhaast Angelsaksische enorme (na)druk die Nederland (dat zeer hoog scoort op de internationale rankings van impactscores) legt op de publicatie van impactscores, dan leidt dit tot pseudowetenschap die niet alleen gewantwoord wordt door medisch hoogleraren, maar ook steeds meer ook door politiek, media en publiek. Dat lijkt me het begin van het einde, want waar anders dan op deugdelijk onderzoek moet je beleid baseren? Of medische behandelingen? Of wetenschappelijk onderwijs? If you think education is expensive, try ignorance. Zonder betrouwbare wetenschap gaan we terug naar de duistere Middeleeuwen.

Zelfreinigend vermogen

Ondanks dit zorgelijke beeld, waarin de affaires van Stapel en Smeesters het topje van de ijsberg zouden kunnen zijn, denk ik dat het probleem oplosbaar is. Mijn wereldje is dat van het onderwijsonderzoek. Onderwijsonderzoekers willen graag impact hebben op de onderwijspraktijk. Ze zijn ook niet bang om kritisch te zijn over onderzoek van collega's.

Dat zelfreinigende vermogen is er en waar mogelijk moet dat versterkt worden, in plaats van verzwakt door onderlinge competitie met nadruk op snelscorende onderzoekpublicaties met positieve resultaten.

Vrijwel alle onderzoekers die ik ken zijn oprecht nieuwsgierig, willen goed onderzoek doen en daar eerlijk over publiceren. En dat kan best, ook zonder fraude. De hierboven genoemde promovendi zijn inmiddels allen gepromoveerd en leven waarschijnlijk nog lang en gelukkig. Ook met echte data, met alle onvolkomenheden en tegenstrijdigheden, kun je dus wel degelijk promoveren.

Velen klagen over de grote impactscores-competitie in Nederland. Die competitie is ontstaan door jarenlange bezuinigingen op onderzoek en het alsmatig versterken van het competitie-model als Angelsaksisch geschoeid bedrijfsmatig 'afrekenprincipe', bijvoorbeeld door verschuiving van eerste geldstroom- naar tweede geldstroommiddelen, waardoor het produceren van impactlijstjes doel in plaats van middel is geworden. En daar betalen we nu de prijs voor.

Wie medisch onderzoek of onderwijsonderzoek belangrijk vindt, zal dus gewoon voldoende ruimte en autonomie – en daarmee ook onderzoeksmiddelen! – bij onderzoekers zelf moeten durven te leggen en zal kritische samenwerking moeten bevorderen in plaats van individualistische competitie op te eenzijdige criteria. Dan krijgt onderwijsonderzoek ook weer impact op de onderwijspraktijk, in plaats van gefixeerd te zijn op impactscores.

Het valt dus te hopen dat het nieuwe regieorgaan voor onderwijsonderzoek dat bij NWO wordt ingericht, zich deze les goed in de oren knoopt.

- Huisman, P. (1996). *Kennis gewogen*. Assen: van Gorcum.

- Martens, R. (2010). *Zin in onderzoek. Docentprofessionalisering*. Oratie. Heerlen: Open Universiteit. pdf via: www.ou.nl/look.

- Volkskrant (2012). Doping voor de data is overal. *Volkskrant*, 7 juli 2012.

