

## Eduard Looijenga

Mathematisch Instituut  
Universiteit Utrecht  
Postbus 80.010  
3508 TA Utrecht  
E.J.N.Looijenga@uu.nl

### Onderzoek Meten van kwaliteit

# Dr. Pangloss neemt de maat

De status van een onderzoeker wordt door veel aspecten bepaald. Nu steeds meer wetenschappelijk werk op internet verkrijgbaar is, is de administratie daarvan ook beter geregeld. Wie kijkt er nu niet graag naar lijstjes zoals de top tien van de citatie-index? Eduard Looijenga geeft zijn mening over het beoordelen van collega's.

Ondanks zijn tekortkomingen is *peer review* nog steeds de beste manier om wetenschappelijk werk te beoordelen. Maar adelplicht en wil deze methode stand houden, dan vereist dit van de kant van deskundigen een serieuze en gewetensvolle inspanning. Want als zij die niet kunnen opbrengen, wie dan wel?

#### Oordelen en beoordeeld worden

Iedereen heeft in zijn werkzame leven meer dan eens te maken met beoordelingen, aanvankelijk passief, later veelal ook actief. In het wetenschappelijk bedrijf gebeurt dat vaker dan elders, want daar doen zich nu eenmaal veel gelegenheden voor waarbij een keus gemaakt moet worden: allereerst wanneer het om een dienstverband gaat, zoals het bepalen wie een promotieplaats, een postdoc-schap, een vaste positie, een leerstoel krijgt, maar ook bij het beantwoorden van vragen als: wie nodigen we uit als spreker voor een conferentie, wie wordt een beurs, een mooie prijs, het lidmaatschap van een geleerd gezelschap aangeboden? Beoordelingswerk is in onze professie dan ook zo oud als het beroep zelf. En tot voor kort werd het praktisch uitsluitend toevertrouwd aan zijn beoefenaars. Met reden: goede wiskundigen kunnen in het algemeen goed oordelen (een uitspraak die minder tautologisch is dan hij lijkt) of vin-

den elkaar gemakkelijk als er vertrouwen is in elkaars deskundigheid en objectiviteit. En dan kan het ook buitengewoon efficiënt gaan. Een treffend voorbeeld is het Max Planck Instituut für Mathematik te Bonn: ieder jaar vragen circa 150 wiskundigen een bezoek aan (van enkele weken of langer), waarvan ongeveer een derde kan worden gehonoreerd. Het niveau van de aanvragers is hoog en dat maakt de selectie een complexe aangelegenheid. Toch slaagt de desbetreffende commissie (een gezelschap van zo'n tien wiskundigen) deze taak in een dagdeel te verrichten en wel zo dat bijna alle leden na afloop hun beslissing in ieder individueel geval kunnen rechtvaardigen.

De aanleiding voor dit stukje is tweeledig: enerzijds de neiging van de laatste tijd om in zulke zaken grotere bemoeienis toe te staan van niet-deskundigen en anderzijds de (mogelijk schuivende) kwaliteit van wetenschappelijke oordelen door vakgenoten. Deze twee zijn op een voor de hand liggende manier met elkaar te verbinden zonder dat er sprake hoeft te zijn van oorzaak en gevolg.

#### Een gebed verhoord

Zou het niet mooi zijn als er een eenvoudig algoritme bestond waarmee onmiddellijk de kwaliteit van een individuele onderzoeker is

te bepalen? En zou het niet nog mooier zijn als kwaliteit een ééndimensionaal attribuut is, waardoor de output eenvoudig in een getal is uit te drukken? Welnu, wat ooit een heilijker wens was van menig bestuurder is sinds een jaar of drie een verhoord gebed: er is zo'n algoritme en die heet de *h-index*, bedacht door de fysicus Jorge E. Hirsch (waar de *h* voor staat weet hij alleen) en gepresenteerd in een artikel uit 2005, *An index to quantify an individual's scientific research output*, PNAS 102 (46): 16569-16572. De definitie is simpel: een onderzoeker heeft *h-index* minstens *n* als deze persoon *n* artikelen geschreven heeft met ieder minstens *n* citaties. Voor iemand wiens werk voornamelijk door wiskundigen wordt geciteerd is dat inderdaad in een paar seconden te bepalen: ga naar *MathSciNet*, rubriek *author citations*, en type de naam in. Tel nu de lijst af totdat het rangnummer het aantal citaties dreigt te overschrijden en dan heb je de *h-index*. Hirsch legt in zijn artikel uit waarom zijn algoritme beter is dan andere conventies. Zo merkt hij terecht op dat het totale aantal citaties van weinig betekenis is als niet ook bekend is hoe die over de publicaties verdeeld zijn. Om maar een voorbeeld te noemen: als twee auteurs hetzelfde totale aantal citaties hebben voor resp. 50 en 100 publicaties, dan is het niet onredelijk te onderstellen dat het werk van de eerste beduidend meer gewicht heeft dan dat van de laatste. Deze telwijze beperkt verder de bijdrage van boeken (een boek kan deze index met hooguit 1 verhogen en als het maar enigszins een succes is

doet het dat ook). Hirsch schrijft dat voor een *tenure track*-positie aan een “major research university” (en daaronder vallen toch al onze universiteiten) een *h*-index van minstens 12 redelijk is en dat een *full professorship* een *h*-index van tenminste 18 vereist. Een oppervlakkige test leerde mij dat slechts één Nederlandse wiskundige aan het laatste criterium voldoet (L. Peletier met een *h*-index van 22) en dat de daarop volgende waarde (16) door maar drie landgenoten (allen emeriti) bereikt wordt. Dit toont al meteen aan dat deze algoritme voor een publicatiecultuur als de onze, waarbij artikelen hun citaties vaak pas over vele jaren verzamelen, heel onrechtvaardig en misleidend uitpakt voor jonge onderzoekers (dit wordt maar in bescheiden mate gecompenseerd door een ander manco van deze methode: MathSciNet verstrekt alleen de citaties verschenen vanaf 2000 en in sommige gevallen vanaf 1997; bovendien zijn het alleen de citaties voorkomend in tijdschriften). Nadere beschouwing maakt ook duidelijk dat het genoemde kwantitatieve verschil tussen de natuurkunde en de wiskunde een pendant heeft binnen de wiskunde zelf: in sommige deeldisciplines zijn de citatiescores gemiddeld lager zonder dat er reden is om daarin de afspiegeling te zien van een overeenkomstig verschil in kwaliteit. Het blijkt wel een (in de tijd) robuuste maat te zijn, want de periode tussen twee sprongen duurt vaak lang. Ook moet ik bekennen dat als we ons beperken tot senior-onderzoekers die werkzaam

zijn in min of meer hetzelfde vakgebied, het resultaat van deze methode aardig in overeenstemming kan zijn met een evaluatie die op conventionele manier tot stand is gekomen. Maar zelfs als dat betekent dat dit voor vier van de vijf keer het geval is, dan nog schiet deze methode naar mijn mening tekort. Want bedenk wel dat hier gaat om de beoordeling van een individu en niet om die van een groep waarvoor middelingsprocedures een zuiverend effect plegen te hebben. (Een opmerkelijk voorbeeld om in het oog te houden is de Fields laureaat Laurent Lafforgue, wiens *h*-index 5 is.) Mijn conclusie luidt dan ook dat voor het beoordelen van wiskundigen deze *h*-index onbruikbaar is.

### Een koninklijke weg

Hoe dan wel? Mijn antwoord zal onder vakgenoten (waartoe de lezer hoogstwaarschijnlijk behoort) weinig opzien baren: ik vind dat alleen de beroepsgroep zelf dat op een verantwoorde manier kan doen. Maar het adjectief in het laatste zinsdeel heeft dan wel gevolgen, zeker als we onze positie tegenover boekhoudende buitenstaanders wensen te versterken. Het vereist bijvoorbeeld dat de beoordelaar zich enigszins verdiept in het werk van de beoordeelde. Daarbij kan het MathSciNet van dienst zijn op een manier die wat inhoudelijker is dan wat ik zojuist beschreef, namelijk door eenvoudig de reviews van alle artikelen van de persoon in kwestie te lezen. Dit kost niet veel tijd en is meestal zeer informatief.

(MathSciNet verstrekt trouwens ook andere interessante gegevens, zoals het totale aantal *verschillende* citeerders, wat ons iets vertelt over de verspreiding van het werk.) En zelfs als het werkterrein voor de referent relatief onbekend is, dan nog loont het vaak voor de laatste om de artikelen er zelf eens bij te nemen, al was het maar om in een enkel geval de inleiding te lezen. Mijn ervaring is dat kwaliteit op deze manier bijna altijd te herkennen valt. De andere gebruikelijke indicatoren kunnen en horen natuurlijk ook in aanmerking genomen worden: de reputatie van de tijdschriften waarin is gepubliceerd, bijzondere voordrachten op uitnodiging, ontvangen internationale prijzen, een artikel dat een *Featured Review* ten deel is gevallen en dergelijke. Dit alles lijkt een open deur, en voor de beoordelende gezelschappen waar ik deel van uit heb meegemaakt was het dat meestal ook. Maar als ik mij bepaal tot mijn ervaringen in eigen land, dan zijn die toch gemengd. Laat ik het maar zo formuleren: zich voor een oordeel baseren op de Google Scholar score, of in het feit dat volledige objectiviteit nu eenmaal een illusie is reden zien om een serieuze inspanning achterwege te laten, dat is niet wat je noemt de koninklijke weg. Zulk een gemakzuchtige benadering is niet te rechtvaardigen, zeker niet als het om personen gaat en al helemaal niet als we de stelling willen verdedigen dat het systeem van peer review ondanks zijn tekortkomingen superieur is aan het oordeel van leken. ←



Dr. Pangloss uit *Candide* van Voltaire, hier als decorstuk bij een uitvoering van Bernsteins musical.  
Illustratie: Eloi Meylan (Gymnase de Chamblandes à Pully)