

## Sturen met cijfers Een korte reflectie

Hans de Bruijn, hoogleraar bestuurskunde aan de TU Delft  
31 augustus 2022

Dit essay is geschreven op verzoek van de Adviesraad Migratie (ACVZ) en bevat een aantal reflecties op het gebruik van ‘beleidsmatige richtgetallen in het migratiedomein’.<sup>1</sup> Het verzoek van de ACVZ was om op hierop te reflecteren vanuit mijn deskundigheid over kwantitatieve prestatie-sturing.

In het rapport ‘Samenleven in verscheidenheid’ adviseert de WRR om het gebruik van een ‘beleidsmatig richtgetal’ in het migratiebeleid te overwegen.<sup>2</sup> Zo’n richtgetal geeft aan wat de jaarlijkse omvang van de asielmigratie kan zijn, gegeven het incorporatievermogen (absorberend vermogen) van de samenleving. Het richtgetal geeft een politieke ambitie weer – het is geen absolute grens aan het aantal asielmigranten. Een dergelijk richtgetal heeft volgens de WRR twee mogelijke voordelen: het leidt tot een meer expliciete politieke discussie over gewenste aantallen en het helpt bij het treffen van voorbereidingen voor de komst van migranten. Een beleidsmatig richtgetal hoeft zich uiteraard niet te beperken tot asielmigratie, maar kan ook betrekking hebben op andere vormen van migratie.

Een richtgetal is een vorm van sturen met cijfers. Wat weten we daarover, dat relevant kan zijn voor besluitvorming over het wel of niet gebruiken van richtgetallen?

### I Terminologie

Uiteraard vraagt een beleidsterrein als migratie om kwantitatieve inzichten. Het gaat nog een stap verder wanneer cijfers actief worden gebruikt bij sturing en verantwoording. De basisterminologie is eenvoudig:

- Ken- en stuurgetallen. Er zijn cijfers die een feitelijke situatie weergeven, dat noemen we kengetallen. Vervolgens zijn er cijfers die die de gewenste situatie aangeven, dat noemen we stuurgetallen. Ken- en stuurgetallen tezamen noemen we een prestatie-meetsysteem, het gebruik ervan heet prestatiesturing.
- In plaats van over stuurgetallen, kan ook worden gesproken over streefgetallen (c.q. richtgetallen). Die zijn zachter dan een stuurgetal, bijvoorbeeld omdat ze minder goed beïnvloedbaar zijn voor bestuurders. Overigens is het onderscheid tussen

---

<sup>1</sup> Dit essay is gebaseerd op Hans de Bruijn (2007), *Managing Performance in the Public Sector*, London: Routledge; Kwink Groep (2010), *Handboek Prestatiemeting*, Den Haag: Kwink; Wouter van Dooren et al. (2015), *Performance Management in the Public Sector*, London: Routledge; E. Gerrish (2016), *The impact of performance management on performance in public organizations: a meta-analysis*, in *Public Administration Review*, 76(1), 48-66.

<sup>2</sup> WRR. (2020). *Samenleven in verscheidenheid. Beleid voor de migratiesamenleving*. Den Haag: WRR. p. 247.

harde stuurgetallen en zachte(re) streefgetallen er een voor bestuurlijke fijnproevers – voor veel buitenstaanders zijn het gewoon kwantitatieve doelen, die je wel of niet hebt behaald.

- Output en outcome. Ken- en stuurgetallen kunnen zijn gericht op output of outcome. Bij output gaat het om de ‘producten’ of diensten van de betreffende organisatie. Als een overheidsorganisatie milieuvergunningen afgeeft, is het aantal verleende vergunningen de output van deze organisatie. Outcome is het uiteindelijk beoogde effect: een schoner milieu. Idealiter liggen output en outcome in elkaars verlengde, maar dat is niet per definitie zo. Je kunt een prima output hebben en tegelijk een slechte outcome, of andersom.

In de volgende paragrafen maak ik geen onderscheid tussen stuur- en streefgetal, omdat de beschreven mechanismen voor beide soorten getallen gelden.

## **II Relevantie, meetbaarheid en toerekenbaarheid als voorwaarden voor sturen op cijfers**

Het begint allemaal met een ambachtelijke vraag. Het gebruik van ken- en stuurgetallen betekent dat er iets wordt geteld, bijvoorbeeld de instroom van migranten, gedifferentieerd naar typen migranten. Als er ook gestuurd wordt op cijfers, kan zo’n systeem van ken- en stuurgetallen alleen functioneren als aan drie voorwaarden is voldaan:

1. meetbaarheid
2. relevantie
3. toerekenbaarheid

Een rechter wijst vonnissen. Zijn die meetbaar en relevant? Het antwoord op beide vragen: ja en nee. En dat is het antwoord in vrijwel alle publieke sectoren, ongetwijfeld ook bij migratie.

Ja: het aantal vonnissen is meetbaar, je kunt ze tellen. Nee: er is altijd de vraag wat nu precies een vonnis is? Er zijn vonnissen en tussenvonnissen, allerlei verschillende soorten zaken en dus verschillende soorten vonnissen. Er is vrijwel altijd gedoe met definities: definities overlappen, doen nooit volledig recht aan de variëteit in de werkelijkheid, er is altijd de vraag tot welke categorie je een concreet geval moet rekenen, et cetera.

Zijn vonnissen relevant voor de beoordeling van de prestaties van een rechter? Ja en nee: ze zeggen iets over wat een rechter doet, maar de prestatie van een rechter gaat over veel meer dan alleen het vonnis (kwaliteit, bejegening, snelheid).

Zijn vonnissen toerekenbaar aan de rechter? Deels. Het aantal vonnissen dat een rechter wijst, is ook afhankelijk van de houding van de procederende partijen of van de vraag of het Openbaar Ministerie (OM) op tijd volledige dossiers aanlevert.

Het voorbeeld van de vonnissen maakt duidelijk dat er eigenlijk altijd wel issues zijn met meetbaarheid, relevantie en toerekenbaarheid. Dat is niet per definitie een probleem. Je

kunt je voorstellen dat bestuurders en professionals ook wel begrijpen dat je, bijvoorbeeld, de perfecte definitie niet hoeft te hebben, om bestuurlijk toch goed gebruik te kunnen maken van die definitie, inclusief bijpassende ken- en stuurgetallen. Zolang definities en getallen niet worden verabsoluteerd, is dat geen probleem.

Wanneer cijfers ook een belangrijke rol gaan spelen in het publieke debat, kan dit wel problematisch zijn – of het nu om harde stuurgetallen gaat of zachte streefgetallen. Een streefgetal dat wordt overschreden, kan de indruk wekken dat een bewindspersoon *out of control* is, ook al was het geen stuurgetal. De tolerantie voor ambiguïteit is in de bestuurlijke arena vaak wat groter dan in de politieke en publieke arena.

### III Ken- en stuurgetallen kunnen perverteren

Stel, er is een systeem van ken- en stuurgetallen. Wat zijn de effecten hiervan? Er wordt in de literatuur een onderscheid gemaakt tussen positieve en perverse effecten. De essentie van deze effecten laat zich als volgt samenvatten. Allereerst de positieve effecten:

1. Er ontstaat meer transparantie. Ken- en stuurgetallen leveren transparantie op: beleidsdoelstellingen worden er scherper door, de uitvoering laat zich beter controleren. Er is cijfermatig inzicht in de feitelijke en gewenste situatie, er is een norm waarop bestuurders kunnen worden aangesproken, er zijn vergelijkingen in de tijd mogelijk, cijfers kunnen worden gebruikt om al te gemakkelijke redeneringen door te prikken, et cetera. Er wordt ook wel gesproken over de elegantie van prestatiemeting: een ingewikkelde prestatie wordt samengevat in een of enkele cijfers.
2. Prikkel tot presteren. *What gets measured gets done*, zo luidt het gezegde. Wanneer er stuurgetallen zijn afgegeven, zal een overheidsorganisatie zich in de regel maximaal inspannen om die stuurgetallen te realiseren. Stuurgetallen leiden zo tot een focus op bepaalde prestaties, en die focus kan bevorderlijk zijn voor het realiseren van de prestatie. Nauw hiermee verbonden: stuurgetallen kunnen een prikkel zijn voor meer samenhangend beleid. Wie een hard cijfer afgeeft over het aantal vonnissen in strafzaken voor rechtbanken, wordt als vanzelf gedwongen om ook naar het OM te kijken: draagt dat voldoende zaken op tijd aan?
3. Prikkel voor leren en innoveren. Ken- en stuurgetallen nodigen uit tot vergelijken: in de tijd, met andere landen of met andere beleidsterreinen. Vergelijkingen zijn bevorderlijk voor leren. Als uit een benchmark blijkt dat vergelijkbare organisaties heel verschillende resultaten boeken, is het interessant om te bezien wat er van de beter presterende organisaties kan worden geleerd.

Deze voordelen klinken ook door in het WRR-advies. Maar daarnaast zijn er de perverse effecten, en die zijn min of meer het spiegelbeeld van de positieve effecten:

1. Geen transparantie, maar schijntransparantie. Ken- en stuurgetallen lijken transparantie te brengen, maar dat is niet per definitie zo en daar kunnen tal van redenen voor zijn. De gekozen indicatoren, of de definities van die indicatoren, zijn aanvechtbaar. Of het proces van data-vergaring is discutabel: er zijn vervuilde data,

concrete situaties zijn niet eenduidig (als asielzoekers na inwilliging van hun asielverzoek direct hun familie laten overkomen, zijn dit dan ‘asielmigranten’ of ‘gezinsmigranten’?). De optelsom van dit soort factoren kan ertoe leiden dat een cijfer uiteindelijk niet zo veel zegt. Een bekend voorbeeld is dat van de rankings van ziekenhuizen, op grond van prestatie-meet-systemen. Je kunt ziekenhuizen rangschikken op grond van hun professionele prestaties. Daar komen echter allerlei definitie- en datavraagstukken in mee. Het gevolg daarvan: er zijn ziekenhuizen die in de ene ranking top zijn en in de andere onderaan bungelen, en andersom.

2. Geen prikkel voor presteren, maar *gaming the numbers* en onprofessioneel gedrag. Gaming the numbers is het verschijnsel dat de cijfers op papier belangrijker worden dan de werkelijkheid daarachter, en dat voor alles wordt geprobeerd die cijfers op papier er goed uit te laten zien. Weer een voorbeeld: als de wachttijden op de spoedeisende hulp omlaag moeten, kun je dat eenvoudig realiseren. Een patiënt komt binnen, er vindt snel een intake plaats, en daarmee in de behandeling begonnen en de wachttijd gereduceerd. Als de patiënt vervolgens na de intake nog moet wachten, telt dat niet mee als wachttijd. Dit is een wel heel opzichtig voorbeeld, maar gaming the numbers verloopt vaak subtieler en is van bovenaf niet altijd zichtbaar. De-professionalisering is het verschijnsel dat cijfers belangrijker worden dan professionele standaarden. Als een fraudepercentage omlaag moet, kun je zo streng gaan controleren, dat je de billijkheid en redelijkheid uit het oog verliest. Maar op papier doet je het geweldig omdat het fraudepercentage omlaag gaat. Er is daarmee geen prikkel voor samenhangend beleid, maar juist het tegenovergestelde: een fixatie op een enkel stuurgetal, dat koste wat het kost gerealiseerd moet worden.
3. Geen innovatie, maar bureaucratie en standaardisatie. Prestatiemeting kan leiden tot innovatie, maar net zo goed tot meer bureaucratie. Onderschat de bureaucratiekosten van een prestatiemeting niet. Als de bovenstaande perverse effecten zich manifesteren, is de reactie vaak meer controles door de partij die de cijfers ontvangt, meer verfijningen van definities, uitbreiding van de set ken- en stuurgetallen om zo een beter beeld te verkrijgen. Uiteindelijk kan dat leiden tot een systeem dat zijn elegantie verliest, zie de klachten over prestatie-meetsystemen in de zorg en het onderwijs.

#### **IV Het risico van *high impact* prestatiebesturing**

Als ken- en stuurgetallen voor- en nadelen hebben, is de cruciale vraag: Onder welke condities gaan die nadelen zich manifesteren? Wanneer overvleugelen de perverse effecten de positieve effecten?

Belangrijk is hier de impact van een systeem van ken- en stuurgetallen. De vuistregel: als de impact hoog is, neemt de kans op pervertering toe. Een eenvoudig gedachte-experiment: stel dat we het salaris van een docent deels afhankelijk maken van de scores in onderwijs-evaluaties, vanuit het idee dat dit tot beter onderwijs leidt. De impact van een prestatie-meetsysteem is dan zo hoog, dat er een prikkel voor pervertering ontstaat: omdat het systeem zo belangrijk wordt gemaakt, eisen docenten dat er preciezer moet worden

gemeten en wordt het systeem eindeloos verfijnd (bureaucratie) of docenten laten bepaalde studenten niet meedoen aan de evaluaties om zo resultaten te sturen (gaming the numbers). Dat gaming the numbers behoeft overigens niet altijd moreel fout te zijn. Een systeem van prestatiemeting geeft altijd een beperkte afspiegeling van de werkelijkheid. Als zo'n systeem te belangrijk wordt gemaakt (te veel impact krijgt), kan een professional (hier: de docent) het volkomen legitiem vinden om met het systeem te spelen. Einde gedachte-experiment.

Dus hier ontstaat een belangrijke paradox. Cijfers geven een beperkte afspiegeling van de werkelijkheid. Hoe meer je met de cijfers wilt sturen, hoe groter de kans op pervertering; hoe meer pervertering, hoe minder je stuurt.

## **V Hoge impact is ook: veel politiek-bestuurlijke attentie**

Bij 'hoge impact' gaat het niet alleen om financiële consequenties – beloningen of sancties – maar ook om bestuurlijke of politieke attentie. Wanneer een stuurgetal veel bestuurlijke of politieke aandacht krijgt (de leiding van een organisatie is er alert op, het parlement heeft er veel interesse voor), dan krijgt een cijfer daardoor een hoge impact. Immers, indien streefcijfers niet worden gerealiseerd, kan dat tot allerlei negatieve effecten leiden: meer bemoeienis met operationele processen, reputatieschade, verlies van legitimiteit, et cetera.

Stel, een bepaald stuurgetal staat in het brandpunt van de politieke belangstelling, er is dus sprake van een hoge politieke attentie. Het getal geeft bijvoorbeeld een bovengrens aan, en de ambitie is om die niet te overstijgen.

Bedenk: omdat bij prestatiemeting de complexe werkelijkheid wordt gereduceerd tot een eenvoudig cijfer, kun je altijd kritiek hebben op de manier waarop dat cijfer tot stand komt. Bij veel politieke attentie is dit de worst case:

- Degenen die het cijfer controleren, kunnen altijd beweren dat het cijfer een verkeerde afspiegeling van de werkelijkheid is, dus dat zij niet de juiste informatie hebben gekregen.
- Degenen die het cijfer produceren, hebben er belang bij dat de bovengrens niet wordt overschreden. Dat kan een sterke prikkel opleveren voor gaming the numbers, onprofessioneel gedrag en andere perverse effecten. Zoals gezegd, gaming the numbers is niet per definitie moreel fout, vanuit professioneel perspectief. Hoe meer politieke belangstelling, hoe sterker de prikkel hiertoe.

## **VI De noodzaak van gematigd gebruik en de spanning met een gepolitiseerde context**

Hoe kun je omgaan met de beperkingen die ken- en stuurgetallen met zich meebrengen? Het antwoord op die vraag is eenvoudig: het gaat niet om de cijfers, maar om de werkelijkheid achter de cijfers. Zie de getallen dan ook niet als een sluitstuk van een beoordelingsproces, maar als het begin ervan. Gebruik cijfers om een gesprek te faciliteren. Een

enkel cijfer zegt in principe niets. Het voorbeeld hier: mortaliteitscijfers van een afdeling in een academisch ziekenhuis. Het feit dat dat cijfer hoog is in vergelijking met afdelingen in andere academische ziekenhuizen, zegt nog niets over de kwaliteit van de verleende zorg. Het is wel opmerkelijk en dus het begin van een gesprek. Misschien blijkt dat de betreffende afdeling top-artsen in dienst heeft en daardoor veel patiënten trekt die elders zijn uitbehandeld. Maar het kan ook blijken dat artsen slecht samenwerken, elkaar niet corrigeren – en dat daardoor het mortaliteitscijfer hoog is.

Een systeem van prestatiesturing werkt dus vooral als het enigszins gematigd wordt toegepast: beperk de impact, zie het als een tool voor een gesprek, het cijfer is niet het sluitstuk van een beoordelingsproces, maar het begin ervan.

Wie het gebruik van prestatiesturing overweegt, moet zich vervolgens altijd afvragen of de politiek-bestuurlijke context zich daarvoor leent. Naarmate een onderwerp sterker is gepolitiseerd, zijn de kansen op dit gematigd gebruik kleiner. Wie vanuit een sterke politieke overtuiging met een cijfer wordt geconfronteerd dat ingaat tegen de eigen politieke opvattingen, is niet altijd in voor een genuanceerde discussie. Weer een gedachte-experiment. Er is een politicus die van uit een diepe morele overtuiging vecht tegen armoede. Er worden door de regering vijf streefcijfers afgegeven inzake armoede en vier daarvan blijken na twee jaar niet te zijn gehaald. Een minister kan zich verdedigen met het argument dat de werkelijkheid genuanceerder is. Het is de vraag of onze politicus in die nuancering is geïnteresseerd. Cijfers kunnen communicatief heel krachtig zijn – ‘het kabinet scoort vier dikke onvoldoendes’ – en dan is de verleiding groot om ze te gebruiken en zo meer aandacht voor armoedebestrijding te vragen. De communicatieve kracht van cijfers maakt ze een goed pressiemiddel richting overheid en samenleving. Einde gedachte-experiment. De moraal van dit verhaal: hoe meer gepolitiseerd een vraagstuk, hoe minder kans op het noodzakelijke gematigde gebruik van ken- en stuurgetallen.

## **VIII Hoe te gebruiken in gepolitiseerde context? Voorbeelden**

Ten slotte, stel dat een context is gepolitiseerd en dat toch het gebruik van stuurgetallen gewenst is. Dan gaat het er om strategieën te ontwikkelen voor gebruik in een dergelijke context. Een paar voorbeelden:

- Voorkom de fixatie op een of enkele cijfers. Hoe? Bijvoorbeeld door met bandbreedtes te werken, en niet een enkel getal. Of door naast jaarlijkse cijfers, ook voortschrijdende gemiddelden over de jaren heen te laten zien. Of door naast absolute cijfers, ook relatieve cijfers te laten zien – een getal en een percentage dus. Een belangrijk principe is hier ook ‘variëteit’. Gebruik een set met een variëteit aan ken- en stuurgetallen, die tezamen zo veel mogelijk de complexiteit van de onderliggende prestatie weerspiegelen. Als bijvoorbeeld sprake is van een dienst, die zich op twee verschillende manieren laat definiëren, kan het zinvol zijn om beide definities te hanteren. Dat levert een meer genuanceerd beeld van de prestatie op en beperkt de mogelijkheden om één cijfer heel erg uit te vergroten.
- Verbind met andere getallen. Een streefgetal kan als snel binair worden - het is overschreden of onderschreden, de realisatie is goed of fout. Verbreed de set van

ken- en stuurgetallen, waardoor cijfers over immigratie ook verbonden kunnen worden met cijfers over bijvoorbeeld arbeidsmarkt, onderwijs, huisvesting. Verbreed de set ook door aandacht te vragen voor outcome, de uiteindelijke bijdrage van migranten aan de Nederlandse samenleving. Dan ontstaat een breder, meer integraal beeld en dat kan bevorderlijk zijn voor de kwaliteit van het politieke debat.

- Investeer bij de presentatie van de prestaties (dus wanneer duidelijk wordt hoe de werkelijke prestatie zich verhoudt tot het stuurgetal) in een scherp kwalitatief verhaal over de interpretatie van de cijfers. Voorkom een situatie van scherpe cijfers en een wollig verhaal, want dan trekken de scherpe cijfers alle aandacht.
- Zorg dat een streefgetal breed wordt gedragen en een *negotiated performance indicator* is. Immigratie is een collectief vraagstuk, waarbij alle overheidslagen zijn betrokken. Wanneer een streefgetal wordt vastgesteld is het van belang dat ook deze andere overheden daar een stem in hebben. Zij spelen immers een cruciale rol bij opvang en integratie. Wellicht is de betrokkenheid van anderen (huisvesting, onderwijs) ook mogelijk. Hoe breder een streefgetal wordt gedragen, hoe beter. Mijn indruk is dat dat de essentie van de Oostenrijkse drempel is: een bovengrens die is afgesproken tussen verschillende overheden en onderdeel van een package deal. Een collectieve afspraak dus: als de drempel onverhoopt wordt overschreden, is dat dan ook een collectief probleem. Het proces van totstandkoming van een getal is minstens zo belangrijk als het getal zelf.
- Zoals gezegd, bij bepaalde vormen van migratie is een stuurgetal inzake instroom voor een belangrijk deel niet toerekenbaar aan een bewindspersoon. Door externe ontwikkelingen kan er meer instroom zijn dan gepland. Wanneer sprake is van een negotiated indicator, kunnen er wellicht ook gezamenlijke scenario's worden afgesproken voor zo'n situatie van meer instroom. Wordt het streefgetal overschreden, dan treden die scenario's in werking. Dan is er een streefgetal, maar ook de erkenning van externe invloeden, plus de anticipatie daarop.
- Probeer wat leer- en experimenteerruimte te creëren om na te gaan of stuurgetallen werken. Het zal duidelijk zijn: het gebruik van ken- en stuurgetallen roept allerlei vragen op. Er is de vraag van toerekenbaarheid. De vraag of de positieve of juist de perverse effecten zich zullen manifesteren. Het kan ook onduidelijk zijn wat de functie van de stuurgetallen wordt: wordt er hard op afgerekend of worden de getallen meer genuanceerd gebruikt? Dit soort onzekerheden rond het gebruik van stuurgetallen, betekent dat er ruimte moet zijn om te leren. Door eerst wat te experimenteren met het gebruik van ken- en stuurgetallen, zonder ze onderdeel te laten zijn van de formele beleidscyclus, kan worden geleerd of ze doen wat ze moeten doen. Daarbij is het uiteraard wel weer de vraag of een bewindspersoon in een sterk gepolitiseerde context voldoende ruimte wordt gegund voor leren en experimenteren.