

Column

One size fits all?

Waarom willen we zo graag standaardiseren?

Melanie Ehren*

Standaardisering, het vergelijkbaar maken van een complexe set aan kwaliteiten op eenzelfde meetlat, vormt een belangrijk uitgangspunt in het werk van de rijksinspecties. Door prestaties van organisaties op een gestandaardiseerde manier te meten verwachten we objectief te kunnen vaststellen wie of wat er beter is en of er aan een acceptabele basisnorm wordt voldaan. Dergelijke beoordelingen moeten inzicht geven in vragen zoals: Welke school biedt de beste onderwijskwaliteit? Wordt het minimumaantal lessen aangeboden? In welk ziekenhuis is de geleverde zorg op orde en zijn de wachttijden en het aantal heropnames beperkt?¹

Standaardiseren van vaak complexe diensten is zo gewoonged geworden dat we ons nauwelijks afvragen of deze vorm van beoordelen wel antwoord geeft op deze vragen, en of het bijdraagt aan verbetering van de publieke diensten.

Waarom hechten we zo'n belang aan het standaardiseren en vergelijkbaar maken van prestaties? En hoe informatief en accuraat zijn dergelijke ranglijsten en prestatierapportages?

Het presteren van leerlingen op een basisschool en de kwaliteit van het geleverde onderwijs, de kwaliteit van een zorginstelling: het zijn alle voorbeelden van zeer complexe en veelomvattende kwaliteiten die bovendien in de loop der tijd veranderen. Onze verwachtingen van goed onderwijs veranderen bijvoorbeeld in reactie op ontwikkelingen in de samenleving: burgerschap is de laatste jaren veel nadrukkelijker op de agenda gekomen,

naast de aandacht voor taal en rekenen. Een ziekenhuis bestaat uit verpleegkundigen, artsen en ondersteunend personeel, waarbij sommige teams excellente specialistische kankerzorg bieden, maar wellicht weinig oog hebben voor de emotionele zorg die nodig is voor het algehele welbevinden van de patiënt. Standaarden, uitgedrukt in meetbare en vergelijkbare cijfers, zijn vaak moeilijk in staat om die nuance en diversiteit goed te vangen. Waarom maken ze dan zo'n prominent deel uit van hoe we toezicht houden op publieke dienstverleners en hoe we hun verantwoording organiseren?

Waarom: beperkte cognitieve capaciteit en gepercipieerde objectiviteit

Het antwoord ligt in onze beperkte vermogens om grote hoeveelheden aan kwalitatieve informatie goed te verwerken en vergelijken. Zelfs als we de kennis hebben over wat 'kwaliteit' inhoudt, dan hebben we niet de cognitieve capaciteit, tijd of interesse om uitgebreide rapportages te lezen en te interpreteren. Liever een lijstje met scholen of ziekenhuizen vergelijken op een aantal scores dan uitgebreide rapportages te lezen over wat de unieke kwaliteiten zijn van het onderwijs in verschillende vakken en aan verschillende doelgroepen leerlingen in een school, of hoe instellingen presteren in het aanbieden van geïntegreerde zorg aan chronische patiënten.

Het is echter niet alleen deze *bounded rationality* die onze voorkeur voor kwantificeerbare standaarden bepaalt. Onderzoek laat ook zien dat cijfers en getallen meer overtuigingskracht hebben dan kwalitatieve beschrijvingen; ze worden vaak als objectiever, betrouwbaarder en meer valide gezien. In de *sociology of quanti-*

* Prof. dr. M. Ehren is hoogleraar in Educational Governance aan de Vrije Universiteit Amsterdam en directeur van Research Institute LEARN!

1 www.ziekenhuischeck.nl/vergelijken/.

fication² wordt dit fenomeen uitgebreid onderzocht. Uit dat onderzoek blijkt dat de mogelijkheid om vergelijkingen te maken tussen zeer verschillende eenheden – denk aan een klein ziekenhuis met specialistische zorg en een groot universitair medisch centrum – appelleert aan een gevoel van efficiëntie en effectiviteit; getallen impliceren rationaliteit in hoe iets wordt beoordeeld en een vorm van neutraliteit in hoe we wegen en evalueren.

De informatiewaarde van standaarden

Niets is natuurlijk minder waar: door iets of iemand te beoordelen op een vergelijkbare set aan standaarden reduceren we weliswaar een grote hoeveelheid aan relevante kwaliteiten en informatie tot een overzichtelijke set aan indicatoren, maar aan die reductie liggen ook subjectieve keuzes en voorkeuren ten grondslag, veelal gebaseerd op wat we makkelijk kunnen meten in plaats van wat we echt belangrijk vinden. Een voorbeeld uit de onderwijscontext vinden we in de gestandaardiseerde toetsen die worden gebruikt om leerprestaties van leerlingen vast te stellen en daarmee (in een geaggregeerde score) de kwaliteit van hun school. Dergelijke toetsen meten geen mondelinge spreekvaardigheid omdat die niet goed op grote schaal is te meten en is uit te drukken in een getal. Dat betekent dat deze vaardigheid buiten de inspectiebeoordeling van onderwijskwaliteit blijft terwijl we wel verwachten dat leerlingen zich goed leren uitdrukken. Standaardiseren betekent keuzes maken over hoe en wat we meten en prestaties die makkelijk zijn te kwantificeren hebben daarbij veelal prioriteit; het is ‘laaghangend fruit’.

Dat er keuzes worden gemaakt impliceert dat toezichtstandaarden normatief zijn en bovendien een normerende werking hebben: ze bepalen hoe we naar individuen en organisaties kijken en wat we belangrijk vinden. Met het toetsen van taal en rekenen hebben deze twee vakken aan belang gewonnen, en wie zal nog betwisten dat een goede school geen kwaliteit levert? Toch blijft een kritische blik op de wijze waarop we standaardiseren en vergelijken van belang. Niet alleen om recht te doen aan de enorme diversiteit aan kwaliteiten van mensen en organisaties, maar ook omdat onze eigen wensen en behoeften zeer uiteenlopen en bovendien veranderen in de loop der tijd. Waar de ene ouder een bepaalde school de beste vindt voor zijn/haar kind, maakt een andere ouder een andere keus. Oftewel: *one size does not fit all*.

Een opdracht voor het toezicht

Wat betekent dat nu voor het toezicht? Oog voor het ‘hoe’ en ‘waarom’: toezichthouders die uitspraken willen doen over de echte wereld waarin variatie, padafhankelijkheid, menselijke zwakten, groepsdynamiek en nog vele andere grilligheden een rol spelen, zullen naar aanleiding van de cijfers een realiteitscheck moeten doen. En die kan heel eenvoudig beginnen door de vraag te stellen naar het hoe en waarom achter de cijfers. De

school die goed scoort in de gemeten en gestandaardiseerde statistieken kan op ongemeten kwaliteiten (bijvoorbeeld sfeer en omgangsvormen) voor sommige ouders en leerlingen heel onaangenaam zijn. Standaarden zullen daarom altijd met een zekere nuance en bescheidenheid moeten worden geïnterpreteerd en dat vergt ook tijd voor gesprek en observatie in plaats van de volledige automatisering van beoordelingen, bijvoorbeeld in risicomodellen. De cijfers en scores op gestandaardiseerde meetlatten kunnen een goed begin zijn van een gesprek over het ‘waarom’ van die scores of het waarom van bepaalde verschillen, maar geven nooit het volledige beeld van de werkelijkheid.

Dan komen we bij het tweede punt dat van belang is: toezichthouders hebben discretionaire handelingsruimte nodig om dat gesprek te voeren en standaarden te interpreteren. Toezichthouders zonder handelingsruimte zijn per definitie veroordeeld tot een reductionistische blik en doen daarmee per definitie de werkelijkheid geweld aan. Handelingsruimte is noodzakelijk om na de realiteitscheck op de cijfers (de ‘waaromvraag’) in de informatiefase, ook een specifieke context, achtergrond of verleden te kunnen laten meewegen in de beoordelings- en interventiefase van het toezicht. Om naar de geest van de wet te kunnen handelen en niet alleen naar de letter, om als toezichthouder niet bij te dragen aan het probleem maar aan de oplossing.

Dat laat onverlet dat er in de maatschappij wel behoefte is aan toegankelijke en vergelijkbare informatie over kwaliteiten van dienstverleners. Maar aan die vraag zouden we moeten voldoen door instrumenten te ontwikkelen waarmee meer gedifferentieerde en kwalitatieve informatie goed wordt ontsloten. Nieuwe technologie, zoals ChatGPT en *large language*-modellen, bieden goede perspectieven om toezichtinformatie met meer nuance te ontsluiten en antwoord te geven op een variëteit aan vragen, passend bij de behoeften en opvattingen over kwaliteit van burgers en andere informatiegebruikers. Recht doen aan de complexiteit en diversiteit van onze samenleving betekent een andere vorm van wegen en meten: geen liniaal, maar een intelligent *dashboard*.

2 R. Diaz-Bone & E. Didier, ‘The sociology of quantification. Perspectives on an emerging field in the social sciences’, *Historical Social Research* 2016, nr. 2, p. 7-26.