Inhoudstafel

1. Inleiding 7

> *1.1. Situering* *12*

> *1.2. Onderzoeksobject: Internationale rangschikkingen van publieke sectoren* *17*

> 1.2.1. Soorten, maten, gewichten 19

> 1.2.2. Concepten 21

> 1.2.3. Hoe overheidskwaliteit meten? 24

> 1.2.4. Waarom publieke sectoren rangschikken? 29

> *1.3. Ranking World* *33*

> *1.4. Governance Rankings: Gehypt of beleidsrelevant?* *38*

2. Probleemstelling 43

> *2.1. Governance Rankings: veel geblaat, weinig wol?* *45*

> *2.2. Critici en problemen* *47*

> *2.3. Onderzoeksvragen* *52*

> 2.3.1. Wie produceert de rankings? 53

> 2.3.2. Wat zijn de doelstellingen en het doelpubliek? 54

> 2.3.3. Belang van het meetmodel 57

> 2.3.4. Wat is de kwaliteit van de rankings? 63

> 2.3.5. Hanteerbaarheid voor het (Vlaamse) beleid? 65

> 2.3.6. Hoe groot is de kloof tussen vraag en aanbod? 66

3. Onderzoeksdesign 68

> *3.1. Onderzoeksopzet* *70*

> 3.1.1. Exploratie van kritische literatuur 71

> > 3.1.1.1. Synthese kwaliteitsproblematiek 71

> 3.1.2. Empirisch onderzoek 72

> > 3.1.2.1. Extractie kwaliteitscriteria 74

> > 3.1.2.2. Proefdraai 75

> > 3.1.2.3. Empirische datagaring 76

> > 3.1.2.4. Synthese onderzoeksresultaten 80

> 3.1.3. Klovenanalyse 81

> *3.2. Onderzoeksdoel* *83*
> 3.2.1. Vlaanderen krijgt een zicht op de kwaliteit en (on)bruikbaarheid van rankings van bestuur 84
> 3.2.2. Kwaliteitsbeoordeling voedt het maatschappelijke debat 86
> 3.2.3. De studie legt de ideologische achtergrond van rangschikkingen van publieke sectoren bloot 87
> *3.3. Onderzoekskader* *89*
> 3.3.1. Methodologie 90
> 3.3.2. Theoretisch kader 91

4. Kwaliteitsproblemen 96
> *4.1. Foutief gebruik van subjectieve indicatoren: objectieve feiten versus subjectieve meningen?* *98*
> *4.2. Non-representativiteit van respondenten in surveys* *103*
> *4.3. Conceptuele tekortkomingen in de meting* *108*
> *4.4. Meetfouten: Vergeven en vergeten?* *113*
> *4.5. Onderbelichting van politiek-administratieve dimensie* *116*
> *4.6. Onderhuidse bias* *118*
> *4.7. Over-aggregatie* *122*
> *4.8. Gebrek aan transparantie* *125*
> *4.9. Vergelijkbaarheidsproblemen* *128*
> *4.10. Conclusie: Meten, maar met verantwoordelijkheid voor producenten en gebruikers* *132*

5. Empirisch onderzoek 134
> *5.1. Onderzoeksplan* *136*
> *5.2. Selectie internationale governance rankings* *138*
> 5.2.1. De Worldwide Governance Indicators 139
> 5.2.2. Het World Competitiveness Yearbook 145
> 5.2.3. De Global Competitiveness Index 148
> 5.2.4. De Corruption Perceptions Index 150
> 5.2.5. De Country Risk Briefing 153
> 5.2.6. De Country Policy and Institutional Assessments 154

> 5.2.7. De International Country Risk Guide 156
> 5.2.8. De Public Sector Performance 158
> *5.3. Analyse* *161*
> 5.3.1. Conceptuele dimensie 162
> 5.3.1.1. Validiteit 164
> 5.3.1.2. Opvatting publieke sector 166
> 5.3.1.3. Aard van indicator 168
> 5.3.2. Technische dimensie 172
> 5.3.2.1. Statistische kwaliteit 175
> 5.3.2.2. Data-oorsprong 178
> 5.3.2.3. Analyse-eenheid 180
> 5.3.2.4. Timeframe 183
> 5.3.2.5. Databeschikbaarheid 185
> 5.3.2.6. Duurzaamheid 190
> 5.3.2.7. Survey-eigenschappen 193
> 5.3.3. Politiek-administratieve dimensie 198
> 5.3.3.1. Producent 199
> 5.3.3.2. Visie 202
> 5.3.3.3. Doelstelling(en) 204
> 5.3.3.4. User 208
> *5.4. Conclusies* *213*

6. Klovenanalyse 217
> *6.1. Naar wat zoeken de beleidsmakers?* *218*
> *6.2. Wat kunnen en willen producenten leveren?* *220*

7. Slotconclusie 222
> *7.1. Algemene bevindingen* *225*
> 7.1.1. Aanbevelingen gebruikers 227
> 7.1.2. Werkpunten producenten 231
> 7.1.3. Aanbevelingen Vlaamse beleid 234
> *7.2. Toekomstig onderzoek* *238*
> 7.2.1. De governance rankings van morgen 238
> 7.2.2. Voortzetting wetenschappelijk onderzoek 243

8. Bijlage 244

9. Bibliografie 245

Lijst van figuren

Figuur 1: Input Output Model 25
Figuur 2: Bestuur/Beleid Schema 93
Figuur 3: Verhouding objectieve/subjectieve indicatoren 171
Figuur 4: Interdependentie governance rankings 179
Figuur 5: Landensamples 182
Figuur 6: Uitsplitsing subjectieve indicatoren naar survey-type 196
Figuur 7: Uitsplitsing (aard) producenten 201
Figuur 8: OESO-aanpak 'Government at a glance' 242

Lijst van kaders

Kader 1: Kenmerken van een *'well governed country'* (Besançon, 2003) 23
Kader 2: Kenmerken van een '*well governed country*' (Kauffman & Kraay, 2007b) 23
Kader 3: Selectie van governance rankings voor de meta-analyse 73
Kader 4: Secundaire bronnen governance rankings 78
Kader 5: Kwaliteit van data en statistieken 96
Kader 6: Millennium Challenge Account 208
Kader 7: Evaluatie governance rankings 213
Kader 8: Listing pijnpunten governance rankings 225

Abstract

Het onderzoek 'Internationale vergelijkingen vergeleken: Een meta-analyse van rangschikkingen van publieke sectoren' kadert in het Steunpunt Bestuurlijke Organisatie Vlaanderen. Het vormt een antwoord op de nood op Vlaams beleidsniveau aan een kwaliteitsdifferentiatie van de bestaande benchmarkingsmodaliteiten. Op wetenschappelijk vlak kadert deze studie in de kritische literatuur omtrent de governance rankings.

Op periodieke basis compileren multilaterale organisaties, wereldwijde adviesbureaus en internationale NGO's een brede waaier aan internationale vergelijkingen van de publieke sector. Enerzijds zijn er rangschikkingen van beleidsprestaties, zoals onderwijs, milieu of gezondheid. Anderzijds worden er ook meer en meer bestuursrangschikkingen opgesteld. Deze rankings stoelen op performantiemetingen van het overheidsapparaat. Het zijn deze laatste rankings, de zogenaamde governance rankings, die het voorwerp van deze studie uitmaken.

Dergelijke internationaal-comparatieve oefeningen zijn in de tweede helft van vorige eeuw ontstaan en hebben vooral gedurende de laatste twee decennia enorm aan populariteit gewonnen. De media maken er nieuws van, ministers stofferen er hun beleid mee en de oppositie ziet er het ideale wapen in. Het moge duidelijk zijn: Rankings are here to stay. Samen met de verhoogde belangstelling voor de kwaliteit van de overheid, heeft dit voor een 'ranking boom' gezorgd. Sommige auteurs gewagen zelfs van een 'mushrooming industry'.

De proliferatie van de governance rankings heeft er echter voor gezorgd dat de vergelijkingen van publieke sectoren tot een onoverzichtelijk landschap zijn vergleden waarvan niemand de waarde nog kan inschatten. Overheden die zich internationaal willen positioneren of willen benchmarken zijn genoodzaakt aan 'ranking shopping' te doen, zonder te weten wat de kwaliteit van de metingen echt is. De cijfers worden meestal zonder veel discussie aangenomen, hoewel de kwaliteit van de internationale rangschikkingen van publieke sectoren vaak te wensen over laat. Onze studie wenst in te spelen op deze problematiek door een soort vergelijking der vergelijkingen te maken.

Het uitgangspunt van de auteurs luidt immers dat internationale rangschikkingen van publieke sectoren vanuit bestuurskundig oogpunt als beleidsinstrument een bijzonder belangrijke functie hebben te vervullen. Als beleidsinformatie moeten dergelijke metingen de beleidsmaker bijstaan in het internationaal positioneren van zijn overheidsapparaat en hem de kans bieden best practices in andere landen te identificeren. Op die manier zijn governance rankings instrumenteel in het opvoeren van externe druk op de overheid om de efficiëntie en effectiviteit van de publieke sector te verhogen. De juiste beleidslessen kunnen echter pas getrokken worden als duidelijk is wat er effectief gemeten wordt, welke technisch-statistische kwaliteit de data herbergen en hoe betrouwbaar de vergelijkingen zijn. Met andere woorden: Wat is de intrinsieke waarde van data uit de rankings?

Om hier een zicht op te krijgen en een tegengewicht te bieden voor de ongenuanceerde conclusies die momenteel uit de governance rankings worden getrokken, hebben we in dit onderzoek een kwaliteitsdoorlichting van dergelijke internationale vergelijkingen opgemaakt. In onze studie zijn drie pijlers te onderscheiden. Ten eerste hebben we een meta-synthese opgemaakt van de kritische literatuur omtrent metingen en rangschikkingen van publieke sectoren. Ondanks het relatief prille bestaan van de governance ranking, heeft het gepercipieerde kwaliteitsdeficit reeds een horde internationale auteurs tot kritische analyses bewogen. De kritieken waren echter nog diffuus en stonden vaak op zich. Aan de hand van een diepgaande literatuurstudie hebben we een listing opgemaakt van de kwaliteitsgebreken die in internationale studies reeds aan het licht zijn gekomen. In deze meta-synthese zijn we tot een lijst van negen pijnpunten gekomen. Het is onze overtuiging dat deze gebreken dienen meegenomen te worden voor een goed begrip van de governance data.

Ten eerste is het een problematisch punt van de meeste internationale rangschikkingen van publieke sectoren dat zachte data (percepties, ervaringen, inschattingen) -die in grote mate ingang vinden in de governance rankings- foutief geïnterpreteerd worden als feitelijkheden. Men beweert de prestaties van de overheid in kaart te brengen, maar kwantificeert veeleer het imago van de publieke sector. Bovendien vertolken de respondenten die bevraagd worden ter invulling van deze subjectieve indicatoren in vele gevallen niet de publieke opinie. Deze tweede kritiek spitst zich toe op de scheefgetrokken surveysamples die aan de basis liggen van veel governance data. Zo komt het vaak voor dat de

efficiëntie van de overheid in feite de inschatting van het overheidsapparaat door (internationale) bedrijfsleiders weerspiegelt. Op die manier krijgt de 'efficiënte overheid' wel een erg enge en vertekende invulling. Ook de kleine samplegroottes en de ondoorzichtige methodologie bij dergelijke bevragingen doen de betrouwbaarheid van de rankings geen goed. Ten derde is er een groot gebrek aan conceptuele kaders aan de basis van de overheidsmetingen. Wat er eigenlijk gemeten wordt, lijken veel rankers zelf niet te weten. De inhoud van de metingen is bij gebrek aan theoretische basis enkel af te leiden uit de indicatorenselectie. De keuze voor deze of gene variabele berust echter vooral op praktische gronden, zoals beschikbaarheid. Het vierde obstakel voor betrouwbare internationale rangschikkingen van publieke sectoren wordt gevormd door de grote meetfouten, die bovendien in de meeste gevallen doodgezwegen worden. Slechts twee governance rankings uit ons empirisch onderzoek maken gewag van meetfouten. In beide gevallen verminderen de opgemaakte betrouwbaarheidsintervallen in grote mate de comparabiliteit binnen de rankings. Landen uit de 'messy middle' die gelijkaardige scores vertonen, kunnen daardoor niet op significante wijze van elkaar onderscheiden worden. De andere metingen van publieke sectoren kampen meer dan waarschijnlijk nog met grotere measurement errors. De producenten van deze rankings opteren er echter voor ze te verzwijgen uit schrik voor een devaluatie van hun product.

Ten vijfde klagen we het gebrek aan aan zichtbaarheid van -wat we noemen- de politiek-administratieve dimensie van de rankings. De vragen 'Wie meet er?', 'Waarom wordt er gemeten?' en 'Voor wie wordt er gemeten?' komt onzes inziens nog te weinig aan bod in de rankings en ook de kritische literatuur gaat grotendeels voorbij aan deze beïnvloedende factoren. We trachten ten zesde ook de onderhuidse biases die in veel rankings schuilgaan, bloot te leggen. Veel rankings gaan gebukt onder een vertekening in een of andere richting. Arbeidsrechten of milieurichtlijnen die als positieve of negatieve punten op het performantierapport van een overheid verschijnen, betreft een normatieve keuze. Of je ermee akkoord gaat of niet, is betwistbaar. Belangrijk is echter vooral dergelijke politieke keuzes te expliciteren, zodat er rekening mee kan gehouden worden. Een zevende kritisch punt behandelt de over-aggregatie binnen de governance rankings. Doordat producenten steeds naar een uni-dimensionale rangschikking willen werken, worden alle onderliggende variabelen geaggregeerd tot een index waarvan de betekenis nauwelijks nog kan gevat

worden. Bovendien gaat er een hoop informatie verloren in het aggregatieproces, wat verder versterkt wordt door de nadrukkelijke focus op de rankings als 'horse races'. Niet de negatieve punten op het vlak van de overheidsefficiëntie komen onder de aandacht, het verwordt meer tot een hitparade waarin een ander land 'verslaan' het doel is. Vervolgens kaarten we ten achtste het gebrek aan transparantie van de internationale rangschikkingen van publieke sectoren aan. Zowel op het vlak van databeschikbaarheid, als methodologische openheid is er nog veel vooruitgang te boeken door de producenten van de diverse internationale rangschikkingen van publieke sectoren. Ook de hoge abonnementsprijzen voor de datatoegang die bij meerdere governance rankings wordt gevraagd, vormt een belemmering voor de transparantie en komt ook de leermogelijkheden voor het beleid niet ten goede. Tot slot maken we in een negende punt enkele kanttekeningen bij de idee van vergelijkingen tussen landen op het vlak van de overheidskwaliteit. Zo laten we enkele auteurs aan het woord die van mening zijn dat de diversiteit in culturele, politieke en omgevingsfactoren tussen landen het niet toelaat publieke sectoren betekenisvol over landsgrenzen heen te vergelijken.

Naast de synthese van de kwaliteitsproblematiek rond de governance rankings, hebben we in ons onderzoek ook een empirische analyse van rangschikkingen van publieke sectoren opgemaakt. Aan de hand van een standaard analyse-instrument -de meta-fiche- hebben we een selectie van acht frequent gebruikte en hoogaangeschreven governance rankings aan een beschrijvend en beoordelend onderzoek onderworpen. Onder meer op basis van de methodologische documenten, interviews met producenten en secundaire bronnen hebben we getracht de kwaliteit van de rankings in kaart te brengen. Onder de conceptuele dimensie zijn we nagegaan wat er beweerd wordt te meten in de internationale rangschikking van publieke sectoren en dat afgezet ten aanzien van wat er werkelijk gemeten wordt. Onder de technische dimensie hebben we aspecten die verband houden met de statistisch-technische kant van de meting geëvalueerd. En in de politiek-administratieve component tot slot hebben we de aard van de producent, diens objectieven en de beoogde doelgroep trachten te vatten.

Resultaat van deze oefening was een meta-fiche voor elke ranking. Daarin wordt ten eerste de governance ranking in zijn geheel besproken, aan de hand van een kwalitatieve beschrijving voor elk van de 33 geïdentificeerde kwaliteitscriteria. Vervolgens wordt we op een tweede niveau het

onderdeel, de component, de indicator binnen deze ranking besproken die een meting uitvoert van de kwaliteit van het overheidsapparaat. Tot slot wordt we dat laatste onderdeel ook tot op indicatorenniveau ontrafeld. Door elke individuele variabele die aan de basis ligt van de meting van 'de efficiënte of effectieve overheid' bloot te leggen en te beschrijven, wordt pas echt duidelijk hoe de overheidsmeting concreet door de producent werd geoperationaliseerd. Dit laatste weerspiegelt de inhoudelijke waarde van de governance ranking; de validiteit van de meting.

De meta-fiche moet beleidsmakers, aan de hand van de beoordelende en beschrijvende evaluatie van acht gerenommeerde governance rankings, een beeld opleveren van de kwaliteit van de verschillende overheidsmetingen. Op die manier moet ze uitgroeien tot een bijsluiter van de respectievelijke internationale rangschikkingen van publieke sectoren. Als cruciaal achtergrondinstrument dient ze gebruikers van governance data te wijzen op de tekortkomingen en biases. Enkel wanneer het beleid de werkelijke waarde van een governance ranking kan inschatten, kan ze een betekenisvol debat opstarten over de conclusies van de meting. Door de plus- en minpunten van de frequent gebruikte rankings bloot te leggen, willen we enerzijds gebruikers waarschuwen voor betekenisloze vergelijkingen, eenzijdige interpretaties en ongenuanceerde conclusies. Anderzijds willen we er de producenten van governance data mee aansporen de kwaliteit van hun producten te garanderen en werk te maken van de vastgestelde pijnpunten.

Naast de synthese van de kwaliteitsproblematiek, de diepgaande analyse in de meta-fiches, hebben we in een derde pijler ook een klovenanalyse opgesteld. In deze klovenanalyse hebben we een blik geworpen op de kloof tussen de verwachtingen van beleidsmakers ten aanzien van governance rankings en de doelstellingen en overwegingen van de producenten achter dergelijke metingen. Meer specifiek hebben we onderzocht wat de wensen en verzuchtingen aan Vlaamse zijde zijn. Vlaanderen heeft de ambitieuze doelstelling in afzienbare tijd aansluiting te vinden op wereldvlak bij het koppeloton van regio's. In het kader van deze doelstelling wenst het zich zoals nationale overheden internationaal te positioneren en tracht het haar beleid te benchmarken. Problematisch is echter het feit dat internationale rangschikkingen van publieke sectoren veelal voorbij gaan aan regio's. Van de acht onderzochte governance rankings in onze empirisch onderzoeksgedeelte, heeft enkel het International Institute for

Management Development in haar World Competitiveness Yearbook ooit haar geografische blik verruimd door enkele regio's in de ranking op te nemen. Dit initiatief stamt echter al van enkele jaren terug en kreeg geen vervolg. Producenten van governance rankings geven in interviews mee dat het te veel tijd, energie en geld kost om regionale data te verzamelen. Zolang de meeste rankers hun rangschikking als een product beschouwen, zal deze trend pas gekeerd kunnen worden op het moment dat er voldoende (gecoördineerde) vraag naar regionale inclusie komt van de (autonome) regio's zelf.

We besluiten dit rapport door in een concluderend hoofdstuk de gedetecteerde gebreken van de internationale rangschikkingen van publieke sectoren van vandaag te vertalen in enerzijds aandachtspunten voor gebruikers en anderzijds werkpunten voor producenten van governance rankings. Op korte termijn kunnen we in beleidsdiscussies immers pas vooruitgang boeken indien er op een omzichtige en genuanceerde wijze met de conclusies uit de governance rankings wordt omgesprongen. Pas indien duidelijk is wat er in de ranking juist gemeten wordt, wie er meet, voor wie er gemeten wordt en waarom er gemeten wordt, en hoe betrouwbaar de vergelijkingen zijn, kan er op een betekenisvolle manier over de inhoud gesproken worden. Tot dit doel heeft deze studie een bijdrage willen leveren. Op middellange termijn is het bovendien eveneens noodzakelijk dat producenten van governance rankings meer aandacht aan de kwaliteit van hun product -de governance ranking- gaan besteden. In eerste instantie dienen zij meer aandacht te hebben voor de primaire functie van dergelijke instrumenten; nl. het creëren van beleidsinformatie. Daarbij is van groot belang dat overheden niet alleen de 'blame and shame' van een slechte score ondervinden, maar dat er ook leereffecten gecreëerd worden. Uiteindelijk moet de benchmarking ervoor zorgen dat slecht bestuurde landen gewezen worden op hun tekortkomingen en aan de hand van best practices uit beter presterende landen hun score kunnen opkrikken.

1. Inleiding

In deze studie laten we ons licht schijnen op de kwaliteit van internationale rangschikkingen van publieke sectoren. Er kan daarbij een onderscheid gemaakt worden tussen rangschikkingen van beleid en rangschikkingen van bestuur. In dit onderzoek focussen we ons op deze laatste. We laten daarbij dus sectorale prestaties (zoals onderwijs, gezondheidszorg, …) buiten beschouwing. We bekijken enkel wat de bestaande governance rankings ons vertellen over de prestaties van het overheidsapparaat, over de kwaliteit van de overheid en over de gepercipieerde *'government effectiveness*'. Met het onderzoek 'Internationale vergelijkingen vergeleken: Een meta-analyse van rangschikkingen van publieke sectoren' trachten we een algemeen beeld te schetsen van de governance rankings die vandaag de dag worden opgevoerd door de media, beleidsmakers, politici, onderzoekers en anderen. Enerzijds doen we dit door middel van een synthese van de kwaliteitsproblematiek aan de hand van een grondig literatuuronderzoek. Anderzijds voeren we ook zelf een grondige analyse uit van acht hoogaangeschreven en frequent geciteerde governance rankings. We ontleden deze rankings, ontrafelen hun onderliggende structuur, leggen hun achtergrond bloot en onderzoeken hun gebruik. Een blik achter de cijfers heeft als doel zicht te bieden op de kwaliteit van deze individuele internationale governance rankings. Het doel van deze studie is dan ook het Vlaamse beleid een beter idee te geven van welk materiaal er op het vlak van comparatieve overheidsmetingen bestaat en welke waarde er aan gehecht mag worden. We trachten hier ook de dialoog tussen gebruikers en producenten van governance rankings te voeden. We beginnen met een overzicht van de items die in dit eindrapport aan bod zullen komen.

De inleiding geeft in een eerste paragraaf een historisch overzicht van de snel groeiende markt van internationale rangschikkingen van publieke sectoren. De interesse is groot, de productie navenant. We tonen aan dat men niet blind mag zijn voor de tekortkomingen van de governance rankings. We introduceren ook enkele begrippen die noodzakelijk zijn voor een goed begrip van het onderzoeksobject van deze studie. Na een korte omschrijving van enkele meetconcepten zoals de indicator, de index en de rangschikking, trachten we te schetsen welke aspecten van de overheid nu juist gemeten worden in de door ons onderzochte rankings. Ook de

afbakening van dit onderzoek komt hierbij aan bod. De waaromvraag van het meten van publieke sectoren sluit het onderdeel over het onderzoeksobject af. Vervolgens schetsen we de opmerkelijke opmars van de rankings, die in de loop van de vorige eeuw begonnen op te komen, maar vooral de laatste decennia een forse groei hebben gekend. Ten slotte raken we al even aan de controverse die rond de internationale rangschikkingen van publieke sectoren hangt. We geven aan waar we de critici moeten situeren, alsook welke twijfels zij doorgaans uiten.

Na de inleiding, volgt in hoofdstuk 2 de probleemstelling. In dit onderdeel trachten we deze studie te situeren. We schetsen de problematiek die inherent is aan de thematiek van de rangschikkingen van publieke sectoren. We duiden hun prille bestaan, maar ook de forse groei die ze de laatste 20-tal jaren hebben gekend en het kwaliteitsdeficit dat hen verweten wordt. Ondanks het feit dat de governance rankings pas sinds begin jaren '90 echt op het voorplan kwamen, is er immers toch al een groep auteurs opgestaan die zich bezighoudt met het analyseren en bekritiseren van de bestaande governance rankings. We opperen voorts enkele onderzoeksvragen die doorheen de studie aan bod zullen komen. Zo willen we de politiek-administratieve dimensie van de rankings onder de aandacht brengen: wie meet er, waarom wordt er gemeten en voor wie wordt er gemeten? We poneren de stelling dat de kwaliteit van de internationale rangschikkingen van publieke sectoren pas kan onderzocht worden als we de structuur van de rankings kennen. Enkel zo kunnen we immers de onderliggende indicatoren blootleggen die betekenis geven aan de metingen. Ook de vraag naar de hanteerbaarheid van het Vlaams beleid van deze internationale rangschikkingen wordt beschreven. Tot slot werpen we een klovenanalyse op waarin het verschil wordt onderzocht tussen wat producenten van internationaal comparatief onderzoek van publieke sectoren kunnen en willen bieden enerzijds, en waar gebruikers eigenlijk nood aan hebben anderzijds. We leggen daarbij extra de nadruk op de noden op het Vlaamse beleidsniveau.

Hoofdstuk 3 schept het kader, het design van dit onderzoek. In de eerste paragraaf wordt het onderzoeksopzet uiteengezet. We beschrijven hoe we enerzijds een synthese willen maken van de kritische literatuur over de internationale rangschikkingen van publieke sectoren. Anderzijds geven we een overzicht van de methodologie van ons empirisch onderzoek. Ook het opzet van de klovenanalyse wordt hier gesitueerd. In het onderzoeksdoel

splitsen we de algemene doelstelling van het analyseren van de kwaliteit van de governance rankings uit in drie concrete deelvragen en onderzoeksdoelstellingen. Ten slotte lichten we in het onderzoekskader onze manier van werken toe en geven we het theoretische kader mee waarbinnen we dit onderzoek voeren.

In het daaropvolgende vierde hoofdstuk bespreken we de problemen die de kwaliteit van de internationale rangschikkingen van publieke sectoren aantasten. De oplijsting van deze problemen komt voort uit de kritische literatuur en uit onze empirische analyse van acht internationale governance rankings. De interpretatieproblemen van subjectieve indicatoren komen aan bod in de eerste paragraaf. Daarnaast houden we de representativiteit van de respondenten bij subjectieve metingen tegen het licht. We schetsen vervolgens de tekortkomingen die er op conceptueel vlak zijn. Een vierde paragraaf is gewijd aan de betrouwbaarheid van de meting; hoe er wordt omgegaan met meetfouten. De onderbelichting van de politiek-administratieve dimensie van de rankings, die wij willen toevoegen aan de internationale onderzoeksliteratuur, is een vijfde item dat aan bod zal komen. De onderhuidse bias die vaak over het hoofd gezien wordt, is het volgende probleem dat aangekaart wordt. Ten zevende bespreken we de kwaliteitsproblemen die over-aggregatie van de data in governance rankings met zich meebrengt. Het gebrek aan transparantie is het onderwerp van de achtste paragraaf. De twijfels in verband met de vergelijkbaarheid is het negende probleem dat we behandelen. Dit vierde hoofdstuk sluiten we tot slot we door deze problemen samen te brengen in een conclusie. Het standpunt dat we verdedigen is dat de veelheid aan kwaliteitsproblemen niet het einde van het internationaal vergelijken van publieke sectoren mag inluiden. De diverse tekortkomingen moeten gebruikers wel aanmanen tot voorzichtigheid bij het interpreteren, overnemen en gebruiken van governance rankings, zoals ze vandaag bestaan. Ook de producenten moeten hun lessen trekken uit het vastgestelde kwaliteitsgebrek. Het is hun verantwoordelijkheid om de kwaliteit op te krikken, waardoor de betrouwbaarheid en hanteerbaarheid van hun product omhoog gaat.

In hoofdstuk 5 beschrijven we het empirisch onderzoek dat de kern vormt van deze studie. We beginnen met het uiteenzetten van ons onderzoeksplan, waarin we aangeven hoe we te werk zijn gegaan in onze analyse. Vervolgens geven we een woordje uitleg bij de selectie van de acht internationale rangschikkingen van publieke sectoren. Hierbij omschrijven

we alle acht de internationale rangschikkingen ook even ter kennismaking. Zo komen achtereenvolgens de Worldwide Governance Indicators (World Bank Institute), het World Competitiveness Yearbook (International Institute for Management Development), de Global Competitiveness Index (World Economic Forum), de Corruption Perceptions Index (Transparency International), de Country Risk Briefing (Economist Intelligence Unit), de Country Policy and Institutional Assessments (World Bank), de International Country Risk Guide (PRS Group) en de Public Sector Performance (European Central Bank) aan bod. De analyse geeft een overzicht van de bevindingen op de diverse beschrijvende en beoordelende criteria die we bij elk van de geselecteerde internationale rangschikkingen van publieke sectoren onderzocht hebben. In de conceptuele dimensie gaat het om de validiteit, de opvatting over de publieke sector en de aard van de indicator. Onder de vlag van de technische dimensie hebben we achtereenvolgens de statistische kwaliteit, de dataoorsprong, de analyse-eenheid, het *timeframe*, de databeschikbaarheid, de duurzaamheid en de survey-eigenschappen van naderbij bekeken. De politiek-administratieve dimensie ten slotte vertelt ons meer over de producent, diens visie en doelstellingen en de doelgroep die de rankings voor ogen hebben. In de conclusie zetten we enkele opmerkelijke bevindingen op een rijtje. De output van het empirisch onderzoek is terug te vinden in de acht individuele meta-fiches die aan de hand van kwaliteitscriteria een goed beeld scheppen van de structuur en kwaliteit van de respectievelijke governance rankings. Daar deze tabellen in totaal bijna 50.000 beschrijvende cellen omvatten, is het praktisch onmogelijk om deze fiches in dit rapport te ontsluiten. Omdat het echter grote zonde zou zijn deze omvangrijke bron aan informatie verloren te doen gaan, zullen alle gegevens uit onze analyse op een elektronische informatiedrager bij het rapport geleverd worden. De annexen bij dit rapport zullen de belangrijkste onderdelen van de meta-fiches op papier bevatten. Het gaat daarbij om de analyse van het onderdeel van elke governance ranking dat 'de kwaliteit van de overheid' tracht te meten. De meta-fiches kunnen een vitale rol spelen in het interpreteren van de data van de onderzochte governance rankings. Ze vormen als het ware een soort van bijsluiter bij de data waar ze betrekking op hebben. Net zoals bij medicijnen het geval is, moet verantwoord gebruik van de data uit de governance rankings bevorderd worden met deze bijsluiters.

In hoofdstuk 6 snijden we de klovenanalyse aan. Doorheen het onderzoek hebben we selectieve gesprekken gevoerd met zowel beleidsmakers op het

Vlaams niveau, als producenten die aan de basis liggen van enkele internationale rangschikkingen van publieke sectoren. Aan de hand van deze communicatie en verdere studie geven we een aanzet tot overzicht van de vraag- en aanbodzijde. We stellen ons de vraag waar het (Vlaamse) beleid nood aan heeft op het vlak van internationaal comparatief onderzoek van publieke sectoren. En aan de andere kant proberen we uit te zoeken wat de beweegredenen zijn van de internationale organisaties en adviesbureaus om governance rankings te produceren.

In hoofdstuk 7 zetten we nog even onze slotconclusies op een rijtje. We sommen de gedetecteerde kwaliteitswerkpunten nog eens op en schatten de implicaties ervan in voor zowel gebruikers, als producenten. We vertalen de kwaliteitsproblemen enerzijds in aandachtspunten voor de *users* en anderzijds in werkpunten voor de *rankers*. Ook de gevolgen van het vastgestelde kwaliteitsdeficit voor het Vlaamse beleidsniveau worden summier besproken. Naast deze algemene bevindingen in de eerste paragraaf, werpen we in de tweede paragraaf ook een blik op de toekomst. Enerzijds doen we enkele suggesties voor toekomstscenario's op het vlak van meten en vergelijken van publieke sectoren. Onder andere de werkwijze van het OESO-project *'Government at a glance*', dat normaliter in 2009 van start zal gaan, wordt hier uit de doeken gedaan. Anderzijds geven we een aanzet voor toekomstig (academisch) onderzoek. We geven aan waar de wetenschap nog een bijdrage kan leveren in de discussie rond de rankingproblematiek?

Hoofdstuk 8 bevat de bijlagen, hoofdstuk 9 lijst de bibliografie van dit onderzoek op.

> 1.1. Situering

Je kunt tegenwoordig geen krant meer openslaan zonder te lezen dat ons land op beleidsdomein x of y beter, dan wel slechter presteert dan andere landen. Vooral de berichten die melding maken van een minderwaardige positie van onze overheid op de internationale ladder, krijgen pakken aandacht. Internationale rangschikkingen van publieke sectoren zijn in. Ze lijken simpel, eenduidig en overtuigend en zijn daardoor een geliefd instrument voor journalisten en politici. De vermeende simpliciteit vormt een belangrijke factor in de populariteit van internationale overheidsrangschikkingen. In de vorm van een geaggregeerde score en een positie op de internationale ladder, komen internationale rangschikkingen van publieke sectoren tot een ongecompliceerde, maar vaak ook ongenuanceerde conclusie in verband met de prestaties van een bepaald land. Naast de vormelijke aantrekkelijkheid van de governance rankings, ligt er nog een meer fundamentele tendens aan de basis van de toegenomen aandacht voor rangschikkingen van publieke sectoren. De laatste decennia werden immers gekenmerkt door een stijging van de interesse voor de kwaliteit van de overheid. De blikverruiming van een louter economische focus, naar 'goed bestuur' -*good governance*- bij het beoordelen van publieke sectoren, is het gevolg van diverse factoren.

Schandalen binnen de publieke sector, een immer dalend vertrouwen van de burger in overheidsinstellingen en een toenemende druk op de overheidsfinanciën hebben overheden ertoe aangezet om de efficiëntie en de effectiviteit van de publieke sector hoog op de agenda te plaatsen (Van de Walle et al., 2004). Het verhogen van de overheidsefficiëntie en het verbeteren van de overheidsprestaties zijn vooral de laatste 20 jaar uitgegroeid tot belangrijke werkpunten voor de overheid. Goede illustraties hiervan kunnen we onder andere terugvinden in Engeland en de Verenigde Staten. Over het Kanaal publiceerde en implementeerde de Labourregering in 2002 de *Public Service Improvement Strategy* (Office of Public Service Reform) die de performantie van de overheid op een hoger niveau moest tillen. In de Verenigde Staten moesten de *National Performance Review and Government Performance Results Act* (1993) en de *Management Agenda* van president Bush een zelfde doel dienen (Boyne & Walker, 2005). In ons land beoogden onder andere de Copernicushervorming op federaal en Beter Bestuurlijk Beleid op Vlaams niveau hetzelfde doel. Vanuit dit perspectief groeide de vraag naar internationaal vergelijkend datamateriaal over de

overheidsprestaties ook bij beleidsmakers en hervormers. Internationale rangschikkingen van publieke sectoren pikken rechtstreeks in op deze noden.

Daarnaast werden de internationale metingen van overheidskwaliteit ook vanuit het ontwikkelingsdenken gepromoot. Zo benadrukte voormalig Secretaris-generaal van de Verenigde Naties Kofi Annan niet zo lang geleden het belang van goede metingen van overheidsprestaties: "*Good governance is perhaps the single most important factor in eradicating poverty and promoting development. […] If governance matters, so does the need for more reliable and valid data on key governance processes*" (United Nations, 1998).

Deze dynamieken hebben een forse groei van het aantal governance rankings met zich meegebracht (Arndt & Oman, 2006). Kauffman, Kraay en Mastruzzi (2006) -met de Worldwide Governance Indicators, de auteurs van de bekendste rangschikking van publieke sectoren- spreken zelfs van een *'virtual explosion of datasets'*.

Vanuit beleidsperspectief zijn internationale rangschikkingen van publieke sectoren interessant omdat ze bijdragen tot een kwantitatief onderbouwd beleid. Daar waar beleidsmakers initieel overwegend op basis van anekdotische bewijsvoering -vaak op basis van misleidende en onvolledige informatie- de kwaliteit van het openbaar bestuur inschatten, is de voorbije decennia een steviger gefundeerd overheidsmanagement ontstaan dat zich onder andere baseert op internationale rankings (Kauffman et al., 2000). Het populaire *'Management by numbers'*, waarbij de publieke sector beheerd wordt met behulp van data over de producten (outputs), de effecten (outcomes), de middelen (inputs) en de processen (process), ligt in het verlengde hiervan (Hood, 2007).

In de typologie van Hood, vallen er drie soorten systemen onder de vlag van een *'Management by numbers'*. Ten eerste zijn er de *'target systems'*. Deze managementinstrumenten zetten de effectieve prestaties af ten opzichte van een of meerdere standaarden, uitgedrukt in drempelwaarden. Het tweede soort systeem wordt gevormd door de *'intelligence systems'*, die prestaties meten als achtergrondinformatie, zonder daar een vaste interpretatie aan vast te knopen. Ten derde beschrijft Hood de *'ranking systems'*, die in deze studie onder de loep genomen worden. Deze

rangschikkingen meten de huidige of verleden prestaties van vergelijkbare diensten en plaatsen deze op internationale schaal tegenover elkaar. Dit gebeurt onder andere met het oog op aanvullende informatie om consumentenbeslissingen te vergemakkelijken, als informatie-instrument voor overheidsoptreden, of simpelweg als een manier om de '*saints*' aan te moedigen en de '*sinners*' te bekritiseren.

Daarnaast laten internationale rangschikkingen van publieke sectoren toe om de eigen overheidsprestaties in absolute en relatieve termen te beoordelen. Een land kan zich immers pas internationaal positioneren als het haar resultaten weet af te wegen ten opzichte van andere overheden in de wereld. Ook Vlaanderen ziet hierin een vertrekpunt voor beleidsformulering. In zijn Septemberverklaring (2007) benadrukte Minister-president Kris Peeters het belang van internationale positionering. "Elke mens streeft ernaar gelukkig te zijn. Het is de essentiële opdracht van de Vlaamse regering om de voorwaarden te creëren waaronder elke Vlaming geluk kan vinden en zijn levensplan kan realiseren. Daarom moet de Vlaamse regering vooruitzien. En om vooruit te kunnen zien, moet zij weten waar Vlaanderen staat. Pas dan kunnen we, door reflectie én door actie, de problemen van vandaag oplossen en de uitdagingen van morgen aanpakken." Op basis van het *'best practices*'-principe trachten overheden lessen te trekken uit de nationale verhalen van beter presterende overheden. In dit laatste geval spreekt men ook van 'benchmarken'. *Benchmarking* komt overeen met het vergelijken van een verzameling van vergelijkbare *units* met de best presterende onder hen, met het oog op het verbeteren van de prestaties van de eigen organisatie (De Groot, 2004). Het komt in wezen neer op het vergelijken van verschillende beleidsopties om te zien welk pad tot de beste resultaten leidt. Of zoals het Sociaal en Cultureel Planbureau, dat op periodieke basis een internationale vergelijkingsoefening voor Nederland opmaakt over de ontwikkelingen en veranderingen in de samenleving, het passend verwoordt: "Internationale vergelijkingen helpen om de onzekerheden beter in kaart te brengen en de met verschillende beleidsopties verbonden kansen en risico's nader te bepalen" (Uitterhoeve, 2000).

Daar waar oorspronkelijk overheidsstatistieken als basis gebruikt werden voor de comparatieve *'league tables'*, nemen nu vooral internationale organisaties, private ondernemingen, NGO's en adviesbureaus het voortouw in het opstellen van internationale rangschikkingen van publieke sectoren

(Hood, 2007). De media maken er nieuws van en de ministers stofferen hun beleid ermee. Ook voor beleidsmakers en hervormingsgezinden zijn internationaal vergelijkende indicatoren een zegen.

Gevaarlijk wordt het echter wanneer dergelijke vergelijkingsoefeningen overweldigende aandacht genieten, zonder dat de kwaliteit van de voortgebrachte data navenant is. In dat schuitje zitten de huidige rangschikkingen van publieke sectoren nu net. De internationale rangschikkingen die de werking van de overheid -het goed bestuur of de *'government effectiveness'* zoals men wil- onder de loep nemen, genieten grote aandacht onder politici, media en beleidsmakers. De conclusies worden echter te gemakkelijk voor waar aangenomen, men heeft te weinig oog voor de tekortkomingen van dergelijke internationale rankings (Van de Walle et al., 2004). Ondanks deze vaststelling, blijven de internationaal vergelijkende data op het vlak van de overheidswerking jaarlijks uit de geoliede machine van veel internationale organisaties en private adviesbureaus stromen. *'Rankings are here to stay*', ze maken blijkbaar deel uit van de *'Measurement Culture*' die heden ten dage bestaat (Kuha & Gaskell, 2005).

Het is bijgevolg aan de wetenschap om de taak op zich nemen om de kwaliteit van internationale rangschikkingen van publieke sectoren te evalueren en de gebreken onder de aandacht te brengen. Zoals Thomson het treffend verwoordde op de conferentie *'Where does Britain rank? International Public Service Rankings'* die in 2005 in Londen plaats had: "*Academics could act as a 'parallel source of intelligence' on international rankings, serving to expose those of poorer quality and to act as a stabilizing influence on the 'media frenzy' that attended new publications*" (Thomson, 2005b). Vanuit dat perspectief trachten we een duidelijke nuance in het verhaal van de internationale rankings aan te brengen. We zullen vanuit een kwaliteitsonderzoek van de data, gebruikers en producenten op hun verantwoordelijkheden wijzen. De eersten dienen verder te kijken dan de conclusies die de rankings poneren en moeten omzichtiger te werk gaan bij de interpretatie van de data. Daarnaast willen we de producenten van governance rankings aansporen meer aandacht aan de kwaliteit van hun product te besteden.

Dit onderzoek, dat kadert in het Steunpunt Bestuurlijke Organisatie Vlaanderen, kadert in deze wetenschappelijke doelstelling en heeft tot

concreet doel de kwalitatieve gebreken in het landschap van internationale rangschikkingen van publieke sectoren bloot te leggen. Enerzijds gebeurt dit met het oog op het verminderen van verkeerdelijk gebruik van governance data door beleidsmakers en andere gebruikers. Een goed begrip van wat de bestaande internationale rankings werkelijk meten en hoe betrouwbaar de resultaten in feite zijn, moet een aanzet bieden voor het trekken van correcte en gepaste beleidslessen op verschillende beleidsniveaus. Anderzijds willen we als onderzoekers ook een bijdrage leveren aan een verbeterde constructie van internationale rangschikkingen van publieke sectoren. En kwaliteitsvollere governance ranking moet op haar beurt de algemene betrouwbaarheid en hanteerbaarheid van deze informatiedragers de hoogte instuwen.

> 1.2. Onderzoeksobject: Internationale rangschikkingen van publieke sectoren

De internationale rangschikkingen van publieke sectoren zijn zoals aangegeven uitgegroeid tot een populair instrument en worden te pas en te onpas gebruikt in beleidsdiscussies. De kwaliteit van de overheid krijgt immers meer en meer aandacht, wat ook de vraag naar indicatoren en rankings de hoogte in jaagt. Deze studie gaat na in welke mate de data en de conclusies uit dergelijke vergelijkingsoefeningen betrouwbaar zijn. Een focus op de algemene kwaliteitsproblematiek gecombineerd met een diepgaande analyse van een aantal internationale rangschikkingen van publieke sectoren, moet een betere kijk op de kwaliteit opleveren. Om de studie juist te kaderen, bakenen we eerst het onderzoeksterrein verder af.

Rangschikkingen van publieke sectoren, hetzij vanuit nationaal hetzij vanuit internationaal perspectief, vergelijken de overheidsprestaties van landen in kwantitatieve termen. In het jargon spreekt men ook wel van *'governance rankings*', samengesteld uit *'governance indicators*' (Van de Walle, 2005). Internationale rangschikkingen van publieke sectoren vormen echter een vlag die een grote lading dekt. Van Roosbroek en Van Dooren (2006) maken een onderscheid tussen enerzijds rangschikkingen van beleid en anderzijds rangschikkingen van bestuur. De rangschikkingen van beleid betreffen een vergelijking tussen de beleidsprestaties in een afgelijnd bevoegdheidsdomein. Zo bestaan er rankings die de economie, het onderwijs of de gezondheidszorg in verschillende landen tegenover elkaar plaatsen. Het Programme for International Student Assessment (PISA) van de Organisation for Economic Co-operation and Development (OECD), waarbij landen worden vergeleken op basis van de kennis en vaardigheden op onderwijsgebied van 15-jarigen is een schoolvoorbeeld van dit soort rangschikkingen.

De internationale rangschikkingen van bestuur daarentegen voeren een meting uit van de bestuurlijke prestaties van nationale overheden op een internationale schaal. Voorbeelden hiervan zijn de rangschikkingen die landen vergelijken op basis van de openbare schuld, de corruptie of de E-governance in de publieke sector. Ook de diverse internationale rangschikkingen die de kwaliteit van de overheid -in termen van kwaliteit, efficiëntie of effectiviteit- meten, behoren tot deze soort. In onze empirische analyse, zullen we ons focussen op de (onderdelen van)

rangschikkingen van publieke sectoren, die een meting uitvoeren van de kwaliteit van de overheid. De invulling van het concept 'overheidskwaliteit', zo zullen we later nog uitvoerig illustreren, is echter niet altijd duidelijk en eenduidig af te lijnen. Zo is er een grote conceptuele verwarring ontstaan onder de internationale rangschikkingen van bestuur.

Samengevat spitst deze studie zich dus toe op de analyse van internationale rangschikkingen van bestuur waarin de werking van de (ganse) overheid centraal staat. Het studieobject wordt bijgevolg gevormd door internationale rangschikkingen, geproduceerd door internationale organisaties of commerciële adviesbureaus, die landen vergelijken op het vlak van hun bestuurlijke performantie. We kiezen voor een topdown perspectief waarbij we het huidige rankingslandschap zullen overschouwen. Het gaat daarbij dus om rangschikkingen die de prestaties -de werking, de efficiëntie, de effectiviteit- van het bestuur in de verschillende landen quoteren en met elkaar vergelijken. We houden het daarbij dus ook op een macroperspectief. In een microbenadering worden specifieke publieke diensten binnen een land onderling vergeleken. Zo bestaan er vergelijkende studies van brandweerdiensten, diensten burgerlijke stand, ziekenhuizen, scholen, gevangenissen, rechtbanken, kinderopvang (bv. (Bouckaert, 1992; Bouckaert, 1993; Blank, 2000)) (Callens, 2007).

Van de Walle waarschuwt wel dat er eerder weinig onderzoek naar de *'bureaucratic quality'* en *'administrative performance*' op een geaggregeerd niveau bestaat. Volgens hem blijft het onderzoek eerder beperkt tot ofwel een meting van de administraties binnen een land ofwel een internationale vergelijking van specifieke administraties (Van de Walle, 2005). Desalniettemin spitsen we ons toe op internationale rangschikkingen van publieke sectoren die een meting uitvoeren van de kwaliteit van de overheid. We leggen deze rankings op de rooster en trachten er de kwaliteit van in te schatten. Soms gaat het daarbij om een beoordeling van een index of een ranking zelf, maar veelal zullen we op zoek moeten gaan naar een deelcomponent van een hoger liggend concept zoals '*competitiveness*', '*governance*' of '*country risk*'.

De meting van de 'goede werking van de overheid' die hier in beeld is, wordt in veel gevallen samengevat onder het label '*government effectiveness*', in andere rankings spreekt men van '*institutional effectiveness*'. In nog andere rangschikkingen van publieke sectoren heeft

men het dan weer over '*government (in)efficiency*', al komt dit niet overeen met de bestuurskundige term 'efficiëntie'. Input zoals personeel of budgetten worden immers doorgaans niet in rekening gebracht. In algemene bewoordingen gesteld, peilen de rankings naar de mate van effectief, goed bestuur in de verschillende landen. Al moet gezegd worden dat de producenten meestal zelf geen sluitende omschrijving kunnen geven voor hun onderzoeksobject. De betekenis van de meting valt paradoxaal genoeg net enkel af te leiden uit de onderliggende indicatoren.

> 1.2.1. Soorten, maten, gewichten

Ter inleiding van het onderzoek naar de internationale rangschikkingen van publieke sectoren, starten we met een korte introductie van de verschillende concepten. Een internationale rangschikking van publieke sectoren -of (governance) ranking- is een vergelijking die hetzij de beleidsprestaties (beleid), hetzij het overheidsapparaat (bestuur) van verschillende landen in kwantitatieve termen met elkaar vergelijkt. Een laddersysteem waarbij onderliggende landen een mindere score behaalden dan bovenliggende landen is in de regel hiervan het eindproduct. De basis voor dergelijke rankings wordt gevormd door indices die gebaseerd zijn op de respectievelijke landenscores. Deze index wordt samengesteld op basis van meerdere indicatoren en tracht een achterliggend, multi-dimensionaal fenomeen te meten. Een voorbeeld hiervan is het indexcijfer van de consumptieprijzen (Van Roosbroek & Van Dooren, 2006). Voor deze index wordt ook wel de Engelse term *'composite indicator*' of gewoonweg (totaal-)score of gebruikt.

De index of de (internationale) rangschikking van publieke sectoren stoelt op één of meerdere indicatoren. Een indicator is een variabele die een verschijnsel kwantitatief meet. De omvang van het BBP, de inschatting van de corruptie door bedrijfsleiders of de gepercipieerde administratieve last als gevolg van overheidsregulering zijn hiervan voorbeelden. Bij het evalueren van de internationale rangschikkingen van publieke sectoren die bijvoorbeeld landen een score toebedelen op het vlak van Government Effectiveness (cf. Worldwide Governance Indicators), is het bijgevolg cruciaal te onderzoeken welke indicatoren door de producent (i.c. het World Bank Institute) geselecteerd werden om dat concept te meten. De indicatoren worden in sommige gevallen nog in deelcomponenten

opgedeeld. Deze componenten vormen onderdelen van het gemeten fenomeen. Zo wordt Government Efficiency door het International Institute for Management Development (IMD) opgevat als een component van het hogerliggende fenomeen Competitiveness.

Bij de meting van abstracte begrippen is de constructie tegelijkertijd cruciaal en problematisch. Welke indicatoren gebruiken we bijvoorbeeld om 'een efficiënte overheid' te meten? Wat is een 'goede administratie' of een 'competitieve markteconomie'? Een rangschikking bouwt verder op het concept van de index. De producenten van internationale rankings meten aan de hand van een reeks uitgekozen variabelen (indicatoren) een concept, uitgedrukt in een index (score). Nadien worden deze indices dan met elkaar vergeleken, wat een rangschikking van goed naar slecht of omgekeerd oplevert. De scores en de plaatsen op de ranglijst zijn echter sterk afhankelijk van de constructie van de meting en dus niet zelden onderwerp van controverse.

Binnen de internationale rangschikkingen van publieke sectoren, zijn er nog verschillende soorten, maten en gewichten te onderscheiden. Zo kan de ranking een momentopname (absolute waarde) quoteren of een tijdsperiode (relatieve waarde) overschouwen. Ondanks het feit dat rangschikkingen vaak de neiging hebben om absolute waarden in overweging nemen, is het vaak interessanter om naar verandering te kijken om de prestaties van het huidige beleid en bestuur in te schatten (Van Roosbroek & Van Dooren, 2006).

De producenten van internationale governance rankings zijn meestal internationale organisaties of adviesbureaus. Een belangrijk aspect hierbij is echter wie de basisdata aanbrengt. Vooral wanneer de data van de nationale overheden zelf moeten komen, duikt er een zeker risico op. Aangezien landen er duidelijk veel belang aan hechten welke plaats ze op de internationale ladder bekleden, bestaat het risico zo immers dat slecht scorende overheden hun statistieken artificieel of valselijk opwaarderen. Dit risico verhoogt verder als de meting als *target* wordt gebruikt. "*When a measure becomes a target, it ceases to be a good measure* (Bevan & Hood, 2006)." Dit is onder andere het geval voor enkele governance-metingen die ingezet worden bij de verdeling van ontwikkelingsgelden.

> *1.2.2. Concepten*

Belangrijk is toch om hier enkele concepten nader toe te lichten. De internationale rangschikkingen van publieke sectoren doen per definitie een meting van de prestaties van de overheid. Een probleem dat bij het evalueren van een goede overheidsrangschikking opduikt, is in eerste instantie de vraag wat een overheid is. Wie behoort tot de overheid en wie valt daarbuiten? Op het eerste gezicht een banale vraag, maar niets is minder waar. De concepten, 'overheid', 'staat', 'regering', '*public sector*', '*administration*' en '*government*' worden immers in diverse landen en door verschillende auteurs en organisaties op verschillende manieren ingevuld. Er bestaat bijvoorbeeld vooralsnog geen unaniem aangenomen, eenduidige wijze om te bepalen wie kan worden meegerekend als overheidspersoneel en wie niet (Van de Walle et al., 2004).

De publieke sector kan volgens Kuhry en Van der Torre (2002) op drie verschillende manieren afgelijnd worden; van erg ruim tot zeer strikt. De ruimste invulling kijkt naar de taak van de organisatie om zo de overheid te onderscheiden van andere instellingen en organisaties. Organisaties of instellingen die het algemene publieke belang dienen, worden daarbij tot de overheid gerekend. Het gaat dan over alle instellingen op de terreinen openbaar bestuur, sociale zekerheid, veiligheid, onderwijs, zorg en maatschappelijke en culturele dienstverlening. Men spreekt in dit verband ook van de 'quartaire sector'. Is de taakstelling daarentegen van private of louter commerciële aard, dan maakt ze deel uit van de private of andere sector.

Een tweede potentiële scheidingslijn betreft het financiële criterium. De hamvraag hierbij luidt: wordt de opdracht betaald door belastingen? Indien de kosten ingeschreven staan in de overheidsbegroting, kan de organisatie tot de publieke sector gerekend worden. In het andere geval valt de organisatie daarbuiten.

Het derde, en meest restrictieve criterium is het juridische statuut. Vanuit dit oogpunt, worden alle organisaties met publiekrechtelijk statuut tot de overheid gerekend. Alle privaatrechtelijke organisaties vallen dan buiten de publieke sector.

Een andere veelgebruikt criterium om uit te maken wat wel en wat niet tot de overheid hoort, is de opsplitsing uit het bekende Europees Stelsel van Rekeningen van 1995 (ESR95). In dit document vindt men onder titel S.13 een beschrijving terug van de overheid. Voor de centrale overheid, de deelstaatoverheden, de lagere overheden en de wettelijke-sociale-verzekeringsinstellingen geldt: "De sector overheid (S.13) omvat alle institutionele eenheden die tot de 'overige niet-marktproducenten' behoren en waarvan de output bestemd is voor individueel of collectief verbruik en die voornamelijk gefinancierd wordt uit verplichte betalingen van eenheden die tot andere sectoren behoren en/of alle institutionele eenheden die zich in hoofdzaak bezighouden met de herverdeling van het nationale inkomen en vermogen (ESR95; §2.68)."

Eens vastgelegd wat de publieke sector of een overheid juist inhoudt, is het zoeken naar een goede omschrijving van de '*governance*', die in de governance rankings -de internationale rangschikkingen van publieke sectoren- onderzocht wordt. De eerste concrete geoperationaliseerde definitie van '*governance*' dateert van 1991. Landell-Mils en Serageldin duidden het concept aan de hand van zes minimumkarakteristieken; *political accountability, freedom of associations, sound judicial system, bureaucratic accountability, freedom of information and expression en capacity building* (Tiihonen, 2004).

Een recentere en vaak opgevoerde definitie komt van Besançon (2003). Zij stelt dat governance overeenstemt met '*the delivery of political goods to citizens of nation-states*'. De World Bank (Kauffman et al., 2000) definieert governance dan weer als '"*the traditions and institutions that determine how authority is exercised in a particular country, the process by which governments are selected, held accountable, and replaced, the capacity of governments to manage resources efficiently and formulate, implement, and enforce sound policies and regulations, and the respect of citizens and the state for institutions that govern economic and social interactions among them".* De vaagheid en omvang van deze omschrijvingen geeft passend aan hoe moeilijk het is om dergelijke concepten te definiëren. Rothstein en Teorell gaan nog verder in hun kritiek en stellen dat dergelijke definities omwille van hun vaagheid een concept onbruikbaar maken. "*If governance is everything, than maybe it is nothing*" (Rothstein & Teorell, 2005).

'*Good governance'* is een concept dat in het verlengde van '*governance*' ligt. Waar '*governance'* de brede term van bestuur behelst, is *'good governance*' het streefdoel dat gepredikt wordt. Volgens Besançon is er sprake van bij '*nation-states that perform effectively and well on behalf of their inhabitants'*. Hoe dit moet vertaald worden in deelaspecten en vervolgens indicatoren, is een moeilijke kwestie waar geen eensgezindheid over bestaat. Besançon maakt een oplijsting van 10 punten waar een goed bestuurd land hoog moet op scoren, de World Bank houdt het op zes indices (die overeenkomsten vertonen met de definitie van Landell-Mills en Serageldin). We lijsten deze twee suggesties hieronder op (zie Kader 1 en Kader 2).

Kader 1: Kenmerken van een *'well governed country*' (Besançon, 2003)

- *Rule of law*
- P*olitical and civil freedoms*
- *Medical and health care*
- *Schools and educational instruction*
- *Roads, railways, the arteries of commerce*
- *Communications networks*
- *A money and banking system*
- *A fiscal and institutional context within which citizens can prosper*
- *Support for civil society*
- *A method of regulating the sharing of the environmental commons*

Kader 2: Kenmerken van een '*well governed country*' (Kauffman & Kraay, 2007b)

- *Voice and accountability*
- *Political stability*
- *Government effectiveness*
- *Regulatory quality*
- *Rule of law*
- *Control of corruption*

Kort samengevat doen internationale rangschikkingen van publieke sectoren een meting van de prestaties van (een aspect van) de overheid. Ze meten in brede of enge zin de kwaliteit van de overheid. In onze studie leggen we ons specifiek toe op de internationale rangschikkingen van bestuur die een meting uitvoeren van de kwaliteit van de overheid in zijn geheel. Er wordt dus gekeken naar rankings en indicatoren die een meting uitvoeren van een aspect waar de hele overheid -in positieve of negatieve zin- verantwoordelijk kan gesteld worden.

De validiteit is echter een van de grote pijnpunten van de governance rankings. Producenten schieten steevast tekort als het aankomt op de conceptualisering van hun overheidsmetingen. De internationale metingen van de overheidskwaliteit verschijnen bij verschillende producenten onder verschillende noemers. Waar het World Bank Institute in haar Worldwide Governance Indicators spreekt van 'Government Effectiveness', heeft het World Economic Forum het in haar World Competitiveness Index over 'Government Inefficiency' en de European Central Bank in haar Public Sector Performance over de 'Administrative Performance'. De differentiële *labeling* niettegenstaande, doelen ze echter allen de kwaliteit van de overheid vanuit een bepaald perspectief te meten.

Moeilijk wordt het al helemaal wanneer de verschillende producenten van governance rankings op enkele uitzonderingen na geen definitie of duidelijke omschrijving meegeven van hun gemeten concepten. Landen krijgen wel een score toebedeeld, maar het is niet zelden onduidelijk in te schatten waarop ze nu eigenlijk gequoteerd werden. Op die manier ontstaat een contradictorische situatie waarin de betekenis van de meting dient afgeleid te worden uit de gebruikte indicatoren.

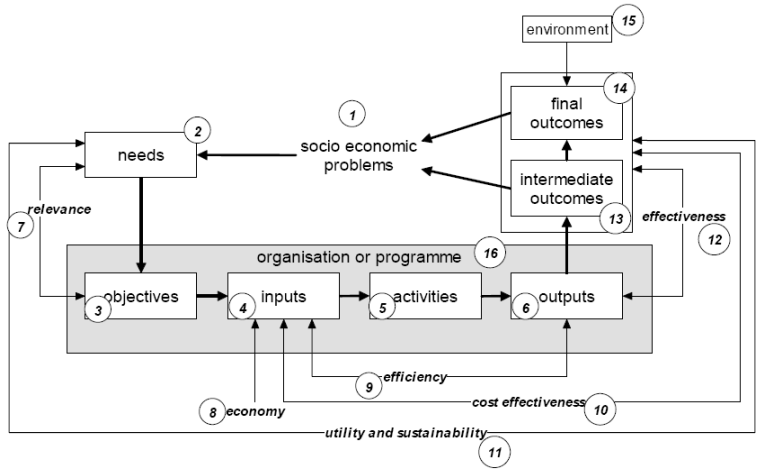
> *1.2.3. Hoe overheidskwaliteit meten?*

Internationale rangschikkingen van publieke sectoren maken een relatieve evaluatie op van de prestaties van nationale overheden. Dergelijke metingen zijn echter een uiterst complexe uitoefening. Een 'goed werkende overheid', 'een goed bestuurd land', 'een kwaliteitsvolle publieke sector'; er bestaan geen 'natuurlijke' standaarden om deze concepten aan af te toetsen. Bovendien zijn er meestal geen prijzen vast te knopen aan de diensten en prestaties van de overheid, zoals dat wel het geval is in de

marktsector (O'Mahony & Stevens, 2002). De gekwantificeerde prestaties van nationale overheden gelden als vergelijkingspunt. Het gaat daarbij om verschillen in productiviteit, kwaliteit, doeltreffendheid of doelmatigheid van de door de overheid georganiseerde dienstverlening (Sociaal en Cultureel Planbureau, 2007).

De geleverde prestaties in de publieke sector kunnen bekeken worden vanuit de ingezette middelen (input), het proces (process), het product (output) of het effect (outcome) (Callens, 2007; Bouckaert & Auwers, 1999; Van dooren, 2006). Men kan vanuit verschillende perspectieven naar de overheidsprestaties kijken. Men kan de productiviteit, de efficiëntie of de effectiviteit van de overheid in beschouwing nemen (Kuhry, 2004). Het *Input Output Model* van Pollitt en Bouckaert (2004a) zet deze perspectieven in schema (zie figuur 1). Vooreerst kunnen governance rankings een beoordeling maken van de productiviteit van de overheid. In economische zin is de productiviteit het volume aan prestaties, dat geleverd wordt per eenheid input. Arbeidsproductiviteit slaat terug op de prestaties per medewerkers. Als alle productiefactoren in overweging genomen worden, spreekt men van *Total Factor Productivity*. In de bestuurspraktijk in de overheid is productiviteit vaak een synoniem voor efficiëntie.

Figuur 1: Input Output Model



(Pollitt & Bouckaert, 2004a)

Efficiëntie of doelmatigheid wordt afgemeten aan de hand van het aantal middelen om een bepaald doel te bereiken. Zoals het in de vierde pijler van Vlaanderen in Actie wordt samengevat: "Een overheid die de doelstelling realiseert van een beter bestuur met proportioneel minder middelen" (Vlaamse regering, 2006). De directeur van de Nationale Bank van België, De Batselier (2007), verwoordt het als volgt: "Efficiëntie komt neer op het boeken van de beste resultaten tegen de kleinst mogelijke kosten." Deze stelling is inaccuraat. De resultaten kunnen immers zowel output (producten en diensten) betreffen, als de effecten van deze producten en diensten. In de overheid gaat het vooral om effecten, en is effectiviteit het ultieme criterium.

Omdat de diensten die door de publieke sector worden geproduceerd doorgaans niet op de vrije markt worden verhandeld, is hun marktwaarde meestal niet bekend. De waarde van de productie kan dan niet in geld worden uitgedrukt. Daarom wordt in de meeste gevallen gebruik gemaakt van fysieke productindicatoren. Dit is een verzamelterm, die betrekking heeft op verschillende typen indicatoren, die als directe of indirecte maat voor de productie kunnen gelden. Hierbij kan een onderscheid worden gemaakt tussen:

1. Prestatie-indicatoren, die te maken hebben met de geleverde eindproducten,
2. gebruiksindicatoren, die te maken hebben met de afnemers van de diensten,
3. procesindicatoren, die te maken hebben met verrichte werkzaamheden of tussenproducten.

Prestatie-indicatoren hebben betrekking op het eindproduct van de dienstverlening. Als zodanig zijn zij het meest geschikt om de doelmatigheid van de dienstverlening te toetsen. Voorbeelden van prestatie-indicatoren zijn het aantal succesvol behandelde patiënten in een ziekenhuis, het aantal behaalde diploma's in het onderwijs of het aantal uitvoeringen bij concerten.

Gebruiksindicatoren hebben niet zozeer betrekking op de eigenlijke productie, dan wel op het aantal afnemers van producten. Daarmee zijn ze bij uitstek geschikt voor analyses en ramingen van de vraag naar diensten. Voorbeelden van gebruiksindicatoren zijn het aantal patiënten in

ziekenhuizen, het aantal leerlingen in het onderwijs en het aantal toehoorders bij concerten. Gebruiksindicatoren zijn in veel gevallen redelijk tot goed bruikbaar als maatstaf voor de productie. Als het aantal opnamen in ziekenhuizen of het aantal leerlingen in het onderwijs als maat voor de productie wordt gekozen, wordt impliciet verondersteld dat de genezingskans of het slaagpercentage constant is.

Procesindicatoren kunnen betrekking hebben op verrichte taken en daarmee een maat zijn voor de inspanningen van het betrokken personeel. Voorbeelden van procesindicatoren zijn het aantal verrichtingen in ziekenhuizen of het aantal gegeven lessen in het onderwijs. In sommige gevallen kunnen procesindicatoren gebruikt worden als indicator voor de productie. In het geval van de thuiszorg, bijvoorbeeld, kan de productie worden gemeten via het aantal contacturen van het personeel. De impliciete veronderstelling is dan dat het personeel per contactuur een constante hoeveelheid werk verzet. Er rijzen echter enkele problemen bij het meten van efficiëntie. Sommige producten of diensten lenen zich veel minder gemakkelijk tot het nagaan van de efficiëntie.

Een eerste probleem betreft de collectieve goederen. Bij zuiver collectieve goederen of diensten is vaak geen goede indicator voor de geleverde prestaties te construeren. Het is bij voorbeeld bijzonder moeilijk voor voorzieningen als openbaar bestuur en defensie aan te geven wat de geleverde prestatie is. Laat staan dat dit voor dé overheid gemakkelijk is.

Een tweede probleem is de heterogeniteit die de meting van de productie in veel gevallen bemoeilijkt. Hiervan is sprake als een bepaalde producent meerdere producten of producttypen voortbrengt. Zo kan in de zorg onderscheid worden gemaakt naar zorgzwaarte of diagnosecategorie van patiënten, bij de politie naar soorten misdrijf, of bij het onderwijs naar doelgroepen. Het is dan noodzakelijk om de productie op één noemer te brengen om een vergelijking met de ingezette middelen mogelijk te maken. In de marktsector is dat geen probleem omdat de afzonderlijke producten kunnen worden gewogen met hun waarde, de marktprijs. Bij de meeste quartaire diensten is dat niet mogelijk en moeten de afzonderlijke productcategorieën gewogen worden opgeteld met behulp van de ingezette middelen (bij voorkeur integrale kosten, soms -bij gebrek aan beter- de personeelsinzet per product).

Ten derde is er het probleem van de kwaliteit. Een van de belangrijkste problemen bij het meten van de prestaties van publieke dienstverlening is dat de kwaliteit van de producten in onvoldoende mate wordt gereflecteerd door de meting van op kwantiteiten gerichte productindicatoren. Daarom moet er op zijn minst een kwaliteitsaanpassing komen van gebruikte kwantitatieve indicatoren (O'Mahony & Stevens, 2002). Kwaliteit is een vaag en complex begrip dat te maken heeft met de kenmerken van het product in relatie tot de gestelde eisen. Het is zinvol om een onderscheid te maken tussen:

1. objectieve en subjectieve kwaliteitsmaten enerzijds
2. productkwaliteit, proceskwaliteit, en systeemkwaliteit anderzijds.

Bij objectieve maten voor de kwaliteit van diensten kan bijvoorbeeld worden gedacht aan het percentage treinen dat op tijd rijdt, de gemiddelde uitruktijd van de brandweer, het percentage uitgevallen lessen, het percentage foutieve beschikkingen of het percentage (gehonoreerde) klachten.

Veel relevante aspecten van kwaliteit zijn echter moeilijk meetbaar, zoals de bejegening van patiënten of de correctheid van medische handelingen. Aanvullende informatie kan worden verkregen via subjectieve kwaliteitsoordelen over het product of over het productieproces. Dergelijke oordelen kunnen worden gevraagd aan de gebruiker, een toezichthouder (inspectie) of het personeel.

Naast de productiviteit en de efficiëntie, kan ook de effectiviteit of doeltreffendheid van de publieke sector gemeten of ingeschat worden. Hierbij worden de geleverde prestaties afgezet ten opzichte van de gerealiseerde doelstellingen. De relatie van productieprocessen met effecten (outcome) is vaak minder direct dan die met output. Het is in dit verband zinvol om een onderscheid te maken tussen directe, via het eindproduct meetbare, doelstellingen en diepere achterliggende maatschappelijke doelstellingen.

Als directe doelstelling van het onderwijs zou bijvoorbeeld het behalen van het einddiploma kunnen gelden, als die van de curatieve zorg het succesvol voltooien van de behandeling en als die van de politie het ophelderen van misdrijven. Een productindicator is in de regel bruikbaarder, naarmate hij

beter het directe doel van het productieproces weergeeft. Bij ziekenhuizen, bijvoorbeeld, blijkt dat analyses vaak aan scherpte verliezen doordat opnames, verpleegdagen of verrichtingen als indicator fungeren in plaats van succesvolle behandelingen. Hetzelfde geldt voor analyses van het onderwijs, waarbij de aantallen leerlingen in plaats van hun schoolresultaten als productindicator worden gehanteerd. Voorbeelden van diepere achterliggende doelstellingen zijn het bevorderen van de mondigheid van burgers en het zorg dragen voor een goed opgeleid arbeidsaanbod (onderwijs), bevordering van een lang en gezond leven (zorg) en bestrijding van de criminaliteit (politie).

Bovendien is het ook interessant om de invalshoek van de producent van de ranking onder ogen te zien. De visie op de voorname taken van de overheid speelt hierbij een belangrijke rol. Zo kiest de World Bank er in haar Worldwide Governance Indicators eerder voor een politieke inslag te geven aan de meting, wat zich reflecteert in de aard van de indicatoren. Een visie die de Europese Unie deelt (Tiihonen, 2004). Het World Economic Forum meet de prestaties van de overheid in haar World Competitiveness Index dan weer eerder vanuit een economische invalshoek. Ook het International Institute for Management Development (World Competitiveness Yearbook) opteert voor een economische inslag.

> *1.2.4. Waarom publieke sectoren rangschikken?*

Internationale rangschikkingen van publieke sectoren baseren zich op nationale prestatie-indicatoren binnen de overheid, die nadien met verschillende landen vergeleken worden. Het gebruik van dergelijke performantie-indicatoren kennen een opwaartse trend sinds de opkomst van de *New Public Management*denkrichting. Efficiëntie en effectiviteit werden belangrijke begrippen in de hervorming van de publieke sector.

Nieuw zijn internationale vergelijkingen van publieke sectoren echter niet. Volgens Hood (2007) is het gebruik van prestatie-indicatoren in de overheid op zijn minst terug te brengen tot het einde van de 18e eeuw. Filosoof en sociaal hervormer Jeremy Bentham lanceerde toen het idee van wat nu *performance accounting* genoemd zou worden, in combinatie met het '*tabular-comparison*'-principe (soort van *league tables*). Internationale rangschikkingen van publieke sectoren bestaan ook al lang, zij het in een

andere, meer rudimentaire vorm dan de rankings van vandaag. Zo zijn er bewijzen van internationale rangschikkingen op het vlak van criminaliteit die op het Algemeen Statistisch Congres van 1853 in Brussel werden gepresenteerd.

In de twintigste eeuw werden de governance rankings voornamelijk geproduceerd door en voor private instellingen en commerciële bedrijven die de investerings- en leenrisico's in verschillende landen in kaart wilden brengen. Zij werden als producent gevolgd door academici die -vanuit dezelfde economische invalshoek- dergelijke rankings en data gebruikten in hun analyses over economische groei.

De Freedom House Ratings, die landen quoteren op het vlak van politieke en burgerlijke vrijheden, worden algemeen beschouwd als de eerste governance indicatoren, die niet in de eerste plaats bedoeld waren om investeringsbeslissingen mee te sturen. Freedom House publiceerde haar eerste data in 1972 (Van de Walle, 2005).

Het vergelijken van prestaties in internationaal verband kent de laatste twee decennia een groeiende belangstelling. Naast de bedrijven die de rankings gebruikten voor investeringsbeslissingen, kwam er ook vanuit andere hoeken interesse voor internationale rangschikkingen van publieke sectoren. Zo ontstond er ten eerste interesse in de ontwikkelingswereld voor de kwaliteit van de overheid. Goed bestuur/*good governance* werd een modeterm in de ontwikkelingsleer. Donoren en internationale organisaties zagen het belang in van een goed overheidsbestuur, dat efficiëntie hoog in het vaandel draagt en corruptie afzweert, in de ontwikkeling van ontwikkelingslanden. Financiële steun bieden aan landen met een slechte overheid als basis, werd gezien als nefast. Zo begon men *good governance* meer en meer als een voorwaarde voor, in plaats van een doelstelling van ontwikkelingshulp te zien. Dit heeft zijn weerklank gekregen in de zogenaamde *'performance-based aid allocation*'. Onder dit principe krijgen ontwikkelingslanden die goed (beter) scoren op de governance rankings een groter deel van de ontwikkelingssteun, aangezien zij het beter zullen besteden. Dit systeem ligt de laatste jaren echter fel onder vuur en heeft een heel deel van haar aanhangers verloren.

Daarnaast kwamen internationale rangschikkingen van publieke sectoren ook meer en meer onder de aandacht van economici. Sinds zij begrepen hoe

belangrijk de overheidsefficiëntie is in het verklaren van een goed werkende economie, sprongen ook zij op de kar van de internationale metingen van de overheidskwaliteit. Deze gedachtegang vond pas echt ingang in de economische wereld begin jaren '90 (Brutsaert, 2006b). Economische indicatoren bestonden er reeds in voldoende mate, de verklarende factor - een efficiënte overheid - werd van dan af overgenomen uit de verschillende internationale rangschikkingen van publiek sectoren. Moesen en Cousserier (2006) stellen in hun beschrijvend model dat de economische prestaties van een land afhangen van arbeid, kapitaal (fysiek en menselijk), technologie en enkele residuele factoren. Van deze restfactoren is '*Institutes*' de belangrijkste factor. Niet alleen economische macro-indicatoren kunnen mede verklaard worden vanuit de kwaliteit van de overheid. Ook fenomenen op individueel niveau zoals subjectief geluk (Frey & Stutzer, 2000; Heliwell, 2003) en steun en vertrouwen in de overheid bij de bevolking (Andersen & Tverdova, 2003) worden positief beïnvloed door een goede overheidskwaliteit. Op het niveau van de gemeenschap werd een negatief verband gevonden met het voorkomen van burgeroorlogen (Fearon & Laitin, 2003) en een positieve samenhang met democratische consolidatie (Rose & Shin, 2001; Zakaria, 2003).

De motieven achter het vergelijken van publieke sectoren op internationale schaal zijn divers. Ten eerste kan er uit dergelijke rangschikkingen opgemaakt worden in welke mate verschillende staatssystemen bijdragen tot een goed presterende publieke sector (O'Mahony & Stevens, 2004). Aan de hand van comparatieve studies kan onderzocht worden of er een correlatie bestaat tussen een staatssysteem van een land en diens plaats op de governance rankings.

Ten tweede genereren internationale rangschikkingen van publieke sectoren leereffecten. De vergelijking van publieke sectoren biedt beleidsmakers de mogelijkheid om te leren van beter presterende overheden. Het stelt ook een standaard voor verbetering, waaraan ambitieuze landen zich kunnen spiegelen (Besançon, 2003). Op deze manier zorgt een dergelijke vergelijking onrechtstreeks tot efficiëntiewinst. Daar waar in de privésector de marktwerking deze taak op zich neemt, kan internationaal vergelijkend materiaal in de overheidssector tot betere prestaties leiden (Callens, 2007).

Ten derde -en aansluitend bij het voorgaand motief- zijn governance data een modaliteit van verantwoording voor nationale overheden. *Monitoring* van de eigen prestaties vormt zo een deel van de *accountability*-taak die de overheid op zich moet nemen. Door de eigen performantie in een internationaal kader te presenteren, kan een overheid aantonen of het in relatieve termen goed bezig is. Niet enkel geldt dit voor de interne verantwoording doordat overheden en regeringen ten opzichte van hun burgers kunnen aantonen dat ze goed werk leveren. Ook extern geldt deze *accountability*, en dan vooral bij ontwikkelingshulp. Donoren kunnen op deze manier nagaan welk (in)direct effect het toekennen van financiële middelen in arme landen heeft gehad.

Ten vierde leveren internationale rangschikkingen van publieke sectoren het materiaal voor academische analyses. Zo is de conclusie van de auteurs van de Worldwide Governance Indicators dat *weak governance* trage economische groei in de hand werkt, terwijl verbeterde *governance* ontwikkeling bevordert (Kauffman et al., 2000). Ondanks het feit dat deze conclusie door sommige auteurs wordt tegengesproken, is het minste dat je kunt zeggen wel dat de internationale vergelijkingen van publieke sectoren statistische input genereren die van pas komt in dergelijke discussies. Daarnaast kan ook een grondige analyse op basis van harde data uitgevoerd worden van de determinanten en de impact van *good governance* op de ontwikkeling van een land.

> 1.3. Ranking World

Internationale rangschikkingen van publieke sectoren bestaan dan wel al even in basisvorm, hun opgang heeft vooral de laatste 20 tot 30 jaar een enorme boost ervaren. Zwaarbeladen begrippen zoals 'goed bestuur', *'good governance*' en 'overheidsefficiënte' zijn hoog op de nationale en internationale agenda komen te staan. Dit heeft aanleiding gegeven tot een forse toename van het aantal governance indicators en andere prestatie-indicatoren die activiteiten binnen de publieke sector meten. Deze indicatoren hebben de laatste decennia ook meer en meer ingang gevonden in internationale rangschikkingen van publieke sectoren, die het onderwerp van deze studie uitmaken. In die mate zelfs dat sommigen de toegenomen productie van internationale rangschikkingen van publieke sectoren omschrijven als een '*mushrooming industry*' (Van de Walle, 2006a). In wat volgt geven we een overzicht van de drijvende krachten achter deze evolutie.

Midden 20e eeuw ontstond er een eerste belangstellingsgolf voor internationale governance rankings, zij het bij slechts een beperkte doelgroep. Initieel werden dergelijke data immers enkel geproduceerd voor investeringsdoeleinden. Bedrijven of privépersonen die financiële middelen in het buitenland investeerden, geld leenden aan ontwikkelingslanden of economische activiteiten op vreemd grondgebied ontplooiden, hadden nood aan harde cijfers om hun beslissingen daaromtrent vorm te geven. De invalshoek was bijgevolg voornamelijk economisch en financieel getint, rangschikkingen van publieke sectoren hielden er een minimalistische opvatting van de overheid op na.

In het verlengde van deze risicoanalyses, die voornamelijk door commerciële organisaties of private adviesbureaus werden uitgevoerd, kreeg ook de economische onderzoekswereld oog voor de rangschikkingen van publieke sectoren. Governance data werden in dat geval gebruikt bij economische analyses en verklarende modellen. Deze aandacht kaderde in de zoektocht van economici naar verklarende factoren van economische welvaart en ontwikkeling.

De laatste decennia hebben de internationale rangschikkingen van publieke sectoren echter een ander dimensie gekregen. Ze zijn gesofistikeerder geworden, verwierven een continu karakter en worden nu vanuit

verschillende oogpunten geproduceerd. Het aanbod is ook fors gestegen. Aan de basis van de explosieve groei van het aantal governance rankings ligt een toegenomen interesse voor de '*quality of governance*' in diverse domeinen.

De eerste internationale rangschikking van onderwijsprestaties dateert van 1964 (International Association for the Evaluation of Educational Achievement), de eerste internationale ranking op het vlak van criminaliteit van 1970 (United Nations) en de eerste internationale competitiviteitsranking van 1979 (World Economic Forum en International Institute for Management Development) (Hood et al., 2007). Besançon (2003) berekende onlangs dat sinds 1970 het aantal governance rankings elk decennia verdubbeld is. Vooral de laatste 15 à 20 jaar is er een forse groei van het aantal internationale rangschikkingen van publieke sectoren waar te nemen. In 2003 telde Besançon maar liefst 49 verschillende governance rankings. Als dit tempo aanhoudt, bereiken we tegen 2030 een onoverzichtelijk governance rankingslandschap van 300 datasets (Besançon, 2003). Hood (2007) merkt wel op dat het aantal indicatoren van de publieke sector niet evenredig is gestegen met het aantal governance rankings. Met andere woorden, producenten van governance rankings gebruiken vaak dezelfde ingrediënten, maar dan wel elk vanuit hun eigen recept.

Aan de basis van deze groei kunnen we twee factoren onderscheiden. Enerzijds is de idee van een goede presterende overheid en een goed werkend bestuur door ontwikkelingsspecialisten, en in hun zog de internationale gemeenschap, opgepikt als één van de sleutelfactoren in het ontwikkelingsproces in de derde wereld (Arndt & Oman, 2006). Anderzijds is ook de interesse gestegen voor de overheidskwaliteit, en de gerelateerde data en rangschikkingen van de Westerse landen zelf.

De ontwikkelingsleer nam de gedachte uit de economische analyses over dat *good governance* een basis vormt voor economische groei en ontwikkeling (Public Governance and Territorial Development Directorate OECD, 2005). Deze idee werd snel opgepikt door de World Bank en verschillende voortrekkende nationale overheden. Zij gingen hun ontwikkelingshulp vanuit deze visie koppelen aan governance indicators. De redenering hier achter is dat ontwikkelingslanden financiële steun enkel goed zullen besteden, als er goed bestuurd wordt (Schmitz, 2007). '*Good governance*' nam de plaats in van behoefte (armoedeniveau) als uitgangspunt bij de verdeling van

ontwikkelingshulp. De mosterd voor deze theorie werd gehaald bij het werk van twee hooggeplaatste actoren binnen de World Bank, David Dollar en Paul Collier: *'Assessing Aid: What works, what doesn't, and why?*'. Derde Wereldlanden dienden eerst schoonmaak te houden in eigen bestuur, vooraleer financiële hulp verdere ontwikkeling zou kunnen steunen. Zoniet zou de ontwikkelingshulp enkel helpen om armoede en achterstelling van bepaalde bevolkingslagen in stand te houden. Het belang van de governance indicators en gerelateerde internationale rangschikkingen van publieke sectoren steeg daardoor gevoelig.

Bovendien vond men in het post-Koude Oorlog tijdperk in de governance indicators een nieuw verdeelmiddel van ontwikkelingshulp (Besançon, 2003). Daar waar de alliantie met het Oost- of Westkamp het criterium was voor het ontvangen ontwikkelingshulp gedurende de Koude Oorlog, viel dit allocatiesysteem na de val van de Sovjet-Unie zo goed als helemaal weg. Tijdens de Koude Oorlog kwam de kwaliteit van het bestuur niet op de onderhandelingstafel bij het verdelen van de financiële steun aan de Derde Wereld. Met het wegvallen van de spanning tussen de Verenigde Staten en de Sovjet-Unie en meer specifiek met het aantreden van een nieuwe directeur bij de World Bank, James Wolfensohn, draaide dit beleid 180°. Het idee van '*lending to improve governance*' kreeg vorm en de World Bank ging zich van dan af wel inmengen met de interne politiek van corruptie en slecht bestuur in de arme landen. Zo steeg het aantal World Bankleningen om goed bestuur te versterken tussen 2000 en 2004 met 11%, World Bankleningen die economische hervormingen aanstuurden zelfs met 14% (Arndt, 2007).

Ook de Verenigde Staten hebben deze theorie een steun in de rug gegeven. Met de invoering van de *Millennium Challenge Account* in 2004 ging president George W. Bush over tot de conditionaliteit van de Amerikaanse ontwikkelingshulp. Maar liefst 5 miljard dollar aan financiële steun voor de Derde Wereld werd verbonden aan de prestaties op 16 Governance Indicators (Arndt & Oman, 2006).

Dit conditionaliteitsidee was gedurende lange tijd *common rule* voor grote financiële instellingen, maar heeft de laatste jaren meer en meer te lijden onder kritieken van experten, ontwikkelingsorganisaties en zelfs producenten van governance rankings zelf (Arndt & Oman, 2006). Zij vragen

zich af of slecht bestuurde landen niet juist geholpen moeten worden in plaats van gestraft.

De overheidskwaliteit in ontwikkelingslanden werd ook om opportunistische, financiële redenen van nabijer opgevolgd. De *Foreign Direct Investments* vervijftienvoudigden in de laatste twee decennia van het vorige millennium van 10 miljard dollar in 1980 tot 150 miljard dollar in 1997. Daarnaast stegen de portfolio-investeringen in diezelfde periode van 2 miljard dollar naar 26 miljard dollar. De stijging van internationale investeringen in ontwikkelingslanden, heeft zich bijgevolg vertaald in meer interesse voor de kwaliteit van de overheid in die landen. De overheidskwaliteit wordt immers als een belangrijke factor beschouwd van het algemene risico waar het geïnvesteerde kapitaal in het buitenland aan bloot gesteld wordt.

Naast de verhoogde interesse voor de kwaliteit van de overheid in minder ontwikkelde landen, groeide ook de belangstelling voor de overheidsprestaties van Westerse landen zelf. Deze bewustwording is 10 à 15 jaar geleden ingezet. Het stijgende belang van het meten van prestaties past in een algemene trend van modernisering van het overheidsapparaat. Budgettaire problemen en het mondiger worden van de burger zorgden immers voor een toenemende druk in de richting van effectiviteit en efficiëntie (Pollitt & Bouckaert, 2004b). Ook de schandalen die in de jaren ’80 en ’90 de politiek beroerden en het dalende vertrouwen van de burger in de overheidsinstellingen leidden tot het opwerpen van het efficiëntie- en effectiviteitsvraagstuk (Van de Walle et al., 2004). Meer en meer worstelen we immers met de vraag of onze publieke diensten wel goed genoeg functioneren. Om dit in kaart te brengen, moet er vergeleken worden (Van Roosbroek & Van Dooren, 2006). Ook in de huidige tijden van toenemende druk op het ambtenarenkorps, is de vraag naar de effectiviteit en efficiëntie een zinvolle oefening. Zoals Vlaams Minister-president Kris Peeters het onlangs in zijn Septemberverklaring nog verwoordde: "De Vlamingen willen een krachtig bestuur. Daartoe is een toegankelijke, voorspelbare en efficiënte overheid nodig en een klantvriendelijke dienstverlening (Peeters, 2007)."

De goede werking van de overheid wordt volgens econoom Bowen (2006b) vanuit de economische wereld met grote interesse gevolgd sinds ze daar begrepen hoe belangrijk deze randvoorwaarde is voor een goed werkende economie. Een goede illustratie hiervan dichtbij huis vindt men terug in het

actieplan van het Verbond van Belgische Ondernemingen 'Strategie 2010'. Hierin wordt het belang van de kwaliteit van de overheid in het creëren van een goed ondernemersklimaat, in één van de 9 actiepunten nadrukkelijk erkend (Brutsaert, 2006a).

Dit voornamelijk economische gedachtegoed kent zijn wortels in de mislukte beleidshervormingen van de jaren '80 en '90. De World Bank kwam tot de conclusie dat haar manier van werken niet de gewenste resultaten opleverde in termen van beleidsverandering, goed bestuur en ontwikkeling. De zogenaamde '*Washington Consensus*' was achterhaald en werd begraven. In plaats daarvan ontstond het devies: '*Strong markets require good governance*' (Arndt & Oman, 2006).

Bovendien mag ook de invloed van Douglas North en *the New Institutional Economics* in deze hele ontwikkeling niet onderschat worden. Deze theoretische denkrichting benadrukte voor het eerst het belang van het governance-systeem van een land –de formele en informele instellingen en hun interactie met het gedrag van economische en politieke ondernemers- in het succes van dat land in termen van economische langetermijn groei, welvaart en de maatschappelijke vooruitgang. De *Tournament Theory* deed op die manier zijn intrede in de publieke sector (Hood et al., 2007). *Performance* werd het sleutelwoord voor ambtenaren en organisaties, zowel op nationaal als op internationaal niveau.

> 1.4. Governance Rankings: Gehypt of beleidsrelevant?

Hierboven hebben we aangetoond dat internationale rangschikkingen van publieke sectoren de laatste decennia in toenemende belangstelling kwamen te staan. De impressie van nieuwigheid en de retorische kracht van cijfers, heeft volgens Hood (2007) prestatie-indicatoren ingang doen vinden in het hedendaagse overheidsmanagement. Doorheen de jaren zijn veel beleidsmakers, ontwikkelingsspecialisten en economen het erover eens geworden dat goed bestuur een vertrekpunt vormt voor een goed werkende overheid, de ontwikkeling van een natie en een groeiende economie. In de hoger besproken paragraaf over het waarom van het meten van de publieke sectoren (cf. 1.2.4) hebben we aangegeven voor welke doeleinden de internationale rangschikkingen van publieke sectoren gebruikt worden. In grote lijnen gaven we daar een overzicht van de mogelijkheden die voorstanders van overheidsmetingen vergelijkingsoefeningen toedichten. Hier zullen we enkele twijfels over het gebruik van governance rankings aan bod laten komen.

De governance rankings volgen de wetten van *supply and demand*, en zolang de vraag groot blijft, zal het grote en diverse aanbod dan ook blijven (Hood et al., 2007). Ook de media springen graag mee op de trein van internationale vergelijkingen. Rankings laten gemakkelijker toe om in een kort bestek simpele vergelijkingen te maken en zwaarwichtige headlines te formuleren. Zo titelde De Tijd in 2003 als gevolg van een uitgebrachte internationale studie nuanceloos: 'Belgische overheid is zeer inefficiënt' (Vervenne). De schijnbaar simplistische en objectieve charme die de ranking zo populair maakt, is echter tevens haar grootste gevaar. De media halen vaak selectief enkele opmerkelijke stukken uit dergelijke grootschalige vergelijkingsoefeningen en verliezen daardoor de nuance van het bredere onderzoeksperspectief. Niet zelden verschijnen in de media zelfs conclusies uit governance rankings, die wel erg bij de haren gegrepen zijn. Zoals dit onderzoek verder duidelijk zal maken, zijn de metingen bovendien niet onverdeeld kwalitatief hoogstaand. Er wordt niet altijd gemeten wat beweerd wordt te meten. En ook de data zelf blijken meer dan eens onbetrouwbaar of onderhevig aan significante bias.

Het gevolg is dat zeker niet iedereen er van overtuigd is dat rangschikkingen van publieke sectoren wel nuttig kunnen zijn voor het beleid van een land. De vraag werpt zich op of governance rankings niet te veel *gehypt* worden

door belanghebbende politici en de media. Ons standpunt -zoals verder zal blijken- is er een tegen de visie dat het meten van publieke sectoren irrelevant is. Er dienen echter wel genuanceerde conclusies verbonden te worden met dergelijke data. Door het blootleggen van de structuur, de bedoelingen en de technische kwaliteit van de rankings, trachten we hiertoe een bijdrage te leveren. Niet de cijfers, maar de mechaniek achter de meting moet inzicht bieden in de betrouwbaarheid van de rankings en in de relevantie van de data.

De relevantie van de internationale rangschikkingen inschatten, is echter geen gemakkelijke klus. De mate van relevantie is immers sterk afhankelijk van de doeleinden die producenten en gebruikers de rankings toedichten. Voor economici moeten de data zo opgebouwd zijn dat ze kunnen gebruikt worden in de analytische modellen van economische groei. 'Goed bestuur' moet dan ook als individuele variabele kunnen gemeten worden, los van andere (onafhankelijke) factoren. Voor de ontwikkelingsexperten en donororganisaties moeten de internationale rangschikkingen op een betrouwbare manier kunnen aangeven welke landen over een vruchtbare overheidsbodem beschikken om ontwikkelingshulp te laten renderen en welke niet. Aangezien aan de governance data vaak grote sommen financiële ontwikkelingssteun worden gekoppeld, is de bruikbaarheidsvraag zeker in deze kwesties van pertinente aard.

De vraag naar de kwaliteit, de betrouwbaarheid en de relevantie van de internationale rangschikkingen van publieke sectoren klinkt de laatste jaren alsmaar luider in de internationale literatuur. Zoals we in de volgende hoofdstukken nog duidelijk zullen illustreren heeft een immer groeiend korps kritische auteurs bij dit proces al heel wat pijnpunten blootgelegd van de huidige governance rankings.

De vragen die zich opwerpen zijn tweeërlei. Enerzijds kan de vraag gesteld worden of de publieke sector *an sich* wel gemeten en vergeleken kan worden. Sommige critici verwerpen dit basisidee door te stellen dat de overheid te abstract is om te meten. Anderen zijn de mening toegedaan dat het gebrek aan een goede indicatorenset ervoor zorgt dat er geen zinvolle meting van de overheidswerking kan uitgevoerd worden. Nog anderen zien het nut van de overheidsmetingen niet omdat in hun ogen landen te verschillend zijn zodat de internationale vergelijking van publieke sectoren er een is van appelen en peren. Zo werpen Hood, Beeston en Dixon (2007)

de fundamentele vraag op: "*Are international rankings of governance and public services a useful tool for measuring public management performance, or just an opportunity for attention-grabbing on the basis of spurious data*?" Daar waar netgenoemde auteurs nog de vraag opwierpen, lijkt de situatie voor Handler, Koebel, Reiss en Schratzenstaller reeds duidelijk: "*Comparisons of public sector productivity over time and countries are nearly impossible and policy recommendations to improve productivity are either common sense or speculative* (Handler et al., 2005)."

Een aantal auteurs richten hun kritiek dan weer eerder op de modaliteit van de rangschikkingen van publieke sectoren, zonder de idee van het meten en vergelijken van overheidsprestaties volledig naar de prullenmand te verwijzen. Zo zijn er critici die een probleem hebben met het feit dat de rangschikkingen van publieke sectoren steevast hun beslag vinden in een index die alle onderliggende informatie teniet doet. Voor hen is de idee van het reduceren van al die complexe, genuanceerde en polyvalente aspecten van de prestaties en de kwaliteit van de publieke sector tot één score die op een ordinale schaal kan geplaatst worden, de ergste vorm van reductionisme (Hood et al., 2007). Ze zien meer in het desaggregeren van de indexen tot de diversiteit van indicatoren en bepleiten een koerswijziging richting een gebruik van de data als achtergrondinformatie (*intelligence*) bij beleidsvorming.

Anderzijds werpt zich de vraag op of de data en de rankings wel bruikbaar zijn. Zijn de rangschikkingen van publieke sectoren met andere woorden zo opgesteld dat er relevante beleidslessen uit te trekken zijn? Kan bijvoorbeeld een arm land dat met een slechte governance score en dito positie in de rangschikking, en dat daardoor uit de boot dreigt te vallen bij de verdeling van conditionele ontwikkelingshulp, aan de hand van de rangschikking inzien wat er fout loopt. Krijgt de overheid aanwijzingen over hoe het haar positie kan opwaarderen? Vraag is ook of de povere kwaliteit van de meting- en vergelijkingsoefeningen van publieke sectoren in het geval van ontwikkelingshulp wel toelaat om er bij wijlen zo een groot gewicht aan toe te kennen (cf. conditionele ontwikkelingshulp, *Millennium Challenge Account*).

Zo is er over de jaren heen felle kritiek geuit tegen verkeerdelijk gebruik van de governance rankings. Veel van die kritieken concentreerden zich op het gebruik van governance data en internationale rangschikkingen van

publieke sectoren in het differentiëren van (ontwikkelings)landen. Zo sturen enkele internationale spelers de toegang tot en de grootte van de ontwikkelingshulp en goedkope leningen (*'conditional aid allocation*'). In eerste instantie zijn er daarbij auteurs die de bedenking maken of landen met een slechte positie op internationale rangschikkingen van publieke sectoren net niet eerder geholpen moeten worden om deze score te verbeteren in plaats van gestraft door het uit de boot te vallen bij de verdeling van ontwikkelingssteun (Van de Walle, 2005). Anderen hebben eerder kritiek op het feit dat de rankings qua structuur, inhoud en betrouwbaarheid niet bedoeld zijn om donor-conditionaliteit te ondersteunen.

De Verenigde Staten hebben een voorname rol gespeeld in het opnemen van governance indicators in de toebedeling van hun ontwikkelingshulp. President Bush lanceerde in 2004 de *Millennium Challenge Account*, waarmee Amerika overstapte naar een conditionaliteit van de nationale ontwikkelingshulp. 16 indicatoren uit diverse internationale rangschikkingen van publieke sectoren (het World Bank Institute, de World Bank, de Freedom House, het International Monetary Fund, de Heritage en het Institutional Investor Magazine) moesten van dan af helpen te bepalen hoe de jaarlijkse steun voor ontwikkelingslanden verspreid moest worden. 'Control of Corruption', grotendeels bepaald vanuit de Corruption Perceptions Index van Transparency International, speelt zelfs de rol van *in-or-out-criteria*. Naast vele ontwikkelingsexperten, zijn het twee van de dataproducenten zelf die zich publiekelijk distantiëren van het gebruik van hun indicatoren voor *'aid conditionality'*. Zowel de World Bank, als Transparency International waarschuwen de Verenigde Staten dat hun rankings niet geschikt zijn voor een dergelijke toepassing (Galtung, 2005). De auteurs van de World Bank vermelden zelfs expliciet dat hun governance ranking niet geschikt is voor het differentiëren van landen om er ontwikkelingshulp mee te sturen. Ze verzetten zich dan ook hevig tegen de gedachte van conditionele ontwikkelingssteun op basis van de Worldwide Governance Indicators (zoals de Verenigde Staten doen in hun *Millennium Challenge Account*).

Andere auteurs houden het erbij dat governance rankings wel een interessant plaatje opleveren van de (veranderende) positie waarin een land zich bevindt, maar dat er voorts geen beleidslessen uit te trekken zijn. Onder andere Pollitt bevindt zich onder deze critici. "*To put it combatively,*

my main conclusion is that the Worldwide Governance Indicators - and similar aggregate measures of good governance – are almost worthless. It is hard to imagine a decision where they could safely be used as a clinching or even a main argument" (Pollitt, 2008).

2. Probleemstelling

Internationale rangschikkingen van publieke sectoren vervullen zoals aangegeven enkele belangrijke taken. In eerste instantie zijn ze een ideaal instrument in een kwantitatief onderbouwd beleid. De governance rankings en de onderliggende indicatoren vormen een maatstaf op basis waarvan nationale overheden zich internationaal kunnen positioneren. Slecht scorende landen moeten proberen lessen te trekken van beter presterende landen. Daarnaast worden de rankings ook gebruikt bij de allocatie van ontwikkelingshulp in de Derde Wereld. Bovendien worden de cijfers ook nog gebruikt door overheden en bedrijven bij investerings- en vestigingsbeslissingen in het buitenland.

De populariteit van de governance rankings is zoals aangetoond groot. De probleemstelling is echter of de kwaliteit van de rankings deze aandacht wel rechtvaardigt. Veel auteurs argumenteren dan ook dat de governance rankings in hun huidige vorm te veel gewicht op de schouders krijgen.

Aan de hand van een analyse van de kwaliteit van het product, van de producenten en van de consumenten, willen we een beeld schetsen van de globale betrouwbaarheid van de internationale rangschikkingen van publieke sectoren. Niet de cijfers die de rankings produceren, maar de machinerie achter deze data, vormen de focus van dit onderzoek. Wie verzamelt wat, waarom en met welke kwaliteit? Kortom, een vergelijking van internationale vergelijkingen.

In de eerste paragraaf geven we een inleiding op de probleemstelling. We schetsen de contradictie tussen, enerzijds, de toegenomen aandacht, interesse en gebruik van internationale rangschikkingen van publieke sectoren gedurende de laatste decennia. Als een gevolg hiervan schieten de governance rankings, geproduceerd door verschillende internationale organisaties, commerciële bedrijven, NGO's en andere instellingen, werkelijk als paddenstoelen uit de grond. Anderzijds is er echter een duidelijk gebrek aan kwaliteit bij de internationale rangschikkingen van publieke sectoren. Een toenemende groep internationale auteurs uit kritiek op de ongenuanceerde besluiten die producenten en gebruikers trekken uit de rankings. In deze studie trachten we een synthese van deze problematiek op te maken. Naast een probleemoplijsting trachten we zowel op algemeen

als op concreet vlak een kritische visie toe te voegen. In het eerste geval werpen we een analyse op van de politiek-administratieve component van de metingen, die in de bestaande kritische studies niet aan bod kwam. In het tweede geval maken we voor acht hoogaangeschreven internationale rangschikkingen van publieke sectoren een doorgedreven analyse op in evenveel meta-fiches.

De tweede paragraaf geeft een beknopt overzicht van de auteurs die zich de laatste jaren hebben ingelaten met het kritisch bestuderen van de internationale rangschikkingen van publieke sectoren.

De derde paragraaf stelt de onderzoeksvragen van deze studie op. Het gaat om deelaspecten van de algemene vraag naar de kwaliteit van de governance rankings. We trachten de problematiek ook te bekijken vanuit het standpunt van de Vlaamse Overheid, die eveneens groot belang hecht aan dergelijke vergelijkingsoefeningen, maar als regio vaak uit de boot valt. Tot slot werpen we de klovenanalyse op, waarin we trachten in te schatten hoe diep het water is tussen wat de internationale rangschikkingen van publieke sectoren vandaag aanbieden en de behoeftes die bestaan bij de gebruikers.

> 2.1. Governance Rankings: veel geblaat, weinig wol?

Governance rankings zijn aan een opmars bezig. Het meten en internationaal vergelijken van publieke sectoren zoals het vandaag gebeurt, is echter geen onverdeeld succes. De probleemstelling hangt samen met de vraag naar de kwaliteit binnen het huidige landschap van internationale rangschikkingen van publieke sectoren. Het gebruik van governance rankings kent de laatste decennia dan wel een opmars, de talloze kritieken in de internationale literatuur op hun structuur, betrouwbaarheid en gebruik - kortom de algemene kwaliteit- doen vermoeden dat dit niet altijd terecht is. De gebreken die de governance rankings van vandaag (World Bank, European Central Bank, International Institute for Management Development, World Economic Forum, e.a.) vertonen, wettigen volgens velen het gewicht niet dat op hun schouders rust. De contradictie tussen de grote aandacht, de immer toenemende interesse en het veelvuldig gebruik van de internationale rangschikkingen van publieke sectoren aan de ene kant en de diverse kwaliteitsproblemen aan de andere kant, is opmerkelijk. Het heeft enkele kritische auteurs in de pen doen kruipen. Hood (2007) suggereert het volgende: "T*he puzzle of apparently growing demand for international rankings based on relatively poor-quality data for which there is no obvious market application, might point to a 'bubble' phenomenon that could burst once its inherent shortcomings came to be widely realized*". Hoewel de luchtbelhypothese bij governance rankings wellicht een brug te ver is, duidt Hood niettemin op treffende wijze de problematiek van de kwaliteitsarme internationale rangschikkingen van publieke sectoren die desondanks grote aandacht genieten. Ook Mahony en Stevens (2002) stellen dat governance rankings veel geblaat, maar weinig wol leveren en stellen: "G*lobal studies of public sector productivity tend to generate more heat than light*."

Het rangschikken van publieke sectoren over de landsgrenzen heen, is al bij al nog een relatief nieuw fenomeen. Dit prille bestaan heeft tot gevolg dat ook het internationaal-wetenschappelijk onderzoek rond dit thema nog in zijn kinderschoenen staat. Desalniettemin heeft het gepercipieerde gebrek aan kwaliteit van de rangschikkingen van publieke sectoren enkele auteurs ertoe aangezet kritische analyses te formuleren van de bestaande internationale governance rankings. De kwaliteit van de rankings vertoont immers grote lekken. Christiane Arndt, Cristopher Hood, Gwyn Bevan, Craig Beeston, Ruth Dixon, Marie, Besançon, Steven Knack, Marcus Kurtz, Charles

Oman, Andrew Schrank, Melissa Thomas, Steven Van De Walle, Wouter Van Dooren en Steven Van Roosbroek zijn enkele van de internationale auteurs die zich in het korte verleden hebben beziggehouden met het kritisch analyseren van internationale rangschikkingen van publieke sectoren. Zij lieten het niet na te illustreren dat governance rankings nog veel werk hebben om solide en betrouwbaar genoemd te mogen worden. Deze analyses hebben echter ook gebreken. Ze zijn vaak anekdotisch en eenzijdig gericht op de kritiek van de rankings. Met deze studie willen we diepgaand en systematisch de rankings analyseren, en ook aangeven hoe de overheid nuttig gebruik kan maken van de rankings.

De vraag die hier gesteld wordt is: wie produceert de rankings, waarom en voor wie? En met welke kwaliteit wordt dat gedaan? We gaan als startpunt na welke kritieken de internationale rangschikkingen van publieke sectoren reeds te verwerken hebben gekregen.

> 2.2. Critici en problemen

In deze paragraaf schetsen we een algemeen beeld van wat de internationale literatuur reeds kritisch heeft onderzocht op het vlak van overheidsprestaties.

Van Roosbroek en Van Dooren (2006) maakten in hun rapport 'Meten is weten?! België/Vlaanderen in internationale indicatoren en indices' een algemene beschouwing van rangschikkingen en indicatoren van publieke sectoren. Ze wijzen onder andere op het verkeerdelijk aannemen van percepties voor feiten. Ook de gebrekkige verantwoording van indicatorenkeuzes en wegingsbeslissingen is voor hen een doorn in het oog. Ze klagen ook het fenomeen aan van over-aggregatie van indicatoren tot een concept waarvan nog moeilijk te onderscheiden is wat nu eigenlijk gemeten wordt. Met name de Worldwide Governance Indicators van de World Bank, met hun 158 onderliggende indicatoren een van de omvangrijkste governance metingen, scoren hier slecht op. De auteurs besluiten met het aangeven van de nood aan een statistische autoriteit die waakt over de cijfers en de statistieken. Bovendien wijzen ze de producenten er op dat ze kleur moeten bekennen door de beleidsvisie achter de index kenbaar te maken. Ook het opnemen van betrouwbaarheidsmarges en het zorgvuldig selecteren van respondenten bij subjectieve metingen behoort tot hun aanbevelingen voor een betere meting van publieke sectoren.

Van de Walle, Sterck en Van Dooren (2004) komen tot de conclusie dat er "op dit moment fundamentele problemen zijn met zowel de kwaliteit van de data, als met de manier waarop data verzameld en beschikbaar worden gesteld". Zowel op het vlak van de validiteit, als op het vlak van de betrouwbaarheid zien deze auteurs nog veel mankementen. Er is een gebrek aan eenduidige definities en de gegevensverzameling beperkt zich te vaak tot het meten van de perceptie van de overheid. De betrouwbaarheid wordt dan weer aangetast door het gebrekkige aantal observaties dat in sommige gevallen wordt verricht, het gebrek aan kwaliteitscontrole van aangeleverde data en het laattijdig bijwerken van de scores.

Van de Walle stelt verder in zijn werk *'International comparisons of Public Sector Performance: how to move ahead*?' (2007) dat de kwaliteit van governance data in de internationale literatuur eensgezind onder vuur wordt

genomen. Hij ziet echter een groot verschil tussen wat hij noemt *Public Administration*-onderzoek aan de ene en onderzoek verricht door economici en econometristen aan de andere kant. In de *Public Administration* vat men de hele problematiek eerder op als een conceptueel probleem: 'Wat is een goede overheid en hoe moet die gemeten worden?' "*We cannot measure government if we cannot define it*", zo zouden we deze visie kunnen samenvatten. De economische wereld beschouwt de overheid en haar objectieven dan weer als een gegeven en spreekt eerder van een meet-probleem. Zij stappen probleemloos over naar de vraag hoe de overheid best gemeten zou kunnen worden.

Rothstein en Teorell bouwen voort op deze zienswijze. De groei van onderzoek naar *good governance* wordt volgens hen ondersteund door empirische bewijsvoering vanwege economici dat goed werkende instellingen de sleutel vormen voor economische groei in ontwikkelingslanden. Het probleem volgens deze twee auteurs is echter dat daarbij een belangrijk punt over het hoofd wordt gezien. Er wordt immers volledig aan de vraag voorbijgegaan wat 'good governance' op conceptueel vlak juist inhoudt. "De definities die economici gebruiken zijn ofwel veel te breed ofwel hebben ze te lijden onder een functionalistische invalshoek waardoor de hanteerbaarheid zo goed als nihil is (Rothstein & Teorell, 2005; 2005)", aldus deze kritiek.

Ook Grindl (2007) stelt vragen bij de bruikbaarheid van de internationale rangschikkingen van publieke sectoren. Landen die een slechte score krijgen toebedeeld en deze wensen te verbeteren, worden volgens deze auteur volledig aan hun lot overgelaten. Zo zijn er nog te veel onbeantwoorde vragen in de aard van welke overheidsinstituties het belangrijkst zijn en welke maatregelen nu het best genomen dienen te worden om ontwikkeling te voeden en governance te verbeteren. De cruciale vraag die meerdere overheden in de wereld bezig houdt, is immers: "*Given limited resources of money, time, knowledge, and human and organizational capacity, what are the best ways to move towards better governance in a particular country context*?" Aan dit nochtans vitaal beleidsrelevant aspect wordt echter weinig of geen aandacht besteed in de verschillende internationale rangschikkingen van publieke sectoren.

Arndt en Oman richten in hun *'Uses and abuses of governance indicators*' (2006) hun pijlen voornamelijk op de '*composite indicators'*, de

geaggregeerde indices, die veel gebruikt worden in internationale rangschikkingen van publieke sectoren. De auteurs stellen dat dergelijke indicatoren steevast transparantie ontberen, aan *selection-bias* lijden en eigenlijk niet geschikt zijn voor vergelijkingen in de tijd of tussen landen onderling. Doordat dit laatste feit vaak in de wind wordt geslagen, spreken deze auteurs van een *'widespread misuse*' van governance indicators.

In 2005 werd er in Oxford een conferentie gewijd aan de problematiek van *international public service rankings*. Hood en Beeston (2005) wezen op de ongecontroleerde proliferatie van governance rankings van de laatste jaren. Doordat er veel kwaliteitsproblemen zijn en weinig kwaliteitslabels zijn uitgereikt, is het voor gebruikers *'ranking shopping*', vaak zonder te weten wat in het mandje verdwijnt. Street (2005) illustreerde er dat op basis van het statistische model van de World Health Organisation zelfs een significant verband tussen de gezondheidszorg en de sterkte van het nationaal voetbalteam kan worden bewezen. Zo toonde hij een sterk significant verband aan tussen de WHO-ranking en de FIFA-ranking (wereld voetbalbond). "Als beleidsmakers de gezondheidsstaat van de eigen bevolking willen opkrikken, moeten er inspanningen geleverd worden om het nationale voetbalelftal te versterken", klonk het ironisch. Hij greep dit voorbeeld aan om de kwaliteit van vele hooggewaardeerde governance rankings in vraag te stellen. Ook Thomson (2005a) wees op het kwaliteitsprobleem van veel internationale rangschikkingen van publieke sectoren, waardoor ze uiteindelijk weinig nuttig blijken voor overheden. Verder ijvert zij voor een onafhankelijke instantie die de kwaliteit van data en statistieken controleert, zoals in Groot-Brittannië gebeurt met de *Audit Commission's Comprehensive Performance Assessments*. Manning (2006) doet een poging om verantwoordelijken aan te wijzen voor het aanleveren van kwaliteitsgebrekkige comparatieve data. Enerzijds identificeert hij de '*entrepreneurs*', een groep analisten die een probleem zag in het gebrek aan mondiale informatie op het vlak van de publieke sector. Hun antwoord op deze gepercipieerde nood was snel simpele statistieken samen te stellen die landen op een gemakkelijke manier rangschikken. Anderzijds ziet hij ook een andere groep een hand hebben in de slechte kwaliteit van de governance rankings. De '*trainspotters'* waren volgens Manning haast bezeten door het vergelijken van publieke sectoren, en wilden daardoor zoveel mogelijk data de wereld insturen, ongeacht de kwaliteit en betrouwbaarheid.

Ook Pollitt twijfelt aan de voordelen van internationale rangschikkingen van publieke sectoren. Hij verwijst daarbij naar *'a huge demand for duff data*' en vraagt zich luidop af of er beslissingen en overwegingen op beleidsvlak kunnen genomen worden op basis van dergelijke rankings (Boyle, 2007). Zo beschrijft hij de Worldwide Governance Indicators illustratief als "*a bunch of nonsense, dressed up as comparative statistics*". De governance rankings dragen volgens hem niets bij aan overheden die over een zwak overheidsbestuur beschikken.

De auteurs van de Worldwide Governance Indicators kunnen trouwens zelf niet om dit deficit heen. Kauffman, Kraay en Zoido-Lobotàn (Kauffman et al., 2000) geven toe dat zelfs de beste internationale indicatoren van governance precisie en betrouwbaarheid missen en weinig zeggen over de specifieke institutionele falingen die voor zwakke governance zorgen in bepaalde landen. De echte uitdaging licht er volgens deze onderzoekers in om samen te werken met landen om deze falingen te diagnosticeren, te identificeren en aan te pakken op nationaal en subnationaal niveau. Toch wel een weinig ambitieuze claim van een van de *leading actors* op het vlak van governance rankings.

Om dit te doen, moeten er indicatoren worden ontworpen die kunnen bijdrage aan *'high pay-off reforms*' (Knack et al., 2003b). Volgens bepaalde critici moet er op zoek gegaan worden naar beleidsrelevante en politiek aanvaardbare kwantitatieve indicatoren van governance. Enkel en alleen op die manier kan governance gemeten worden met als doel het ondersteunen van constructieve verandering.

Er komt ook kritiek vanwege gezaghebbers uit landen die een slechte beurt maken in de governance rankings. Zo uitte Norman Girvan, de Secretaris-generaal van de Caribische Staten zijn ongenoegen over de manier waarop de huidige governance rankings ingevuld worden door diverse internationale organisaties en private bedrijven. Volgens hem kampen de meeste rankings met een impliciete bias in de richting van het Westerse meerpartijensysteem. "De metingen van publieke sectoren zijn zo opgesteld dat de Westerse landen steevast *perfect scores* krijgen op indicatoren zoals Polity, Civil liberties of Political rights", aldus Girvan. "De vraag", zo gaat hij verder, "is dan ook of private of quasipublieke instellingen, die weinig of geen verantwoording ten opzichte van een algemeen aanvaarde

Internationale Gemeenschap moeten afleggen en die procedures hanteren die vaak moeilijk of zelfs niet toegankelijk zijn voor het algemene publiek, de legitimiteit zouden mogen krijgen om een dergelijk grote macht uit te oefenen met hun beoordelingen van landen, bevolkingen en systemen. *The procedure itself may not meet the test of good governance."* (Girvan, 2002)

Arndt (2007) tracht verder te kijken en doet een aanzet voor een verbeterde meting in de toekomst. "Governance indicators", zo stelt ze, "moeten niet zoveel hooi op de vork nemen zoals ze vandaag de dag doen. De governance indicators van de toekomst moeten zich meer toespitsen op een bepaald aspect van de governance, meer transparantie aan de dag leggen en ook niet de ambitie koesteren de hele wereld te willen coveren." Met een beperktere maar nauwgezettere ambitie zullen de governance indicators volgens haar al een heel deel van hun kwaliteitsproblemen kunnen oplossen.

> 2.3. Onderzoeksvragen

Internationale organisaties en adviesbureaus compileren een brede waaier aan internationale rangschikkingen. De vergelijkingen betreffen zowel beleidsprestaties (milieu, economie, mobiliteit, gezondheid) als het overheidsapparaat zelf (governance indicators, vertrouwen in de overheid, *e-gov benchmarks*, competitiviteitindices). In dit onderzoek nemen we enkel de rangschikkingen van bestuur onder de loep. We kijken naar de internationale metingen van *'public administration*'. Internationale vergelijkende indicatoren zijn een belangrijk element in een kwantitatief onderbouwd beleid. Vlaanderen kan zich dankzij deze vergelijkingen internationaal positioneren. Er zijn echter ook problemen. In de vorige paragraaf (cf. 2.2) hebben we al enkele kritieken weergegeven. De conclusie is dat er nog vele kwaliteitspijnpunten in de metingen en de rangschikkingen van publieke sectoren zitten. Desalniettemin worden de cijfers meestal zonder veel discussie aangenomen (Van de Walle et al., 2004).

Onder de algemene doelstelling om de internationale vergelijkingen van publieke sectoren te onderzoeken op hun kwaliteit, betrouwbaarheid en hanteerbaarheid, poneren we hier enkele concrete onderzoeksvragen. Deze onderzoeksvragen vormen de rode draad doorheen de synthese en analyse die in dit onderzoek gevoerd worden. Ten eerste stellen we de vraag naar wie er achter de productie van de governance rankings schuilgaat. Het is tevens een vraag naar de aard van de producent. Hierop verder bouwend trachten we ten tweede na te gaan wie deze producent voor ogen heeft bij het produceren van internationale rangschikkingen van publieke sectoren. Wie behoort tot de doelgroep van de aangeleverde data? In een adem onderzoeken we ook de beweegredenen daarbij. Wat zijn de beoogde objectieven van de meting? Waarvoor moeten de data dienen? Ten derde trachten we de rankings te ontrafelen. Door de structuur tot op indicatorenniveau bloot te leggen, wordt een beter inzicht in de meting mogelijk. Vervolgens stellen we ten vierde de algemene vraag naar de kwaliteit van de governance rankings. We stellen enkele criteria op, op basis waarvan we een kwaliteitsvolle governance ranking trachten af te meten. Ten vijfde trachten we naast de kwaliteit ook de hanteerbaarheid van de rankings voor het beleid na te gaan. Ten slotte gaan we als zesde deelvraag de kloof na tussen wat op de markt verschijnt en wat door gebruikers verwacht en verlangd wordt op het vlak van comparatieve

overheidsmetingen. Aan de hand van deze klovenanalyse trachten we enkele aanbevelingen op te stellen om producenten en consumenten van governance rankings dichterbij elkaar te brengen.

> *2.3.1. Wie produceert de rankings?*

De eerste hoofdvraag die we ons stellen is de vraag naar de dataoorsprong. Aan de bestaande kritische analyses willen we immers een component toevoegen. Het politiek-administratieve aspect van de metingen -zoals we het dopen- tot nog toe grotendeels onderbelicht gebleven. De vragen 'Wie?', 'Waarom?' en 'Voor wie?' willen we invullen, om zo een beter inzicht te krijgen in de kwaliteit van de rankings.

De organisatie, het bedrijf of de instelling die de rangschikkingen opstelt is immers een belangrijke factor in de beoordeling van de rangschikkingen. Als we tot een correcte interpretatie van de governance-metingen willen komen, moeten we duidelijk maken wat de aard van de producent is. De internationale rangschikking van publieke sectoren kan van de hand van een internationale organisatie zijn zoals de World Bank (WB), De United Nations (UN) of de Organisation for Economic Co-operation and Development (OECD). Naast de internationale organisaties nemen ook anderen de taak op zich om overheden op internationale schaal te vergelijken. Zo zijn er de universiteiten en andere onderzoeksinstellingen zoals het International Institute for Management and Development (IMD) die op vrije of commerciële basis governance rankings samenstellen en aanbieden.

Naast de bovenvermelde organisaties, zijn er ook private bedrijven die governance data tegen betaling op de markt brengen. De European Intelligence Unit (EIU) en de Political Risk Services Group (PRS Group) zijn hier voorbeelden van. Het minste dat kan gezegd worden is dat dergelijke private organisaties wetenschappelijke objectiviteit niet altijd als hoogste prioriteit nemen bij het produceren van internationale organisaties.

Men moet overigens kritisch zijn en blijven ten opzichte van de producent en niet enkel afgaan op de naam en faam van de producent. Enkele hoogaangeschreven internationale organisaties zoals de United Nations, de Organisation for Economic Co-operation and Development, de World Bank en andere houden zich bezig met het opmaken van governance data en het samenstellen van internationale rangschikkingen van publieke sectoren,

maar dat is niet altijd een kwaliteitsgarantie op zich. Dat bewijst onder andere de Public Sector Performance, die ondanks de goede reputatie van de European Central Bank in de internationale literatuur veel kritiek oogstte.

Er kan ook een onderscheid gemaakt worden tussen de rankings die een Angelsaksische inslag hebben en internationale vergelijkingen die eerder Europees geïnspireerd zijn. Angelsaksische metingen vertrekken over het algemeen eerder vanuit de *'minimal state*'. Dergelijke rankings hebben een overwegend economische inslag en bezien de overheid vanuit die optiek als een (economische) verkeersregelaar die er vooral op moet toezien dat het een consistente koers aanhoudt op economisch vlak, er voor moet zorgen dat het handelsverkeer optimaal kan verlopen en liefst in een minimale dosering van administratieve lasten voor de economische agenten moet voorzien. Kortom; een bedrijfsvriendelijk klimaat scheppen en tegelijk niet in de weg lopen.

De Europees geïnspireerde rankings zijn meer begaan zijn met democratische, politieke en milieuwaarden. De Europese Unie is in 2000 begonnen met het oprakelen van zogenaamde *'lost issues'* in het governance debat. Europa wilde voornamelijk de nadruk leggen op politiek aspecten van governance, zonder de economische prioriteiten zoals economische groei over het hoofd te zien.

> 2.3.2. Wat zijn de doelstellingen en het doelpubliek?

Naast de analyse van de aard van de producent van internationale rangschikkingen van publieke sectoren, zijn er nog twee andere aspecten van de politiek-administratieve dimensie van governance rankings die we van dichterbij gaan bekijken. In eerste instantie kijken we naar de objectieven van de ranking. Aansluitend bij de aard van de producent kan er onder andere een onderscheid gemaakt worden tussen rangschikkingen die als een commercieel product worden aangeboden en het dus moeten hebben van een zo groot mogelijke afname en de rankings die niet afhankelijk zijn van de verkoopcijfers van hun product.

De abonnementsprijzen die voor de toegang tot sommige internationale rangschikkingen van publieke sectoren worden gevorderd, zijn immers niet van de poes. Zo betaal je voor de Grey Area Dynamics van de Merchant

International Group -risicoanalyses bestemd voor bedrijven- $3.000 per jaar en voor een basistoegang tot de International Country Risk Service van de Political Risk Services Group zelfs bijna $8.000. De bekende Global E-governance Index in zijn totaalpakket, met toegang tot de individuele indicatoren van alle onderzochte landen, loopt zelfs op tot meer dan $100.000. Deze prijzen schaden de transparantie en de *actionability* van de rankings. Niet alle landen kunnen of willen immers dergelijke grote bedragen betalen en worden daarmee uitgesloten van inzage in deze internationale rangschikkingen van publieke sectoren. Een gevolg daarvan is ook dat deze overheden met moeite leereffecten kunnen oppikken van deze rankings, aangezien ze geen toegang hebben tot de basisscores, de indicatoren en de methodologie van deze metingen.

De rankings en de data in hun geheel kunnen ook gedifferentieerd worden op basis van de doelstelling die ze meekrijgen. Sommige rankings, zoals de Country Performance and Institutional Assessments (CPIA) van de World Bank zijn er rechtstreeks op gericht om de allocatie van financiële middelen van de International Development Association (IDA) mee te sturen. Van deze doelstelling achter een ranking hangen onder andere de *country coverage* (Derde Wereldlanden), de toegang (vrije beschikbaarheid) en de leermogelijkheden (groot aandeel van 'actionable indicators') af. Andere veel voorkomende doelstellingen betreffen onder andere het adviseren van bedrijven die in het buitenland opereren (bijv. de Country Risk Briefing van de Economist Intelligence Unit) en het creëren van leereffecten (bijv. de Public Sector Performance van de European Central Bank). Het is belangrijk om deze impliciete of expliciete doelstellingen mee te nemen in de interpretatie en beoordeling van elke governance ranking.

Ten tweede mag het belang van de doelgroep, die de producent van de governance ranking voor ogen heeft bij het samenstellen van de rangschikking, niet onderschat worden. De users die door de ranking beoogd worden, liggen bovendien vaak in het verlengde van de doelstelling(en) van de meting. Zo mag het niet verbazen dat er een bepaalde selectiviteit en bias schuilt in bijvoorbeeld rankings die zich op een selectief doelpubliek richten. Een voorbeeld kan dit illustreren.

Het International Institute for Management Development (IMD) brengt op jaarlijkse basis het World Competitiveness Yearbook uit, een internationale rangschikking van publieke sectoren die grote bekendheid heeft verworven.

De competitiviteit van 55 landen wordt in deze ranking gemeten door de verschillende landen op een aantal overheidscomponenten te quoteren, zoals onder andere 'Government Inefficiency'. Een nadere blik op de samenstelling van de meting van de overheidsefficiëntie brengt echter een grote selectiviteit aan het licht. De meting is immers vanuit een zuiver economische invalshoek opgesteld. Kort samengevat komt het er op neer dat een efficiënte overheid door het International Institute for Management and Development opgevat wordt als een overheid die een gunstige omgeving weet te scheppen voor binnen- en buitenlandse bedrijven. De overheid moet er vooral op letten zo min mogelijk hindernissen op te werpen voor de individuele bedrijven zodat die zorgeloos op nationaal en internationaal vlak met elkaar kunnen concurreren. Als men echter in rekening brengt dat bedrijfsleiders met stip op één staan bij de afname van deze ranking en dat het International Institute for Management Development zich ook expliciet richt tot bedrijven en managers die er bedrijfsoverwegingen mee willen sturen, is dit niet echt verwonderlijk. Tekenend is bovendien het feit dat het World Competitiveness Yearbook stoelt op subjectieve indicatoren, die afkomstig zijn uit een survey onder managers en bedrijfsleiders. De doelgroep bestaat uit de economische wereld, de indicatoren zijn economisch getint en zelfs de subjectieve invulling van de survey is economisch gekleurd. Dit alles zorgt ervoor dat de internationale rangschikking van publieke sectoren van het IMD ook vanuit die optiek moet bekeken worden. Een 'efficiënte overheid' kan daarbij niet volledig gelijkgesteld worden met een efficiënte overheid voor de gehele bevolking, noch komt ze tegemoet aan de bestuurskundige opvatting van 'de efficiënte overheid'.

In onze analyse zullen we dus trachten een beeld te scheppen van de doelstellingen en het doelpubliek die de internationale rangschikkingen van publieke sectoren voor ogen hebben bij de meting van de verschillende overheden. We onderzoeken deze aspecten aan de hand van de objectieven die geëxpliciteerd worden bij de rangschikkingen, op de website van de producenten en in methodologische en begeleidende papers van de governance rankings. Daarnaast kan men uit de methodologie, de aard van de producent en andere eigenschappen van de rankings vaak ook onrechtstreeks afleiden waarom en voor wie ze geproduceerd worden.

> *2.3.3. Belang van het meetmodel*

Producenten van de governance rankings van vandaag pakken maar al te graag uit met hun prestatieladders. De conclusies die op basis van de internationale rangschikkingen van publieke sectoren worden getrokken, blijven echter te vaak beperkt tot het in de schijnwerpers plaatsen van de behaalde relatieve score en de plaatsing in de ranglijst. De rangschikkingen van publieke sectoren steunen nochtans meestal op diverse onderliggende componenten en indicatoren, die geaggregeerd worden tot één index. Jammer genoeg is deze ene score en de ene plaats op de ladder die daarbij hoort doorgaans het enige wat in de verf wordt gezet, terwijl de onderliggende indicatoren een veel grotere en concretere nieuwswaarde hebben.

Een simpel voorbeeldje illustreert deze stelling. Land A scoort op een schaal van 0 to 100 goed op indicator 2 (80/100), terwijl het voor indicator 1 heel wat minder doet (20/100). Land B daarentegen kent geen uitschieters, noch extreme zwakheden op de onderzochte variabelen en scoort tweemaal middelmatig (50/100). Als we enkel en alleen naar de totaalscores en de ranking kijken, zouden we kunnen concluderen dat land A dezelfde prestaties aan de dag legt als land B. De totaalscores verbergen echter fundamentele verschillen op de onderzochte indicatoren.

	Indicator 1 (w=0,50)	Indicator 2 (w=0,50)	Totaalscore
Land A	20	80	40
Land B	50	50	40

Om de kwaliteit van de internationale rangschikkingen van publieke sectoren na te gaan, moeten we dus dieper graven dan de totaalscores en de ranking in enge zin. Belangrijk voor ons, is het proces, de manier waarop de totaalscore en de gerelateerde rangschikking tot stand komt. De structuur van de meting van publieke sectoren is daarbij van cruciaal belang.

Munck en Verkuilen (2002) onderscheiden drie uitdagingen in een meting. Ten eerste is er de *conceptualization*, de conceptualisering van het onderzoek. De producent van een meting van publieke sectoren moet

aangeven wat er juist gemeten wordt. Ten eerste moet ze een naam voor de index bedenken, het label. Vervolgens moeten er kenmerkende dimensies geïdentificeerd worden die onder het concept liggen. Deze dimensies helpen in het specificeren van de betekenis van het concept. Ten derde moeten er componenten van deze onderdelen geselecteerd worden. Dit zijn de onderliggende componenten. Ten vierde moet er een meetniveau gekozen worden. Kiest men voor een nominale, een ordinale, een interval- of een categorisch meetniveau? Als laatste moet een aggregatieregel vastgelegd worden. Hier zijn diverse mogelijkheden voor: er kan geopteerd worden voor een additief systeem, een multiplicatief model of een andere wegingsmethode.

Idealiter ligt er een theorie van governance aan de basis van het conceptuele kader. Deze theorie moet dan een verantwoording bieden voor de gemaakte keuzes. Een minimum zou toch moeten zijn dat de conceptuele keuzes enigszins worden verantwoord. In hoofdstuk 4 gaan we verder in op de conceptuele basis van de internationale rangschikkingen van publieke sectoren. Daar wordt duidelijk dat veel governance rankings niet eens aan deze minimumstandaard voldoen.

Vervolgens is er de uitdaging van de *measurement*, hoe de meting zelf wordt uitgevoerd. Er zijn twee aspecten aan deze fase; enerzijds is er de vastlegging van de indicatoren, aan de andere kant de keuze voor het meetniveau. Dit aspect peilt naar de meetmethode die aan de basis ligt van de rangschikking. Het gaat om de identificatie van variabelen die de conceptuele dimensies, die al eerder bepaald en logisch gestructureerd werden, linken met observaties. Het is een noodzaak om een meetmodel samen te stellen dat onobservaarbare, latente variabelen verbindt met observeerbare variabelen, indicatoren genaamd (Bollen, 1989).

De eerste stap die in de formatie van de meting moet gebeuren, is de selectie van '*the leaves of the concept tree*' -de variabelen op het laagste abstractieniveau. Multiple indicatoren zorgen voor een kleinere bias, maar maken het moeilijker om een onderlinge balans te vinden. Munck en Verkuilen (2002) stellen dat de keuze voor deze of gene indicatoren altijd mede ingegeven wordt door beschikbaarheid en toegankelijkheid. "Je kunt er niet omheen dat dergelijke praktische overwegingen meespelen in deze selectie", aldus deze auteurs. Idealiter worden indicatoren verkozen die een kleine kans op bias vertonen en die kunnen gecheckt worden aan de hand

van meerdere bronnen. Bij de selectie van indicatoren, kan er geopteerd worden voor objectieve indicatoren of voor subjectieve bevragingen of een mix van de beide. Men moet wel opletten voor de misvatting dat objectieve indicatoren ook echt 'objectief' zijn. Er komt immers een subjectieve beoordeling aan te pas bij de selectie van objectieve indicatoren. Men dient ook te kiezen voor een meting van input-, output-, outcome-, efficiëntie- of effectiviteitsindicatoren.

Ten tweede moeten er keuzes gemaakt worden betreffende de meetniveaus. Ook hier speelt de validiteit een belangrijke rol. De ideale selectie van het meetniveau moet aan twee voorwaarden voldoen:

a) ze streeft naar homogeniteit binnen meetklassen met een minimum aan noodzakelijke condities
b) de selectie moet gezien worden als een proces dat zowel theoretische verantwoording als empirische tests vereist.

De variabelen die geselecteerd worden, moeten betrouwbaar (*reliable*) en reproduceerbaar (*replicable*) zijn. Het eerste duidt op het vooruitzicht dat eenzelfde datagaringsproces steeds dezelfde data voortbrengt. Het gaat in feite om de consistentie van de codering. De reproduceerbaarheid meet men af aan de mate waarin onderzoekers de mogelijkheid hebben om het dataverwerkingsproces te herdoen.

De studie van Munck en Verkuilen (2002) werpt in dit kader drie aanbevelingen op voor metingen van publieke sectoren die met expertenquotering werken. Aangezien de meeste internationale rangschikkingen van publieke sectoren gebruik maken van dit soort indicator, geven we ze even mee.

a) De producent moet zijn coderingsregels kenbaar maken. Op zijn minst moet er een lijst van de gebruikte indicatoren vrijgegeven worden, moet het geopteerde meetniveau voor elke variabele vrijgegeven worden en moet er gedetailleerde informatie verspreid worden over de betekenis van elke schaal.
b) Het codeerproces moet ook publiekelijk beschikbaar gesteld worden. Dit proces houdt onder andere in dat alle bronnen die gebruikt worden bij het coderen, het aantal codeurs en het resultaat van inter-codeur-betrouwbaarheidstesten ter beschikking gesteld worden. Metingen die met één enkele codeur werken zijn uit den boze.

c) Ook de gedesaggregeerde data voor elke indicator moeten ter beschikking van publiek gesteld worden.

Het blootleggen van de indicatorenset die aan de basis ligt van een meting van bijvoorbeeld een efficiënte overheid, is bijzonder relevant. Deze indicatorenset vormt immers het inhoudelijke plaatje van de meting. Aangezien een conceptuele theorie meestal ontbreekt bij de meeste internationale rangschikkingen van publieke sectoren die momenteel geproduceerd worden, is het des te belangrijker om aan de hand van de gebruikte indicatoren en het meetproces in te schatten wat nu juist gemeten wordt achter de ranking.

Naast de *conceptualization* en de *measurement*, is ook de *aggregation* -het aggregatieproces- van groot belang bij de meting, vergelijking en rangschikking van publieke sectoren. Waar de fases van *conceptualization* en *measurement* eerder de weg van boven naar benden afleggen -men stelt de vraag wat gemeten wordt en hoe dit best geoperationaliseerd wordt naar dimensies, variabelen en indicatoren- legt de aggregatiefase de omgekeerde weg af. Er wordt immers een methode uitgewerkt om het desaggregatieproces in de conceptualisatiefase om te keren en scores op de individuele indicatoren te transfereren tot geaggregeerde scores in een of meerdere indices. De vraag naar de aggregatiemethode komt in feite neer op de vraag naar de wegingsfactoren. De keuze voor deze of gene wegingsmethode bepaalt immers in grote mate de uiteindelijke totaalscore. Enerzijds moet er een aggregatieniveau worden vastgesteld, aan de andere kant een aggregatiemethode.

De eerste stap houdt een selectie van het aggregatieniveau in. Zoals het bovenvermelde voorbeeld illustreerde, gaat de stap naar een hoger niveau gepaard met een validiteitsverlies, in die mate dat informatie over systematische variatie tussen verschillende metingen verloren kan gaan. De selectie van het aggregatieniveau is een expliciete keuze die verantwoord moet worden in het licht van de nood om het verlangen tot spaarzaamheid te verzoenen met de bezorgdheid om de onderliggende dimensionaliteit en differentiatie te bewaren. De praktijk wijst echter uit dat producenten van internationale rangschikkingen van publieke sectoren enkel en alleen de eerste overweging in beschouwing nemen. Er wordt zo goed als altijd getracht om te aggregeren tot op het hoogste niveau, totdat er slechts één index of score meer overblijft. Hierbij voelen de producenten zich niet

getemperd door een verlies aan informatie. Het World Bank Institute bewijst echter dat het ook anders kan. Het behoudt zes componenten, de zes Worldwide Governance Indicators, die het niet verder in één score tracht te persen. Anderen, zoals het International Institute for Management Development en het World Economic Forum hebben blijkbaar niet aan de verleiding van de aggregatie kunnen weerstaan. Zij opteren voor de weg van de simplificatie en drukken honderden indicatoren tot één index.

De tweede stap van de aggregatiefase wordt gevormd door de aggregatiemethode. Deze stap is van cruciaal belang voor de uitkomst van de data. De keuze voor deze of gene wegingsmethode bepaalt immers in grote mate de uiteindelijke totaalscore. Op theoretisch vlak moet de producent eerst de theorie achter de relatie tussen de verschillende dimensies en componenten expliciteren, waarna de analist overeenstemming moet creëren tussen deze theorie en de geselecteerde aggregatiemethode. Hier komt de selectie van de wegingsfactoren bij kijken.

De wegingsfactoren kunnen een belangrijke rol spelen in metingen van publieke sectoren en een groot aandeel hebben in de constellatie van een uiteindelijke internationale rangschikking. Een grotere wegingsfactor staat voor een groter belang van een indicator. Op die manier weerspiegelen subjectieve, politieke en normatieve overwegingen zich in de wegingsfactoren en zijn ze op die manier ook een deel van de *conceptualization*.

Een simpel voorbeeld kan duidelijk maken in welke mate de wegingsmethode en de gelieerde wegingsfactoren hun invloed kunnen hebben op de meting, de quotering en de rangschikking. Land A scoort goed op indicator 4 (80/100), iets minder op indicator 3 (60/100), nog wat minder op indicator 2 (40/100) en het minst goed op indicator 1 (20/100). Land C presteert net omgekeerd: op indicator 1 scoort het 80/100, op de tweede indicator 60/100, 40/100 voor de derde indicator en 20/100 voor indicator 4. Land B scoort middelmatig (50/100) op elk van de vier geselecteerde indicatoren.

We beschouwen deze scores bij twee verschillende wegingsmethodes. We maken hierbij de expliciete keuze om te aggregeren tot op het hoogste niveau (één index), waarbij alle informatie in één score gewrongen wordt.

In het eerste voorbeeld opteren we voor een gelijke weging (*'equal weighting'*). De uitkomst is dat Land A, land B en land C eenzelfde totaalscore bekomen en dus eenzelfde positie in de rangschikking bekleden.

	Indicator 1 (0.25)	Indicator 2 (0.25)	Indicator 3 (0.25)	Indicator 4 (0.25)	Totaal
Land A	20	40	60	80	50
Land B	50	50	50	50	50
Land C	80	60	40	20	50

In voorbeeld 2 onderwerpen we dezelfde landen, met dezelfde prestaties aan een ander wegingsschema. Indicator 4 (0,40) wordt door ons -producent zijnde- als belangrijker ingeschat dat indicator 3 (0,30), die op haar beurt meer waarde krijgt dan indicator 2 (0,20) en indicator 1 (0,10). Ondanks dat er aan de situatie in de drie landen niets veranderd is, bekomen we hier toch een andere uitkomst van de meting en rangschikking. Land A neemt de leiding in de rangschikking met 60 punten, daar het beter presteerde op indicatoren die hoger worden ingeschat door de producent. Land B (50 punten) verwerft omwille van deze zelfde reden een betere score en positie dan land C (40 punten), dat de rij sluit omdat het vooral goed presteert op de minder gewaardeerde onderdelen.

	Indicator 1 (0.10)	Indicator 2 (0.20)	Indicator 3 (0.30)	Indicator 4 (0.40)	Totaal
Land A	20	40	60	80	60
Land B	50	50	50	50	50
Land C	80	60	40	20	40

Nog dit: de opvatting van sommige rankingsproducenten dat een *'equal weighting scheme'*, waarbij elke indicator en/of component een gelijke weging krijgt toebedeeld, de meest objectieve wegingsmethode is, is erg misleidend. Ook een gelijke weging is een weging en dus verbonden met impliciete normatieve overwegingen. Er is geen reden om aan te nemen dat een gelijke weging in de regel objectiever, betrouwbaarder en accurater is dan een andere aggregatiemethode.

Kortom; het aggregatieproces is per definitie normatief getint. Op zich is daar ook niets mis mee, er is immers geen weg langs. Het probleem is echter dat de wegingsfactoren en dus de normatieve keuzes die worden gemaakt slechts zelden verantwoord en onderbouwd worden. De weging weerspiegelt een keuze, maar de gebruiker kan vaak niet inschatten op welke basis die beslissing genomen wordt. Waarom wordt indicator x belangrijker geacht dan indicator y? Deze overweging blijft te vaak open voor speculatie. Indexproducenten hebben over het algemeen aggregatiemethodes gebruikt op een ad-hoc-manier. Theorie of enige andere verantwoording bleef daarbij steevast achterwege.

> *2.3.4. Wat is de kwaliteit van de rankings?*

De bruikbaarheid van de internationale rangschikkingen van publieke sectoren wordt in de eerste plaats bepaald door de kwaliteit van de meting en de ranking. Kwaliteit betekent hier onderzoekstechnische kwaliteit. We zullen enerzijds de vraag stellen naar de validiteit. Hierbij trachten we na te gaan of men de goede dingen meet. We zullen onder deze noemer onderzoeken of de rankings meten wat beweerd wordt te meten. Kunnen met andere woorden de indicatoren, het meetmodel en de manier waarop gemeten wordt het concept dragen dat gemeten dient te worden?

Daarnaast analyseren we ook de betrouwbaarheid. Bij dit aspect stellen we veeleer de vraag of men goed meet. Hierbij werpen we een blik op de technisch-statistische kant van de meting. Het gaat om de betrouwbaarheid van de indicatoren, van de datagaring en van de analyses.

Naast de validiteit en de betrouwbaarheid, is ook transparantie een belangrijk aspect van een kwalitatieve internationale rangschikking van publieke sectoren. Het is immers belangrijk voor gebruikers en analisten om achter de scores en rankings te kunnen kijken. Niet alleen is het van groot belang dat de dimensies, componenten en indicatoren ter beschikking van de externe analist gesteld worden, ook de conceptuele theorie en verantwoording van normatieve keuzes moeten idealiter helder en consistent worden meegegeven. Dit geldt zowel voor het indicatorenniveau (beschikbaarheid van het geraamte achter de ranking, het meetniveau, de databronnen, …), als voor het dataniveau (gemakkelijke beschikbaarheid en

toegankelijkheid van de data, zowel op het hoogste niveau van de index als op het basisniveau van de individuele variabele).

Een belangrijk punt is bovendien de analyse-eenheid. 'Inter-nationale' rangschikkingen van publieke sectoren doen in de regel een vergelijking tussen naties, tussen landen dus. Ons federaal land zit in deze kwestie in een bijzondere situatie. Vlaanderen, Wallonië en Brussel zijn immers regio's met een verregaande autonomie en zijn bijgevolg ook op zoek naar mogelijkheden ter vergelijking en positionering. In eerste instantie zullen we onderzoeken in welke mate de governance rankings van vandaag rekening houden met de verzuchtingen van regio's zoals Vlaanderen. Zoals verder geïllustreerd zal worden, worden regio's doorgaans gepasseerd in de governance rankings. In tweede instantie gaan we dan ook op zoek naar de redenen van de niet-opname van regio's.

Het resultaat van ons onderzoek naar de kwaliteit van governance rankings is een soort *ranking of the rankings*. Dit doen we enerzijds in algemene zin door de internationale kritische literatuur uit te pluizen. Anderzijds voeren we ook een eigen diepgravend empirisch onderzoek uit waarbij we een selectie van acht rangschikkingen van publieke sectoren beschrijven en beoordelen aan de hand van een waaier aan samengebrachte kwaliteitscriteria. Let wel; het gaat om een *'ranking of the rankings'* in de enge zin van het woord. Een van de zaken die we de governance rankings immers verwijten is dat ze ten alle prijzen hun vergelijkingsbevindingen willen aggregeren om er een uni-dimensionale rangschikking uit te puren. Als conclusie kunnen ze dan stellen dat "land A over een beter presterende overheid beschikt dan land B". Dit is echter een te simplistische weergave van de werkelijkheid en doet de onderliggende verschillen tekort. Onze benadering zal er dan ook een zijn die de relatieve en absolute plus- en minpunten van de rankings tracht te identificeren, om die overzichtelijk mee te geven. We trappen daarbij dus niet in onze eigen val door een rangorde van de onderzochte rankings op te maken.

Het vergelijken van diverse internationale vergelijkingen van publieke sectoren is geen gemakkelijke taak. Er moet vertrokken worden van de vragen 'hoe moet een ideale comparatieve meting of rangschikking er uitzien?' en 'hoe moet een ideale rangschikking van publieke sectoren er uitzien?'. Aangezien het gebrek aan eensgezindheid bij dergelijke vraagstukken net een van de basisproblemen is voor het uitblijven van

kwaliteitsvolle governance rankings, is dit in de praktijk een uiterst moeilijke en genuanceerde opdracht. Maar zoals Hood, Beeston en Dixon in een vergelijkbare oefening, genaamd *'Rating the rankings: assessing international rankings of public sector performance*' (Hood et al., 2007), stellen: "*While ranking a set of rankings might be thought to be a demanding task, it is not obviously more so than the ranking of abstract qualities like governance or some of the more polyvalent aspects of public services like crime control or health care."*

> *2.3.5. Hanteerbaarheid voor het (Vlaamse) beleid?*

Naast de politiek-administratieve component -waarin we de vraag stellen naar 'wie er meet?', 'waarom?' en 'voor wie'- de conceptuele component -die de validiteit en de aard van de indicatoren nagaat- en de statistisch-technische component -waarbij we de betrouwbaarheid van de data en het analyseproces onder het licht houden- trachten we ook de praktische bruikbaarheid, de hanteerbaarheid van governance rankings voor het beleid na te gaan.

Een rangschikking van publieke sectoren kan dan wel nog zo inherent kwalitatief en betrouwbaar zijn, ze moet ook voldoen aan de noden van de gebruikers. We stellen een onderzoek in naar de mate waarin een deelgroep van de potentiële gebruikers, de beleidsmakers, de bestaande governance rankings kunnen gebruiken voor hun beleidsevaluatie, beleidsformulering en andere overheidsmanagementstaken. De Groot (2004) ziet internationale rangschikkingen van publieke sectoren in de eerste plaats immers als een managementinstrument dat moet bijdragen aan het vergelijken van de prestaties, het analyseren van beter presterende landen en het trekken van aanbevelingen uit de *best practices* met als doel het verhogen van de eigen performantie. Leereffecten vormen volgens hem een essentieel kenmerk van een goede internationale rangschikking van publieke sectoren. De bruikbaarheid van de governance rankings is echter vooralsnog een onderbelicht onderwerp.

Het gaat in deze kwestie ook om de zogenaamde '*actionability*' van de indicatoren die gebruikt worden in de (internationale) metingen van publieke sectoren. Een van de kritieken die steevast terugkwamen in de jaren '90 was dat de indicatoren te veel een louter beschrijvend karakter vertoonden, zonder dat slecht presterende overheden er ook gebruik van

konden maken in hun poging om hun score op te krikken. Als reactie op deze kritiek zijn enkele auteurs en producenten op zoek gegaan naar *second-generation*, *action-oriented* of *actionable indicators*, die landen een instrument moeten bieden om hun prestaties te verbeteren (Knack et al., 2003a; Public Governance and Territorial Development Directorate OECD, 2005; Arndt, 2007).

In het bijzonder zullen we in deze studie trachten na te gaan in welke mate de internationale rangschikkingen bruikbaar zijn voor het Vlaamse niveau. Zo heeft de Vlaamse Gemeenschap een autonome bevoegdheid op het vlak van bestuurszaken. Het zou dus interessant zijn voor onze regio om te kunnen inschatten hoe we internationaal scoren op bijvoorbeeld *Public Administration* of de efficiëntie van de overheid. De initiële vraag luidt dus: in welke mate wordt er door producenten van internationale rangschikkingen van publieke sectoren omgekeken naar het soort autonome regio's waarvan Vlaanderen deel uitmaakt. Daar we ons op dit vlak aan een somber plaatje verwachten, proberen we ook te achterhalen wat de bezwaren en (praktische) moeilijkheden zijn om Vlaanderen in de governance rankings op te nemen.

In het verlengde hiervan doen we ook een aanzet tot een betere meting vanuit het standpunt van de Vlaamse overheid. Vlaanderen wil zich immers maar al te graag spiegelen aan andere regio's om op zich op die manier internationaal te positioneren. Bovendien is het de vaak uitgesproken ambitie van Vlaanderen om tot de topregio's te behoren. Deze doelstelling kan verder nagestreefd worden als men enerzijds weet waar Vlaanderen momenteel staat ten opzichte van de benchmarks en door anderzijds lessen te trekken van beter presterende regio's. Door leereffecten te creëren uit de internationale vergelijkingen van publieke sectoren, kan het ook verder werken aan het verhogen van de efficiëntie in de Vlaamse overheid.

> *2.3.6. Hoe groot is de kloof tussen vraag en aanbod?*

Onze laatste doelstelling is het opstellen van een klovenanalyse. In deze klovenanalyse willen we de standpunten van producenten en gebruikers ten opzichte van elkaar afzetten. Doel is om een zicht te krijgen op de kloof die er gaapt tussen de verwachtingen, noden en wensen van de gebruikers van internationaal comparatieve metingen van publieke sectoren enerzijds en de drijfveren van de producenten van governance rankings.

In dit onderzoek willen we een beginnende aanzet doen naar de kloofbeschrijving. Op basis van enkele gesprekken binnen de administratie lijsten we enkele prioritaire punten op die de Vlaamse beleidsmaker zoekt in een geschikte internationale rangschikking van publieke sectoren. Zo schetsen we een beginnend beeld van de meting waar Vlaanderen naar op zoek is. Aan de hand van communicatie met enkele internationale organisaties die governance rankings produceren aan de andere kant, trachten we de beweegredenen van de producent in beeld te brengen bij de productie van internationale rangschikkingen van publieke sectoren. We confronteren hen ook met de wensen op het Vlaamse niveau en trachten uit te zoeken waarom regio's zo moeilijk ingang vinden in de governance rankings.

Het tegenover elkaar plaatsen van beide standpunten, wensen en visies moet een aanzet bieden tot een betere afstemming van vraag en aanbod. Beide partijen kunnen er interessante lessen uit trekken. Het heeft bovendien de potentie om een basis te vormen voor een dialoog die zowel de producent, als de consument van governance data ten goede kan komen.

3. Onderzoeksdesign

In dit hoofdstuk schetsen we de onderzoekswijze die gehanteerd wordt in deze studie. Nadat we in het vorige hoofdstuk de doelstellingen en onderzoeksvragen geponeerd hebben, gaan we hier dieper in op het onderzoeksdesign, de manier waarop we het onderzoek naar de kwaliteit van internationale rangschikkingen van publieke sectoren opvatten.

De eerste paragraaf behandelt het onderzoeksopzet. Het beschrijft de structuur die we aan dit onderzoek hebben gegeven. De studie berust op drie grote pijlers.

In een eerste pijler doen we een grondige studie van de bestaande kritische literatuur omtrent de governance rankings. Er is immers een groeiende groep auteurs die zich bezig houdt met de kwaliteitsproblematiek rond de governance rankings. Uit het literatuuronderzoek extraheren we de grote lijnen van kritiek die bij meerdere internationale auteurs aan bod komen. Er is immers nood aan een synthese van de problemen van internationale rangschikkingen van publieke sectoren. We trachten de punten van kritiek ook te illustreren aan de hand van concrete voorbeelden.

De tweede pijler omvat ons empirisch onderzoek waarin we acht vooraanstaande, hoogaangeschreven en frequent gebruikte internationale rangschikkingen van publieke sectoren aan een diepgaande analyse onderwerpen. Dit gebeurt aan de hand van meta-fiches die in hoofdstuk 5 op inhoudelijk vlak worden beschreven. In deze paragraaf geven we een overzicht van het stappenplan dat we het afgelopen jaar hebben doorlopen om tot de uitgewerkte meta-fiches te komen.

In de derde pijler doen we tot slot een aanzet tot klovenanalyse. Aan de hand van peilingen bij zowel beleidsmakers op het Vlaamse niveau, als producenten uit internationale organisaties, trachten we de noden en wensen van de gebruikers af te zetten ten opzichte van de doelstellingen en voorkeuren van de producenten. Dit kan een eerste stap vormen in de toenadering tussen beide partijen.

De tweede paragraaf beschrijft het onderzoeksdoel. Hier kaderen we het onderzoek in de algemene problematiek rond de kwaliteit van

internationale rangschikkingen van publieke sectoren. We duiden ook de oorsprong van dit onderzoek in de Vlaamse overheid, waar de nood bestaat om tot een differentiatie te komen tussen de verschillende governance rankings. Het onderzoek wil samengevat een antwoord bieden op drie hoofdvragen. Ten eerste wil deze studie Vlaanderen een zicht bieden op (het gebrek aan) de kwaliteit en de (on)bruikbaarheid van rankings van bestuur.

Vervolgens luidt de tweede doelstelling om met deze kwaliteitsbeoordeling het maatschappelijke debat te voeden. Het debat omtrent de kwalitatieve gebreken van de governance rankings moet een aanzet geven tot betere comparatieve metingen.

Ten derde willen we binnen deze studie de ideologische achtergrond van rangschikkingen van publieke sectoren blootleggen. Ook de objectieven achter de rankings en het door de producent beoogde doelpubliek -kortom de administratief-politieke component- moeten in de analyse betrokken worden.

In paragraaf 3 tot slot leggen we het onderzoekskader vast. Enerzijds illustreren we de methodologie die in deze studie gebruikt wordt. Anderzijds schetsen we het theoretisch kader dat aan de basis ligt van onze kwaliteitsstudie van internationale rangschikkingen van publieke sectoren.

> 3.1. Onderzoeksopzet

Het project 'Internationale vergelijkingen vergeleken: Een meta-analyse van rangschikkingen van publieke sectoren' kadert in de nood aan internationale benchmarking voor Vlaanderen. De internationale rangschikkingen van publieke sectoren zijn daartoe het passende instrument. Vooral het onderdeel van de meting dat zich bezig houdt met de *public administration*, de werking van de overheid, wekt onze interesse. Het onderzoeksontwerp stoelt op drie stappen. Deze drie fases vormen samen het onderzoeksplan

Ten eerste doen we beroep op het onderzoekswerk dat diverse internationale auteurs al verricht hebben in de voorbije jaren. Een synthetiserend literatuuronderzoek moet aan het licht brengen welke problemen daarbij werden gesignaleerd.

Ten tweede zullen we ook zelf de analyse van enkele rankings uitvoeren. Daarvoor selecteerden we een achttal cases. Aan de hand van een zelfontworpen analyse-instrument -de meta-fiche- zullen we deze internationale rangschikkingen van publieke sectoren onderwerpen aan de toets van een reeks beschrijvende en beoordelende kwaliteitscriteria. Deze empirische studie moet een extra dimensie aan het bestaande onderzoek naar de kwaliteit van de rankings toevoegen. Met name het opnemen van een politiek-administratieve dimensie is innovatief ten opzichte van de internationale kritische literatuur, maar ook andere aspecten zoals de validiteit, de betrouwbaarheid en de transparantie worden in onze analyse van de acht governance rankings meegenomen.

Ten derde doen we een aanzet tot klovenanalyse. We trachten in deze fase de verschillende visies, noden en wensen op de (virtuele) markt van de governance rankings te analyseren. Aan de ene kant staan de consumenten, de gebruikers. Zij hebben bepaalde verwachtingen van een goede comparatieve meting en rangschikking van publieke sectoren op internationale schaal. Bij deze gebruikers verlenen we vooral een stem aan het Vlaamse beleidsniveau. Aan de andere kant van de lijn stellen we een onderzoek in naar de beweegredenen en doelstellingen van de producenten van internationale rangschikkingen. Het uitgangspunt hierbij is de vraag waarom de internationale organisaties, adviesbureaus en andere producenten van governance data niet volledig kunnen of willen inspelen op

de noden van de gebruikers in de algemene zin en van de Vlaamse administratie in concrete zin.

> *3.1.1. Exploratie van kritische literatuur*

Een eerste werkpakket omvat het exploreren van de stand van zaken in de internationale literatuur over de internationale rangschikkingen van publieke sectoren en prestatiemetingen in diverse organisatiestructuren. Zoals hoger beschreven, heeft een rits auteurs zich het voorbije decennium beziggehouden met het kritisch bestuderen van de diverse internationale governance rankings.

Door de kritische literatuur over de governance rankings er op na te slaan, kunnen we een overzicht opstellen van de kwaliteitsproblemen die door diverse auteurs worden aangekaart. Dit heeft twee verschillende doelstellingen. Enerzijds is het in kaart brengen van de problematiek rondom overheidsmetingen, governance data en internationale rangschikkingen van publieke sectoren een doelstelling op zich. Dat governance rankings nog met kwaliteitspijntjes zitten, is al langer duidelijk, het ontbrak echter nog aan een synthese van deze hele problematiek.

Anderzijds nemen we de kwaliteitsproblematiek zoals die in de internationale literatuur wordt beschreven als een beginpunt voor de aanmaak van onze meta-fiches. De meta-fiches vormen het zwaartepunt van ons empirisch onderzoek. Acht governance rankings worden aan de hand van meta-fiches tot op het indicatorenniveau gedesaggregeerd en onderworpen aan een grondige analyse. Het literatuuroverzicht biedt enerzijds een aantal aanknopingspunten van criteria die we in de analyse kunnen gebruiken en legt ook de lacunes bloot die er nog zijn in de kritische analyse van internationale rangschikkingen van publieke sectoren.

> *3.1.1.1. Synthese kwaliteitsproblematiek*

Een literatuurstudie vormt naar goede wetenschappelijke gewoonte een basis voor een onderzoek als deze. Een scan van de artikels, boeken en papers die op internationaal en nationaal niveau geschreven zijn over de problematiek van de governance data en rankings levert immers een goede synthese op van de kwaliteitsproblematiek. In het volgende hoofdstuk maken we een *roundup* van de kwaliteitsproblemen die in de literatuur zijn

beschreven. Hoofdstuk 4 baseert zich echter niet alleen op de eerder verschenen kritische werken, het wordt ook in belangrijke mate aangevuld, geïllustreerd en soms zelfs tegengesproken onze eigen analytische bevindingen.

De literatuurstudie heeft als betrachting ons wegwijs te maken in de bestaande problematiek. Met dit uitgangspunt in gedachte, is het objectief om een synthese van de kwaliteitsproblematiek te maken aan de hand van een literatuurstudie tevens het beginpunt van het ontwerpen van de meta-fiche voor de aansluitende empirische analyse.

De fase van het schetsen van de kwaliteitsproblemen sluiten we daartoe af met het oplijsten van potentiële kwaliteitscriteria die in de meta-fiches op de internationale rankings kunnen worden toegepast. De tekortkomingen die in de literatuur reeds als problematisch werden omschreven mogen immers in onze analyse niet ontbreken. Sterker nog, de kritieken die in de internationale literatuur opduiken met betrekking tot de internationale rangschikkingen van publieke sectoren zullen als uitgangspunt dienen in onze beschrijving, analyse en beoordeling van de internationale rangschikkingen van publieke sectoren. Een bundeling en structurering van de kwaliteitscriteria zal leiden tot de gestandaardiseerde meta-fiche.

> *3.1.2. Empirisch onderzoek*

De synthese van de kwaliteitsproblematiek als eerste fase in dit onderzoek, wordt gevolgd door een empirisch onderzoek. We selecteren een aantal voorname en weloverwogen cases van internationale rangschikkingen van publieke sectoren die we vervolgens grondig gaan analyseren. Wat maken als het ware een vergelijking van internationale vergelijkingen van publieke sectoren op, '*a ranking of the rankings*'. Hood, Beeston en Dixon (2007) deden in hun *'Rating the rankings*' een soortgelijke oefening, weliswaar beperkter in omvang.

Hood, Beeston en Dixon stellen dat er drie mogelijke pistes bestaan om internationale rangschikkingen van publieke sectoren kwalitatief of kwantitatief te vergelijken. Deze methodes komen overeen met de verschillende modaliteiten die producenten van governance rankings hanteren om publieke sectoren met elkaar te vergelijken. Ten eerste zouden we internationale experten kunnen vragen om de verschillende

rankings te beoordelen. Deze methode is gebaseerd op de *expert assessments* die in vele internationale rangschikkingen van publieke sectoren worden toegepast.

Ten tweede zouden we kunnen opteren voor het quoteren van de verschillende rankings op het vlak van relatief 'objectieve' criteria zoals aantal citeringen of positieve referenties in de media. Dit zou neerkomen op een soort *popularity test*. Onrechtstreeks hebben we deze methode toegepast bij het selecteren van de acht rangschikkingen van publieke sectoren die in ons onderzoek geanalyseerd worden. Voor de analyse opteerden we er immers voor om acht governance rankings uit te kiezen. We focusten ons daarbij op de rankings van bestuur, die een kwantitatieve beoordeling en vergelijking opmaken van de kwaliteit van het overheidsapparaat. De uiteindelijke selectie was ingegeven worden door de mate waarin de governance ranking alom bekend is en gebruikt wordt. Hierbij kijken we onder andere naar de naam en reputatie die de rangschikkingen op internationaal niveau hebben. Daarnaast is het ook belangrijk dat de rangschikking wereldwijd veelvuldig gebruikt wordt. Ten slotte speelde het ook in het voordeel van de ranking als ze frequent geciteerd wordt in de internationale literatuur.

We geven hier de uiteindelijke lijst van de internationale rangschikkingen van publieke sectoren al mee die het resultaat is van de hoger beschreven selectie (zie kader 3). De individuele rankings worden verder toegelicht in hoofdstuk 5.

Kader 3: Selectie van governance rankings voor de meta-analyse

- The Worldwide Governance indicators (World Bank Institute)
- The World Competitiveness Yearbook (International Institute for Management Development)
- The Global Competitiveness Index (World Economic Forum)
- The Corruption Perceptions Index (Transparency International)
- The Country Risk Briefing (Economist Intelligence Unit)
- The Country Policy and Institutional Assessments (World Bank)
- The International Country Risk Guide (PRS Group)
- The Public Sector Performance (European Central Bank)

De derde methode bestaat erin om de rankings zelf te beoordelen op een aantal gekozen criteria. In feite komt onze methode overeen met deze derde beoordelingsmodaliteit. Enkel hebben we ervoor geopteerd geen kwantitatieve score te geven op de verschillende kwaliteitscriteria van de rankings, maar voor elk criterium een kwalitatieve beschrijving te voorzien. Aan de hand van de kwalitatieve beschrijvingen zijn in tweede instantie wel beoordelingen te koppelen. We opteren ervoor als conclusie voor elke ranking enkele positieve en negatieve punten op te sommen, om zo een kwalitatieve differentiatie mogelijk te maken.

We willen in onze analyse vooral niet inboeten op grondigheid. Dit houdt in dat we de rankings zullen ontrafelen tot op indicatorenniveau. Daarnaast geven we een beschrijving en beoordeling van de governance ranking in zijn geheel op basis van de geselecteerde kwaliteitscriteria. Aan de hand van diezelfde kwaliteitscriteria zullen we de dimensie, component of indicator onderzoeken die specifiek het *public administration*-gedeelte meet. Tot slot zullen we ook alle indicatoren die deel uitmaken van de meting van overheidskwaliteit onder de loep nemen.

Hood, Beeston en Dixon (2007) wijzen er tot slot nog op dat er een gevaar van oneindige regressie bestaat in het vergelijken van vergelijkingen. '"*Quis custodiet ipsos custodes*', wie vergelijkt de vergelijking van de vergelijkingen?".

In wat volgt geven we een overzicht van het de verschillende stappen die we in de analyse hebben doorlopen om tot de meta-fiches te komen. We onderscheiden vier grote fasen.

> *3.1.2.1. Extractie kwaliteitscriteria*

De eerste fase -de extractie van kwaliteitscriteria- sluit direct aan bij het literatuuroverzicht. De synthese van de kwaliteitsproblematiek -die het einddoel van de literatuurstudie betrof- vormde zo de ideale basis voor het selecteren van potentiële kwaliteitscriteria.

De oplijsting van al de mogelijke kwaliteitscriteria werd nadien gestructureerd in componenten die vervolgens nog in hogerliggende clusters werden onderverdeeld. Deze clusters komen overeen met de drie

'meetagenda's' die volgens Van Dooren (Van Dooren et al., 2007) moeten convergeren om van een kwaliteitsvolle meting te kunnen spreken.

Volgens deze studie kan vooruitgang in het meten enkel geschieden, indien de volgende drie aspecten in acht worden genomen:

- De conceptuele agenda: Deze cluster onderzoekt de vraag 'wat moet er gemeten worden?'. Het moet duidelijk zijn wat de producent van de meting wil meten. Deze invulling moet zich ook weerspiegelen in de samenstelling van de meting (indicatoren, weging, aggregatieniveau); met andere woorden: 'Is wat men meet ook in overeenstemming met wat men beweert te meten?'.
- De technische agenda: Dit aspect peilt naar de vraag 'hoe moet er gemeten worden?'. Het gaat hierbij om de statistische kant van de meting. De berekeningswijze van de variabelen, de significantie, causale relaties en andere statistisch-technische aspecten komen hier aan bod.
- De politiek-administratieve agenda: Wie ligt aan de basis van de meting? Wat zijn de doelstellingen van de meting? Waar moet het doelpubliek gesitueerd worden? Deze schijnbare randaspecten blijken een grote invloed te hebben op de invalshoek en betekenis van de meting.

Binnen deze drie dimensies hebben we 33 beschrijvende kwaliteitsindicatoren geïdentificeerd, die op hun beurt in 14 componenten zijn onderverdeeld. Deze lijst van basisvariabelen hebben we in eerste instantie uitgetest in de testfase. De proefdraai moest technische en praktische moeilijkheden aan het licht brengen en leidde tot de gestandaardiseerde meta-fiche.

> *3.1.2.2. Proefdraai*

Om de lijst van potentiële kwaliteitscriteria te stroomlijnen tot een gestandaardiseerde meta-fiche, voerden we een testfase in waarin we een proefdraai uitvoerden op een representatieve en geschikte testcase. Het uiteindelijke doel van deze test was een verdere optimalisering van het scoringssysteem.

De potentiële kwaliteitscriteria die een uitloopsel vormden van de synthese van de kwaliteitsproblematiek, werden op die manier uitgeselecteerd tot een consistente en logisch gestructureerde meta-fiche. Het was niet onze ambitie een exhaustieve lijst van kwaliteitscriteria samen te stellen. Wel

werden al de potentiële eigenschappen, kenmerken en kwaliteitsitems die uit de literatuur voortvloeiden, opgenomen in de lijst. Dit had tot gevolg dat er een grote en diverse waaier aan kwaliteitsaspecten van internationale rangschikkingen van publieke sectoren aan de start van de proefdraai stond.

Voor de proefdraai hadden we een testcase nodig. Binnen de selectie van acht governance rankings, moest er een uitverkoren worden om de testfase te voltrekken. Wij hebben gekozen voor de Worldwide Governance Indicators van het World Bank Institute. Deze keuze was echter weloverwogen. De Worldwide Governance Indicators worden algemeen aanzien als de governance ranking met de grootste, breedste en meest diepgaande governance data die er bestaat. Zo aggregeren ze maar liefst 310 verschillende individuele variabelen tot 6 governance-indices. In totaal worden bovendien niet minder dan 212 landen in de ranking opgenomen.

De uitgebreide lijst van potentiële kwaliteitscriteria werd tijdens de proefdraai uitgezuiverd, aangepast en gestructureerd. De meeste items doorstonden de test en kregen een plaats in de gestandaardiseerde meta-fiche. De criteria die niet relevant genoeg werden bevonden in het kader van internationale rangschikkingen van publieke sectoren, die praktisch onhaalbaar bleken om na te gaan en de criteria die een duplicaat van een andere variabele zijn, werden niet weerhouden.

De proefdraai mondde uit in een gestructureerde en gestandaardiseerde meta-fiche. Dit is het analyse-instrument dat we zullen aanwenden om de kwaliteitskenmerken van de selectie van acht rangschikkingen van publieke sectoren te beschrijven, te beoordelen en te vergelijken.

> *3.1.2.3. Empirische datagaring*

Eens het scoringssysteem en de daarbij horende meta-fiches op punt stonden, kon in een derde fase de aanvulling van de meta-fiches gebeuren op basis van de verzamelde informatie over de verschillende internationale rangschikkingen van publieke sectoren en de achterliggende initiatiefnemende organisaties. Het einddoel van deze fase was de finalisering van acht meta-fiches, die als bijsluiters moeten dienen bij de governance ranking waar ze betrekking op hebben.

De invulling van de respectievelijke meta-fiches gebeurde op basis van diverse informatieve stukken. In eerste instantie zijn er de primaire bronnen. Deze bronnen zijn uiteindelijk het belangrijkst, direct en meest relevant voor de kwalitatieve inschatting van de governance rankings. Het gaat hierbij om een resem van stukken. In het multimediale tijdperk waarin we ons bevinden, worden de governance data en rankings vaak via diverse kanalen en toepassingen aangeboden en van begeleiding voorzien.

Enerzijds zijn er de jaarlijkse rapporten die meestal een *update* van de rangschikking bevatten, alsook de opgefriste data en de methodologie van de meting van publieke sectoren. Het gaat hier dus om de rankings zelf, eventueel aangevuld met onderliggende data en onderzochte (basis)indicatoren.

Doordat meerdere rangschikkingen van publieke sectoren enkel tegen (relatief hoge) betaling in hun geheel kunnen geraadpleegd worden, hebben we niet altijd de mogelijkheid gekregen om de rangschikkingen van publieke sectoren tot op de basis te verkennen. Dit heeft ons echter niet al te veel parten gespeeld, aangezien meestal enkel de beschikbaarheid van de data (tot op het basisniveau) daardoor de weg werd versperd. Wij focussen ons echter eerder op het systeem achter de cijfers dan op de datascores zelf, wat deze praktische hindernis niet problematisch maakte voor deze studie.

Aan de andere kant hebben we de hoge prijs die door sommige producenten wordt gevraagd voor de toegang tot data uit hun governance ranking wel meegenomen in de beoordeling van de databeschikbaarheid van de internationale meting van publieke sectoren. Het simpele feit dat gebruikers, maar ook overheden -denk maar aan Derde Wereldlanden- soms aanzienlijke sommen moeten neertellen om dergelijke data te bekomen, beperkt onzes inziens de databeschikbaarheid en de transparantie en hypothekeert daarbij ook de leereffecten.

Vaak wordt het hoofdrapport nog aangevuld of verder verduidelijkt met informatie op de website. De elektronische toepassingen spelen een belangrijke rol in de hedendaagse diffusie en voorstelling van de governance data. Meer en meer worden interactieve toepassingen voorzien om de gebruiker op een zo overzichtelijk mogelijke manier de data voor te schotelen. Bovendien kan de user in veel gevallen gepersonaliseerde meetsystemen en rankings samenstellen.

Ook interessant zijn de begeleidende papers die veelal door auteurs van governance rankings of door aan de producent gelinkte onderzoekers worden opgesteld. Deze papers bieden vaak extra inzicht in de methodologie, in de basistheorie achter de ranking en in de conclusies van de resultaten. Het gaat om onmisbare bronnen in de analyse van de rankings, aangezien de hoofdrapporten omwille van marketingoverwegingen meestal niet overlopen van details, nuances en tekortkomingen.

Naast de primaire bronnen, putten we in tweede instantie ook uit de secundaire bronnen. Hier gaat het om onafhankelijke kritische studies die onderzoek instellen naar de kwaliteit van internationale rangschikkingen van publieke sextoren en daarbij concrete analyses hebben verricht op een of meerdere van de door ons geselecteerde rankings. Hieronder geven we een overzicht van een aantal auteurs die zich hebben ingelaten met het kritisch bestuderen van één of meerdere rangschikkingen van publieke sectoren die ook in onze selectie voorkomen (zie Kader 4).

Kader 4: Secundaire bronnen governance rankings

Governance Ranking	Producent	Secundaire literatuur
Worldwide Governance Indicators	WBI	– (Van Roosbroek & Van Dooren, 2006) – (Van de Walle, 2005) – (Arndt & Oman, 2006) – (Knack, 2006) – (Kurtz & Schrank, 2005) – (Thomas, 2007) – (Boyle, 2007) – (Van de Walle, 2006a) – (Besançon, 2003) – (Rothstein & Teorell, 2005) – (Knack & Kugler, 2002) – (Arndt, 2007; Hood et al., 2007)

		– (Moesen & Cousserier, 2006)
World Competitiveness Yearbook	IMD	– (Van de Walle, 2005) – (Boyle, 2007) – (Van de Walle, 2006a)
Global Competitiveness Index	WEF	– (Van Roosbroek & Van Dooren, 2006) – (Van de Walle, 2005) – (Van de Walle, 2006a) – (Hood et al., 2007) – (De Witte & Moesen, 2007)
Corruption Perceptions Index	TI	– (Van Roosbroek & Van Dooren, 2006) – (Van de Walle et al., 2004) – (Arndt & Oman, 2006) – (Kuha & Gaskell, 2005) – (Galtung, 2005) – (Killias, 2005) – (Hood et al., 2007) – (Schmitz, 2007) – (Demidov, 2006)
Country Risk Briefing	EIU	– (Arndt & Oman, 2006)
Country Policy and Institutional Assessment	WB	– (Arndt & Oman, 2006) – (Schmitz, 2007)
International Country Risk Guide	PRS Group	– (Arndt & Oman, 2006)
Public Sector Performance	ECB	– (Van Roosbroek & Van Dooren, 2006) – (Van de Walle et al., 2004) – (Van de Walle, 2005) – (Van de Walle,

		2006a) – (Moesen & Cousserier, 2006)

Legende:

- WBI: World Bank Institute
- IMD: International Institute for Management Development
- WEF: World Economic Forum
- TI: Transparency International
- EIU: Economist Intelligence Unit
- WB: World Bank
- PRS Group: Political Risk Services Group
- ECB: European Central Bank

Naast de primaire en secundaire bronnen hebben we ons ook geïnformeerd bij de producenten zelf. Aan de hand van telefonische contacten, selectieve interviews en mailverkeer met diverse internationale organisaties en private adviesbureaus, hebben we getracht in bepaalde kwesties meer helderheid te scheppen. Het bood ons de mogelijkheid om meer duidelijkheid te brengen in de voorhanden zijnde data en begeleidende informatiebronnen. Bovendien konden we op die manier gerichte vragen en bemerkingen op directe wijze met auteurs en producenten van de governance rankings communiceren.

> *3.1.2.4. Synthese onderzoeksresultaten*

De vierde en laatste stap die we dienden af te leggen in ons empirisch onderzoeksplan, omvatte het in kaart brengen van de onderzoeksbevindingen. De meta-fiches bevatten in hun gefinaliseerde versie meer dan 50.000 beschrijvende cellen die de rankings, de openbaar bestuurscomponenten en de individuele indicatoren beschrijven. Zij vormen het eindproduct van de analyse.

Aangezien de gefinaliseerde meta-fiches in geprinte versie meerdere vierkante meters papier in beslag zouden nemen, hebben we ervoor gekozen om de meta-fiches in hun volledige versie niet in dit rapport te omsluiten. In dit rapport trachten we echter de belangrijkste, meest relevante en best geïllustreerde bevindingen mee te geven. Hoofdstuk 5 zal in grote mate geïllustreerd worden aan de hand van de meta-fiches.

In een annexbundel die we aan het rapport toevoegen, zullen we de evaluatie van het onderdeel dat de *Public Administration*-component meet, omsluiten. Op die manier kunnen we de doelstelling realiseren om de meta-fiches in te zetten als cruciaal achtergrondinstrument bij het inhoudelijk beoordelen van data en conclusies van de jaarlijks wederkerende governance rankings. Het kan ook als hulpmiddel bij toekomstig wetenschappelijk onderzoek gebruikt worden. Bovendien kan het als bijsluiter bij internationale rangschikkingen van publieke sectoren dienen ten behoeve van beleidsmakers, onderzoekers en media.

De vastgestelde mankementen gebruiken we ook als aanknopingspunt voor het meegeven van concrete aandachtspunten voor de gebruikers teneinde tot een correctere interpretatie van de governance rankings te bekomen. Ook voor de producenten trachten we enkele aanbevelingen uit de meta-fiches te destilleren. We willen de producenten op hun verantwoordelijkheid wijzen om de kwaliteit, transparantie en betrouwbaarheid van hun governance rankings hoog te houden. Aan de hand van enkele voorstellen ter promotie van de kwaliteit, hopen we een bijdrage te leveren aan de kwaliteitsverbetering van de huidige internationale rangschikkingen van publieke sectoren en de kloof tussen producenten en gebruikers kleiner te maken.

> *3.1.3. Klovenanalyse*

Na de twee kernfases van dit onderzoek, de literatuurstudie ter synthese van de kwaliteitsproblematiek van governance rankings en de analyse waarin we acht internationale rangschikkingen van publieke sectoren aan een diepgaand hebben empirisch onderzoek onderworpen, is er ook een kleiner derde onderdeel voorzien waarin we een klovenanalyse hebben uitgewerkt.

In deze klovenanalyse geven we een aanzet tot het in kaart brengen van de wensen, noden en standpunten aan beide zijden van vraag en aanbod op het vlak van comparatieve data over de werking van de overheid. Enerzijds wilden we inschatten wat de noden zijn bij de gebruikers op het vlak van comparatieve data en rangschikkingen van publieke sectoren. Welke categorieën van gebruikers zijn er en welke noden en wensen hebben zij? We hebben ons daarbij voornamelijk gefocust op de noden en wensen van

het beleid. Aan de hand van de literatuur en van selectieve gesprekken met vertegenwoordigers van de Vlaamse administratie, hebben we getracht ook meer specifiek aan te geven waar Vlaanderen naar op zoek is op het vlak van internationale vergelijkingen van publieke sectoren.

Aan de andere kant hebben we ook geprobeerd ons in te leven in het standpunt van de producenten. Commerciële, praktische of ideologische obstakels verhinderen immers in grote mate dat producenten volledig kunnen inspelen op de noden en wensen van de gebruikers. Aan de hand van enkele selectieve gesprekken met vertegenwoordigers van de producenten (internationale organisaties, private adviesbureaus en NGO's) hebben we ons dus een beeld proberen te vormen van de standpunten aan de aanbodzijde van governance rankings.

De betrachting van deze klovenanalyse is een betere afstemming te bewerkstelligen tussen de producenten en de consumenten van internationale rangschikkingen van publieke sectoren. De uittekening van de kloof tussen vraag en aanbod kan een basis vormen voor verdere toenadering tussen gebruikers en producenten. Meer specifiek dient Vlaanderen het initiatief op zich te nemen om met haar wensen en noden naar de internationale organisaties te trekken die zich bezighouden met het samenstellen van governance rankings.

> 3.2. Onderzoeksdoel

Met cijfers kan men veel bewijzen. Dat wist ook Benjamin Disraeli, Brits staatshoofd uit de 19e eeuw, al toen hij de beroemde uitspraak deed: "*There are three kinds of lies: lies, damned lies and statistics*". Deze zinsnede is nog altijd brandend actueel en komt er op neer dat uiterst voorzichtig moet omgesprongen worden met data en statistieken.

Doorgetrokken naar ons onderzoeksobject, de internationale rangschikkingen van publieke sectoren, levert dit de wijze les op om de conclusies uit de governance rankings niet zomaar over te nemen en de vergelijkingen over de landsgrenzen heen met een korrel zout te nemen.

De doelstelling van dit onderzoek is dan ook een beter inzicht in product en producent te verwerven opdat de pijnpunten in de rankings meegenomen kunnen worden in de interpretatie en het gebruik van deze beleidsinformatie. Deze studie beschrijft daartoe de problematiek van de gebrekkige kwaliteit van de internationale rangschikkingen, legt de zwaktes bloot en doet enkele suggesties voor een betere bruikbaarheid.

'*Nam et ipsa scientia potestas est*': kennis is macht. Een uitdrukking die ook in de discussies rond internationale rangschikkingen van tel is. Pas wanneer een blik achter de cijfers een kwalitatief oordeel oplevert van de aard van de internationale governance ranking, kan een diepgaande inhoudelijke dialoog starten over de conclusies van de rangschikking.

Opgesplitst in drie onderzoeksdoelstellingen betekent dit ten eerste dat we Vlaanderen een zicht willen bieden op de kwaliteit en beschikbaarheid -of moeten we zeggen onbeschikbaarheid- van internationale rankings van bestuur. Vlaanderen is op zoek naar comparatieve data en rangschikkingen om zich als regio in een snel groeiend Europa en bij uitbreiding op wereldvlak te positioneren. Het is echter de vraag of de governance rankings wel kwalitatief genoeg zijn om deze taak op zich te nemen. Bovendien stelt zich ook de vraag in welke mate internationale rangschikkingen in een wereld van naties een regio als Vlaanderen wijzer kunnen maken op het vlak van de kwaliteit van publiek sectoren.

Door de pijnpunten van de rankings aan de oppervlakte te brengen, willen we de gebruiker op het Vlaamse beleidsniveau een beter inzicht bieden in

de internationale rankings. Op die manier zijn beleidsmakers beter gewapend tegen ongenuanceerde uitspraken op basis van de governance data.

Ten tweede kan onzes inziens een kwaliteitsbeoordeling van internationale rangschikkingen van publieke sectoren het maatschappelijke debat verder voeden. In deze scharen we ons achter de bestaande wetenschappelijke kritische literatuur op dit vlak en pogen we via deze kritieken vooral de dynamiek te steunen om een betere meting van publieke sectoren op gang te brengen.

Ten derde zullen we ook nieuwe aspecten toevoegen aan de huidige kritische literatuur van governance rankings. We doen immers een poging de aard en de ideologische achtergrond van de internationale rangschikkingen van publieke sectoren te achterhalen. Dit is een aspect van governance rankings dat in het verleden onderbelicht is gebleven.

De politiek-administratieve dimensie van de producent -zoals we ze noemen- is nochtans een belangrijk onderdeel van de kwaliteit van de geproduceerde data. De politiek-administratieve kenmerken zitten vervat in de antwoorden op de vragen 'Wie is de *ranker*?', 'Wat zijn de doelstellingen van de producent?' en 'Voor wie zijn de data bedoeld?'.

> *3.2.1. Vlaanderen krijgt een zicht op de kwaliteit en (on)bruikbaarheid van rankings van bestuur*

Internationale rangschikkingen van publieke sectoren maken deel uit van de beleidsinformatie die België/Vlaanderen in staat stelt om zich op het internationale toneel te positioneren. Bovendien kunnen beleidsmakers op diverse niveaus leereffecten putten uit dergelijke internationaal-comparatieve data. '*Best practices'* kunnen dan vervolgens in eigen land geïntroduceerd worden om zo het beleid in positieve zin bij te sturen.

Deze trend heeft zich in Vlaanderen vooral de voorbije jaren stevig doorgezet. In 'Vlaanderen in actie', waarmee de Vlaamse regering Vlaanderen naar het koppeloton van regio's op wereldvlak wil brengen, wordt het als volgt verwoord: "In 1995 was de cijfermatige, statistische en kwantitatieve onderbouw van het Vlaams beleid nog een grote beleidsmatige lacune. Een efficiënte overheid dient echter te beschikken

over de nodige instrumenten om te kunnen vooruitblikken, keuzes te maken, deze uit te voeren en op te volgen, indien nodig bij te sturen. (Vlaamse regering, 2006)"

Omdat de werking van de overheid als een belangrijk aspect van een goed bestuurd land en een welvarende natie wordt beschouwd, kan het van bijzonder nut zijn om Vlaanderen te 'benchmarken' -internationaal te toetsen- aan wat er gebeurt in vergelijkbare regio's in Europa en de wereld. Voor economische performantie-indicatoren is dat sinds enige tijd reeds een gangbare praktijk. "Doordat steeds meer regio's en landen beschikken over gelijkaardig statistisch materiaal (*in het kader van de sociaaleconomische performantie, ML*), wordt het mogelijk om de cijferreeksen met elkaar te vergelijken en zo op een objectieve wijze de performantie van regionale economieën af te toetsen", luidt het in 'Vlaanderen in actie!' (Vlaamse regering, 2006). Of dit momenteel ook mogelijk is op het vlak van *Public Administration* hebben we in dit onderzoek verder uitgezocht. Het onderzoek 'Internationale vergelijkingen vergeleken: Een meta-analyse van de rangschikkingen van publieke sectoren' heeft immers onderzocht in welke mate rangschikkingen van publieke sectoren de mogelijkheid bieden om vergelijkingsoefeningen te maken op het vlak van de overheidskwaliteit. Deze vraag werd gesteld zowel vanuit het perspectief van een nationale overheid (België), als vanuit een regionaal oogpunt (Vlaanderen).

Onze studie heeft zo als doel een wetenschappelijke bijdrage te leveren aan het actuele debat over de internationale positionering van ons land België en onze regio Vlaanderen op het vlak van de prestaties van de overheid. Signalen uit de internationale literatuur en bevindingen uit onze empirische analyse noodzaken ons de vergelijkbaarheid van landen -en meer nog regio's- op het vlak van de overheidsperformantie te nuanceren. De rangschikkingen vertonen nog te veel zwakke punten om deze taak betrouwbaar op zich te nemen.

Door de kwaliteitsproblemen aan te geven, hebben we een poging gedaan een 'objectiever' beeld te geven van de governance data. Op die manier kunnen beleidsmakers in de Vlaamse overheid een beter begrip hebben van wat de cijfers wel en niet aantonen; kortom wat de data intrinsiek waard zijn.

Een van de vaststellingen die we hebben gedaan is dat Vlaanderen als regio in de regel geen ingang vindt in de grote internationale rangschikkingen van publieke sectoren. Het is voor de producenten van dergelijke rankings klaarblijkelijk niet interessant genoeg om Vlaanderen, noch andere regio's een afzonderlijke plaats toe te kennen in de rankings. Op het vlak van internationale vergelijkingen van publieke sectoren, blijft Vlaanderen dus nog op zijn honger zitten.

Er zijn welk enkele kleinschalige initiatieven geweest om Vlaanderen op internationaal vlak op het gebied van overheidsefficiëntie te positioneren, maar die draaiden ofwel uit op een mislukking ofwel waren ze een kort leven beschoren. Een voorbeeld van een dergelijk kortstondig experiment is de herberekening van de Belgische scores in de Global Competitiveness Index van het World Economic Forum naar Vlaamse statistieken (Bowen et al., 2007). Het bleef echter bij een eenmalige oefening in opdracht van toenmalig Vlaams minister van Economie, Ondernemen, Innovatie, Wetenschap en Buitenlandse Handel, Fientje Moerman (VLD).

De oefening bleek immers te tijdrovend en bovendien erg complex, doordat er voor Vlaanderen vaak geen gelijkaardige data voor handen bleken te zijn. Er werd dan ook beslist geen vervolg te knopen aan deze oefening. Bovendien stamt de oefening nog van de tijd dat de Global Competitiveness Index 'slechts' 36 indicatoren telde. Ondertussen heeft de methodologie van het World Economic Forum al tweemaal veranderingen doorgemaakt, waardoor er momenteel drie maal zoveel indicatoren zijn opgenomen om tot de algemene index te komen.

> 3.2.2. Kwaliteitsbeoordeling voedt het maatschappelijke debat

Gelet op de grote aandacht die media, beleidsmakers en oppositie steevast aan de governance rankings besteden, was er grote behoefte ontstaan aan kwaliteitsonderzoek van dergelijke meetinstrumenten. Door mee te stappen in het kritische analyseproces van overheidsrangschikkingen, hebben we getracht aanbevelingen te formuleren voor producenten en consumenten van de data en op die manier een bijdrage te leveren aan betere meet- en vergelijkingsoefeningen op het vlak van de kwaliteit van de overheid.

Deze oefening moet niet alleen de interpretatie van de data faciliteren, ze voedt ook het maatschappelijke debat over overheidseffectiviteit. De

impliciete zoektocht naar de (hypothetisch) perfecte internationale rangschikking van publieke sectoren heeft ook halt gehouden bij de vraag welke aspecten die ideale governance ranking juist zou moeten meten. Dit komt neer op de vraag stellen: Wat maakt een overheid tot een goed presterende overheid? Het antwoord op deze vraag is politiek getint en kan niet losgezien worden van het kerntakendebat, waarin men zich de vraag stelt wat de overheid wel en niet op zich moet nemen. Als men het eens is over deze taakinvulling is de stap naar de manier waarop dat dan gemeten moet worden al een stuk kleiner.

Door een overzicht te geven van de kwaliteitsproblemen en de onopgeloste vraagstukken uit het huidige landschap van internationale rangschikkingen van publieke sectoren, willen we vooral een aanzet geven tot een verbetering van de situatie. Daar waar deze studie zich voornamelijk heeft toegelegd op het detecteren en beschrijven van de tekortkomingen in de internationale rangschikkingen van publieke sectoren, willen we als conclusie ook enkele aanbevelingen formuleren voor eventueel toekomstig onderzoek waarin nieuwe meetinitiatieven worden aangestuurd of uitgewerkt.

> *3.2.3. De studie legt de ideologische achtergrond van rangschikkingen van publieke sectoren bloot*

Op conceptueel vlak kampen alle internationale overheidsmetingen met grote gebreken. Niet alleen worden in de verschillende rankings andere concepten gemeten, het is vaak ook onduidelijk wat die concepten daadwerkelijk inhouden. Naast het feit dat niet altijd het juiste gemeten wordt, wordt ook vaak niet juist gemeten. Te kleine samplebevragingen, statistisch weinig zeggende vergelijkingen en ongestaafde normatieve keuzes in de constructie van de indicatoren voeden het wantrouwen tegenover de data. Bovendien worden inschattingen of opinies te vaak aangenomen als objectieve feiten. Deze en andere problemen worden in de meta-fiches in kaart gebracht onder de ‘conceptuele dimensie’ en de ‘technische dimensie’.

We willen echter een toegevoegde waarde leveren aan de huidige kritische literatuur betreffende governance rankings. In deze studie hebben we daartoe getracht ook de ideologische achtergrond van de rankings in beeld brengen. Door te kijken naar de aard van de producent, diens objectieven

en het doelpubliek dat ze voor ogen heeft, hebben we de ideologische achtergrond betrokken in het beoordelingsproces van de internationale rangschikkingen van publieke sectoren.

Een ranking of indicator opmaken zonder bias of normatieve overwegingen, is een haast onmogelijk opdracht. Het is zelfs niet zozeer noodzakelijk. Het grote probleem is echter dat de vertekening binnen de metingen meestal niet zichtbaar zijn waardoor de gebruiker er doorgaans geen rekening mee kan houden. Zo worden bijvoorbeeld bescherming van het milieu en arbeidsrechten in veel rangschikkingen aanzien als indicatoren met een negatief effect op 'goed beleid'. Je kunt er mee akkoord gaan of niet, maar het problematische ligt er hem in dat vele gebruikers zich niet bewust zijn van deze biases (Arndt, 2007). Daar zit hem het grootste gevaar.

> 3.3. Onderzoekskader

Dit project kadert in de onderzoekslijn 'beleidsinformatie' en behandelt de internationale rangschikkingen van publieke sectoren als een informatief beleidsinstrument. De these achter dit onderzoek is dan ook dat kwantitatieve elementen zoals data, indicatoren en rankings een meerwaarde kunnen opleveren voor een kwantitatief onderbouwd beleid. Vlaanderen kan zich optrekken aan de hand van *'best practices*' die uit internationaal comparatieve onderzoeken en gerelateerde rankings van publieke sectoren gedestilleerd kunnen worden. De kwaliteit van de aangereikte informatie -*in casu* de internationale rangschikkingen van publieke sectoren- moet wel van voldoende hoog niveau zijn vooraleer er gepaste en betekenisvolle beleidslessen uit kan getrokken worden.

Dit onderzoek past in de doelstelling om internationale vergelijkingen van publieke sectoren kwalitatief op een hoger peil te brengen. Het onderzoek is daartoe beschrijvend, beoordelend en aanbevelend. In eerste instantie hebben we immers de rangschikkingen van publieke sectoren doorgelicht en hebben we beschreven hoe de machinerie achter de cijfers in elkaar zit. Vervolgens hebben we ook een oordeel verbonden aan de kwaliteit en structuur van de onderzochte rankings door positieve en negatieve kenmerken uit de beschrijving te lichten. In laatste instantie hebben we geprobeerd vanuit die wetenschap een aantal aanbevelingen op te stellen die een optimalisering van de productie en het gebruik van governance data moeten voortbrengen.

We hebben daartoe gebruik gemaakt van zowel deductieve als inductieve onderzoeksmethoden. Enerzijds hebben we een synthese opgemaakt van bestaande theorieën, inzichten en analyses. Aan de hand van de gevestigde internationale kritische literatuur hebben we een overzicht opgesteld van de kwaliteitsproblemen die de internationale rangschikkingen van publieke sectoren in algemene of individuele zin worden verweten. Anderzijds zijn we ook inductief te werk gegaan. Aan de hand van wetenschappelijke observaties in onze empirische analyse hebben we getracht deze kwaliteitsproblemen te illustreren. De meta-fiches hebben als evaluatief instrument gediend in het beoordelen van de kwaliteit van een selectie van acht governance rankings die we onder de loep nemen.

Een eerste deel van dit onderzoek behandelt de opkomst, het gebruik en de problemen van de internationale rangschikkingen van publieke sectoren. Aan de hand van een grondig internationaal literatuuronderzoek zal een beeld geschetst worden van de groeiende interesse voor de kwaliteit van de overheid, die aanleiding heeft gegeven tot de forse stijging van het aantal geproduceerde governance rankings gedurende de laatste decennia.

Internationale rangschikkingen van publieke sectoren die over de landsgrenzen heen vergelijken maken, zijn nog een relatief nieuw fenomeen. Een kritische meta-analyse van de internationale rangschikkingen van publieke sectoren werd als een lacune in het onderzoek bestempeld. Dit project heeft geprobeerd in het verlengde van de bestaande individuele kritische analyses tot een soort van overkoepelende ‘rangschikking der rangschikkingen' te komen.

Een tweede deel van de kwaliteitsbeoordeling van de internationale rangschikkingen van publieke sectoren, voert een diepteanalyse uit van een selectie van acht hoogaangeschreven en welgekende governance rankings. Dit onderzoek is gebaseerd op een analyse-instrument ontworpen voor het beoordelen van die rangschikkingen. Voor het onderzoek werd een gestandaardiseerde meta-fiche gecreëerd die in een uitvoerige beschrijving en beoordeling van elke internationale rangschikking voorziet.

Het doel van deze analyse was om een beter beeld te krijgen op de kwaliteit van de verschillende internationale rangschikking. Door de bril van de meta-fiche werd er met een wetenschappelijk oog naar de verschillende rangschikkingen gekeken. Het uiteindelijke doel was een kwalitatieve uitspraak te doen over de activiteiten van de *rankers*. Met deze kwaliteitsbeoordeling in het achterhoofd, moet het gemakkelijker worden om de data in een juist perspectief te plaatsen.

> 3.3.1. Methodologie

In dit onderzoek zijn drie grote fases te onderscheiden waarin telkens een andere methodologie gehanteerd werd. Ten eerste is er de literatuurstudie waarin we een *quickscan* hebben gemaakt van de bestaande kritische literatuur over de productie en het gebruik van internationale rangschikkingen van publieke sectoren. Voor dit onderdeel hebben we ons

vooral gebaseerd op bestaande boeken, artikels en papers van wetenschappers, analisten, experten en andere auteurs op dit domein.

Ten tweede hebben we een empirische analyse uitgevoerd op een selectie van acht gerenommeerde internationale rangschikkingen van publieke sectoren. Voor deze analyse hebben we gebruik gemaakt van rapporten en websites van de organisaties die de rankschikkingen opmaken. In tweede instantie zijn we op zoek gegaan naar secundaire literatuur. We doen daarbij een beroep op bestaande kritische studies over de internationale rangschikkingen.

Ten derde hebben we ook een aanzet gegeven tot een klovenanalyse. Hierbij hebben we getracht de kloof te ontleden die er gaapt tussen enerzijds de noden en wensen onder gebruikers van governance rankings en anderzijds de doelstellingen en beweegredenen van de producenten achter dergelijke rangschikkingen. Om een eerste inzicht te krijgen in de standpunten van beide partijen, hebben we enkele selectieve interviews afgenomen langs beide kanten en elektronische communicatie opgebouwd met enkele organisaties die betrokken zijn bij de internationale rangschikkingen van publieke sectoren.

> *3.3.2. Theoretisch kader*

De relevantie van internationale rangschikkingen van publieke sectoren stoelt op de theoretische inzichten met betrekking tot 'beleidsinformatie'. Het vergelijken van overheidsprestaties valt onder de grote noemer van de beleidsanalyse (Donahue et al., 2000). De algemene visie achter deze theorie gaat er van uit dat prestatiemetingen en relatieve rangschikkingssystemen de overheid bijstaan in het nemen van beslissingen, het overwegen van beleidsopties en bij de verantwoording van het gevoerde beleid. In die zin spelen governance rankings een belangrijke rol in een kwantitatief onderbouwd overheidsbeleid.

In haar ambitieuze doelstelling om efficiënter en effectiever bestuur te leveren en uit te groeien tot een topregio in Europa en de wereld, tracht Vlaanderen haar beleid te kwantificeren en te benchmarken. "Door gebruik te maken van deze verzameling van interne, kwantitatieve indicatoren kan de vooruitgang van het beleid van de overheid en de effectiviteit ervan opgevolgd worden. Omgekeerd kan diezelfde overheid de doelstellingen

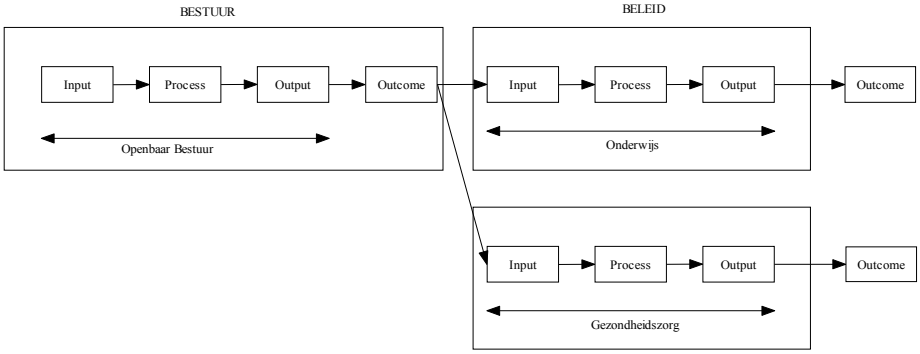
voortaan niet enkel kwalitatief formuleren, maar ook kwantitatief. Deze manier van werken zorgt voor een grotere beleidstransparantie en responsabiliseert de regering en haar administraties ten overstaan van burgers, ondernemingen en organisaties", luidt het in haar toekomstplan 'Vlaanderen in actie!' (Vlaamse regering, 2006).

Onze wetenschappelijke studie kadert dan weer in de internationale kritische literatuur die de huidige governance rankings aan een grondige analyse onderwerpt. Dergelijke kwalitatieve analyses zijn immers belangrijk indien internationale rangschikkingen van publieke sectoren een kwaliteitsvolle bijdrage als beleidsinformatie willen leveren.

Daartoe hebben we een meta-analyse gemaakt, een overkoepelende en alles samenbrengende studie van de kwaliteitsproblemen waar governance rankings mee te kampen hebben. We hebben ervoor geopteerd een selectie van de internationale rangschikkingen van publieke sectoren te overschouwen. We pikten er immers de rankings uit die landen quoteren en vergelijken op het vlak van bestuur, waarbij de kwaliteit van het overheidsapparaat centraal staat. Hierbij sloten we de rangschikkingen uit die een sectoriële inslag hebben en zich bijvoorbeeld focussen op onderwijs of gezondheidszorg.

In onderstaand schema (zie figuur 2) illustreren we dit onderscheid. Enerzijds is er de bestuurscomponent (*public administration*), anderzijds de beleidscomponent(en) zoals bijvoorbeeld de onderwijssector, of de gezondheidszorg. De outcome, de effecten die het openbaar bestuur voortbrengt, komen vooral tot uiting als input in de diverse beleidssectoren. We kijken naar de metingen van deze bestuurscomponent (rangschikkingen van publieke sectoren), de kwaliteit van de beleidsoutput wordt nagegaan in rankings van beleid.

Figuur 2: Bestuur/Beleid Schema



Enerzijds hebben we dus een synthese maken van de kritische literatuur die er in de voorbije jaren verschenen is over dergelijke rankings. Door de verschillende visies, kritieken en suggesties samen te brengen, hebben we op overzichtelijke wijze de grote lijnen van kritiek trachten te schetsen die bestaan ten aanzien van de rankings.

Daarnaast hebben we ook verder gebouwd op de bestaande kritische studies door een eigen empirische analyse te maken van internationale rangschikkingen van bestuur. Hiervoor hebben we acht gerenommeerde governance rankings geselecteerd, waarvan we de (deel-)meting van het bestuursapparaat er tekens uit hebben gelicht. Het analyse-instrument voor de kwaliteitsbeoordeling werd gevormd door onze meta-fiche. Deze fiche is een gestandaardiseerde beoordelingstabel waarin we een reeks kwaliteitscriteria toetsen aan de onderzochte rankings. Op die manier hebben we getracht een algemeen inzicht te verwerven in de kwaliteit van deze governance rankings.

We hebben de governance rankings beoordeeld op drie grote dimensies. De eerste is de conceptuele pijler. In dit onderdeel worden vragen gesteld naar de validiteit van de ranking: Wordt er in de rangschikking van publieke sectoren gemeten wat er beweerd wordt te meten?

We hebben in deze dimensie ook nadrukkelijk gekeken naar de indicatoren die aan de basis liggen van de prestatiemetingen. Deze indicatoren vormen immers de bouwstenen van de ranking. Bij gebrek aan welomlijnde

omschrijvingen van de gemeten concepten, wordt de inhoud van de meting voornamelijk bepaald door de keuze van indicatoren. De performantie van de overheid kan op verschillende wijzen beoordeeld worden. De indicatoren kunnen subjectief zijn of objectief. Ze kunnen bovendien peilen naar verschillende aspecten (taken) van een overheid.

Zo stelt Hood (1991) in zijn theoretisch werk dat de taken van de overheid in drie categorieën zijn te onderscheiden. Zo komt hij tot sigma-waarden, theta-waarden en lambda-waarden in de publieke sector. De sigma-waarden staan voor een doelgerichte en effectieve overheid. Het gaat om harde waarden die productiviteit en zuinigheid hoog in het vaandel dragen. Vervolgens identificeert hij ook theta-waarden, die eerder de zachte dienstverlenende kant van de overheid in de verf zetten. Integriteit, legitimiteit en transparantie staan hier voorop. Het benadrukt verder de overheidsfunctie om klantgericht te zijn, openbaarheid te prediken en inspraak te bieden aan haar burgers. De lambda-waarden tot slot vallen terug op de aanpassingscapaciteiten van de overheid. Het gaat hierbij om de robuustheid van de publieke sector, die grote crisissen weet te overleven en zich weet aan te passen aan wisselende omstandigheden.

Volwassen democratieën komen in meer of mindere mate aan alle drie groepen kwaliteitsmaatstaven tegemoet. Er zijn mechanismen om openheid te garanderen (bv. wetgeving openbaarheid van bestuur) en om efficiëntie en effectiviteit te verhogen (bv. e-gov toepassingen in de back office). De meeste overheden willen ook klaar te staan in moeilijke tijden (bv. door rampenplannen te simuleren). De *New Public Management*hervormingen van de jaren ’90 benadrukten in de eerste plaats de sigma-waarden (efficiëntie en effectiviteit) en dit als antwoord op de economische moeilijkheden in de jaren ’80. De rangschikkingen van publieke sectoren focussen in navolging hiervan ook vooral op deze sigma-waarden. Men mag echter de andere 'zachtere' waardecategorieën niet uit het oog verliezen bij de evaluatie van een overheid.

De tweede dimensie betreft de technische dimensie. In dit onderdeel hebben we het statistisch-technische gedeelte van de overheidsmeting nagekeken. Vragen naar de transparantie, meetfouten en de analyse-eenheid komen hier aan bod. De vraag 'of er goed gemeten wordt' vormt de rode draad doorheen deze onderzoeksdimensie.

Aan deze twee dimensies, die in de bestaande internationale literatuur veelvuldig ter sprake komen, hebben we nog een derde dimensie toegevoegd. De politiek-administratieve dimensie moet een blik werpen op wie er meet, met welk doel voor ogen de producent dat doet en welke doelgroep ze daarbij op het oog heft.

4. Kwaliteitsproblemen

De explosieve toename van het aantal internationale rangschikkingen van publieke sectoren heeft geleid tot een grote nood aan kwaliteitstoetsing van de databronnen achter dergelijke comparatieve data. Aan de hand van een grondig literatuuronderzoek enerzijds en een diepgravende empirische analyse van een selectie van governance rankings anderzijds hebben we ons licht laten schijnen over deze kwaliteitsproblematiek.

De kwaliteitsproblemen die in dit hoofdstuk aan bod komen, hebben betrekking op de onderzoekstechnische kwaliteit van de governance rankings in brede zin. Het gaat daarbij onder andere om de kwaliteit van de indicatoren die aan de basis van de meting liggen, de structuur van de ranking en de kwaliteit van de vergaarde data. De statistische tak van de Europese Unie, Eurostat, heeft voor het beoordelen van de kwaliteit van data en statistieken een aantal aandachtspunten geïdentificeerd. In onze synthese van de kwaliteitsproblemen hebben we ons onder andere gebaseerd op dit theoretische kader (zie kader 5).

Kader 5: Kwaliteit van data en statistieken

- *Relevance*: Er wordt op voldoende mate rekening gehouden met de noden van de gebruikers.
- *Accuracy*: Hoe goed de data de realiteit weerspiegelen.
- *Timeliness and punctuality*: De tijd die er tussen de verspreiding en de dataverzameling zit.
- *Accessibility and Clarity*: De manier waarop de data beschikbaar worden gesteld en de kwaliteit van de verklaringen.
- *Comparability*: Het gevolg van het gebruik van uniforme concepten en definities, meetmethoden en procedures.
- *Coherence*: De data zijn consistent.
- *Completeness*: De mate waarin de gevraagde data beschikbaar zijn.

(Working Group 'Assessment of quality in statistics' - Eurostat, 2003)

In dit hoofdstuk onderscheiden we uiteindelijk negen grote punten van kritiek. Ten eerste klagen we het foutieve gebruik van subjectieve indicatoren aan. In tweede instantie bespreken we het gebrek aan

representativiteit bij surveys. Ten derde signaleren we de conceptuele tekortkomingen in de meting. Ten vierde hebben we het over het voorkomen en in acht nemen van meetfouten in de governance rankings. Vervolgens kaarten we in vijfde instantie de onderbelichting van de politiek-administratieve dimensie van de rankings aan. De zesde kritiek betreft de bias die in vele internationale rangschikkingen van publieke sectoren schuilt. Als zevende hebben we het over de *composite indicators*, die als nadeel hebben onderliggende informatie van de basisindicatoren te verliezen. De achtste kritiek op de kwaliteit handelt over het gebrek aan transparantie. De negende en laatste kritiek tot slot stelt de vergelijkbaarheid in vraag van publieke sectoren over nationale grenzen heen.

Tot slot maken we in paragraaf 10 een slotbeschouwing van de kwaliteitsproblematiek. We geven aan welke conclusies aan het vastgestelde kwaliteitsdeficit moeten vastgeknoopt worden. Als auteurs verzetten we ons immers tegen de stroming die beweert dat de hoofdconclusie luidt dat er niet gemeten kan worden als het gaat over de publieke sector in verschillende landen. Wel is het noodzakelijk de tekortkomingen bloot te leggen opdat de conclusies met de nodige nuance komen. Deze visie vormt de rode draad doorheen ons onderzoek.

> 4.1. Foutief gebruik van subjectieve indicatoren: objectieve feiten versus subjectieve meningen?

Governance rankings stoelen op een samenvoeging van één of meerdere basisindicatoren. De indicatoren die aan de basis liggen van de meting van een concept zoals governance -of in ons geval specifieker *government effectiveness*- kunnen van objectieve of subjectieve aard zijn.

Objectieve indicatoren zijn gebaseerd op een concrete meting, een registratie van objectieve feiten. Het begrotingsoverschot, de nationale schuld of het aantal geregistreerde misdaden in een land zijn hier voorbeelden van. Deze cijfers komen vaak uit officiële statistieken en krijgen daardoor -terecht of onterecht- een zekere betrouwbaarheidsstempel met zich mee.

Subjectieve indicatoren daarentegen zijn gebaseerd op percepties. In het geval van de overheidskwaliteit gaat het om de survey-antwoorden van burgers, experten of andere respondenten over hun ervaringen met, percepties over, of evaluaties van het overheidsapparaat.

De grote fout die vaak gemaakt wordt door zowel producenten als gebruikers van internationale rangschikkingen van publieke sectoren is dat objectieve cijfers vaak verward worden met percepties, inschattingen en opinies. Objectieve indicatoren zijn *as such* niet beter geschikt dat subjectieve indicatoren, maar meten niet hetzelfde fenomeen als *'perception-based indicators*' (Parks, 1984).

Zo stelden Rothstein en Teorell (2005) in hun onderzoek in vele gevallen geen een-op-een relatie vast tussen de kwaliteit van overheidsdiensten gemeten op basis van de percepties van de bevolking, de economische wereld of experten, en objectieve prestatie-indicatoren. Ook Parks (1984) komt tot die conclusie en beschrijft hoe vele auteurs omwille van deze vaststelling openlijk de relevantie van subjectieve indicatoren in vraag stellen.

Men komt vaak bij subjectieve indicatoren uit wanneer men een moeilijk meetbaar concept wil meten. Governance rankings -en dan vooral in het onderdeel dat de kwaliteit van de overheid tracht in te schatten- maken in hun metingen in grote mate gebruik van subjectieve indicatoren. Meten is in

dit geval een minder goede terminologie, aangezien het eerder over een peiling of een inschatting van een opinie van de overheidskwaliteit gaat. Daar er geen harde cijfers beschikbaar of mogelijk zijn voor sommige indicatoren, gaat men over tot een bevraging van 'getuigen' in een survey. Zo beschikken ontwikkelingslanden vaak niet over '*facts-based indicators*' of missen de data elke geloofwaardigheid (Arndt & Oman, 2006; Besançon, 2003). In dergelijke gevallen moet de producent een keuze maken tussen zogenaamde 'objectieve indicatoren' waarbij deze landen in vele gevallen uit de boot van de ranking vallen of subjectieve indicatoren met een grotere *country coverage* tot gevolg.

Harde data over de kwaliteit van de publieke sector zijn sowieso al moeilijk te vinden. Objectieve prestatie-indicatoren bestaan wel, maar beperken zich meestal tot één land en/of binnen één beleidsdomein. Zo werden er onder meer vergelijkende studies van brandweerdiensten, diensten burgerlijke stand, ziekenhuizen, scholen, gevangenissen en rechtbanken opgezet (Bouckaert, 1992; Bouckaert, 1993; Blank, 2000).

Zowel het World Bank Institute als de European Central Bank stellen de prestaties van de overheid te meten, maar hun metingen zijn in feite voornamelijk gebaseerd op bevragingen van burgers of bedrijfsleiders of op de opinies van experten. De Worldwide Governance Indicator 'Government Effectiveness' bestaat hoofdzakelijk uit gegevens verzameld aan de hand van enquêtes, die van de European Central Bank voor een belangrijk deel ervan. Er moet echter voor ogen worden gehouden dat beide indicatoren in eerste instantie de perceptie van prestaties meten en niet de feitelijke prestaties van de verschillende overheden (Van de Walle et al., 2004). Toch worden de resultaten uit dergelijke onderzoeken verkocht als feiten, wat te kort door de bocht gaat.

Subjectieve indicatoren moeten omwille van diverse redenen met extra omzichtigheid benaderd worden. Ten eerste staan zogenaamde '*performance indicators*', gemeten aan de hand van percepties, onder betrekkelijke invloed van achtergrondkenmerken, waarden en normensystemen. Sociale en culturele factoren spelen daarbij een belangrijke rol (Van de Walle et al., 2004). Dit impliceert dat verschillende landen niet van op gelijke hoogte vertrekken, aangezien deze achtergrondkenmerken verschillen met het land en de cultuur.

Bijgevolg kan een negatieve houding van de bevolking ten aanzien van de overheid tot een negatieve perceptie van het presteren van de overheid leiden. Een dergelijke negatieve attitude kan het gevolg zijn van een algemene culturele context; los van het objectieve prestatieniveau. In dit geval scoort een overheid veeleer slecht omwille van een negatief imago dan dat het ook werkelijk slechte prestaties aflevert (Van de Walle et al., 2004).

Dit laatste komt volgens Laporta, Lopez-de-Silanos, Shleifer en Vishny (1999) bijvoorbeeld tot uiting in de vele *expert assessments* die aan de basis liggen van governance rankings. In een dergelijk systeem wordt de kwaliteit van de verschillende overheden ingeschat door experten, die een score toekennen op basis van de formele en informele informatie die hen ter beschikking staat. "*Subjective indicators tend to move together. Hence, expert opinion on a broad range of public sector related issues may simply reflect some general underlying sentiment toward a country*", aldus deze auteurs.

De invloed die de (slechte) reputatie van een overheid heeft op de subjectieve beoordeling van de overheidskwaliteit wordt ook wel het *Halo-effect* genoemd. De negatieve perceptie van een instituut als de overheid is daarbij mogelijks te wijten aan een slecht prestatieniveau op een bepaald (minimaal) aspect dat onterecht gegeneraliseerd wordt. Sommige critici beroepen zich op dit effect in hun kritiek op *expert opinions*. In hun ogen gaan experten er te snel van uit dat rijke en welvarende landen ook over een goede overheidskwaliteit beschikken (Dill & Soo, 2005).

Ten tweede speelt het verwachtingspatroon een belangrijke rol in de perceptie van het presteren van de overheid. Tevredenheid en perceptie worden mee bepaald door de verwachtingen. In landen met een traditioneel sterke dienstverlening zullen burgers een hoog verwachtingspatroon hebben. Indien dit verwachtingspatroon niet wordt ingelost, zal de subjectieve prestatie-indicator in het rood duiken, terwijl het objectieve prestatieniveau allicht hoger zal zijn dan in de landen die de omgekeerde beweging maken (Van Roosbroek & Van Dooren, 2006)

Ten derde is er het transparantieprobleem dat inherent is aan het gebruik van subjectieve indicatoren (Kuha & Gaskell, 2005). Subjectieve

indicatoren, en daarbij in eerste instantie *expert opinions*, zijn vaak moeilijk reproduceerbaar en missen een duidelijke en heldere onderbouwing. Dit zorgt ervoor dat de impact van de overheid in grote mate wordt ingeperkt. Voor overheden is het immers moeilijk in te schatten hoe men de perceptie van de bevolking, de economische wereld of experten kan beïnvloeden. Zeker als er geen zekerheid is of dit met objectief betere prestaties wel mogelijk is. Zo wordt het voor een overheid soms vechten tegen de bierkaai om de eigen rangschikking op het vlak van de overheidsperformantie te verbeteren.

Ten vierde speelt bij het gebruik van subjectieve indicatoren ook de vraag op wie er bevraagd moet worden. Dit aspect zorgt immers vaak voor kritische bedenkingen. Zo mag de opinie van de economische wereld van managers en bedrijfsleiders niet gelijkgesteld worden met de publieke opinie. Dit en andere bedenkingen met betrekking tot de representativiteit van surveys bespreken we in de volgende paragraaf.

Een meting enkel en alleen gebaseerd op objectieve indicatoren is echter vaak ook niet in staat om de globale realiteit te vatten. Harde data vatten immers vaak enkel de *de jure* realiteit (de formele regels) en zien daarbij de even belangrijke informele en ongeschreven regels over het hoofd (Arndt & Oman, 2006). Sommige auteurs stellen zelfs dat percepties van de overheidskwaliteit vaak accurater de reële situatie vatten dan objectieve prestatie-indicatoren (Kauffman et al., 2000).

Ondanks de term 'objectieve indicator' dient men toch op zijn hoede te zijn voor valkuilen zoals meetfouten of een verschillend gebruik van definities bij vergelijkingen (Van Roosbroek & Van Dooren, 2006). Bovendien verliest men vaak uit het oog dat er ook bij de selectie van zogenaamde objectieve indicatoren subjectieve en normatieve keuzes aan te pas komen. Zowel de keuze van de vast te stellen feiten, als de interpretatie van hoe variatie in die feiten de kwaliteit van de overheid beïnvloedt, vergt immers subjectieve beoordeling (Arndt & Oman, 2006).

Het ideale meetscenario geschiedt volgens veel auteurs dan ook door een combinatie van subjectieve en objectieve indicatoren op te nemen in de meting van de overheidskwaliteit (Van de Walle et al., 2004; Besançon, 2003; Parks, 1984; Ministère des Affaires étrangères et Européennes

Française, ; Callens, 2007). Objectieve indicatoren zijn niet slechter of beter dan subjectieve indicatoren, ze vullen elkaar net perfect aan.

> 4.2. Non-representativiteit van respondenten in surveys

Zoals hoger aangegeven, maken de meeste internationale rangschikkingen van publieke sectoren gebruik van subjectieve indicatoren bij het meten van een concept als governance of de kwaliteit van het openbaar bestuur. Dit heeft twee voorname redenen. Ten eerste is het niet gemakkelijk om harde data te vinden over de kwaliteit van de overheid. De economie, de gezondheidszorg of het onderwijs kunnen gemakkelijker met objectieve prestatie-indicatoren gemeten worden, een goed werkend overheidsapparaat of een kwaliteitsvolle overheid zijn veel moeilijker meetbare concepten.

Ten tweede is het ook belangrijk voor de meeste governance rankings om over een brede *country coverage* te beschikken. Veel landen beschikken echter niet over vergelijkbare statistieken op het vlak van de overheidswerking, waardoor deze landen uit de boot dreigen te vallen bij een meting die enkel gebaseerd is op objectieve indicatoren. Het gaat hier bovendien vooral over ontwikkelingslanden. De *good governance*-theorieën stellen echter net dat juist de Derde Wereldlanden het meest gebaat zijn bij een meting van de publieke sector. De rankings leveren immers een interne en externe druk op om tot beter bestuur te komen.

Ook om andere redenen is de opname van de minder ontwikkelde landen in vele internationale rangschikkingen van publieke sectoren van cruciaal belang. Zo zijn er de governance rankings die gebruikt worden voor de allocatie van financiële ontwikkelingssteun en goedkope leningen. Dat hierin de scores van de arme landen niet mogen ontbreken, hoeft geen betoog. Daarnaast zijn veel internationale governance rankings ontwikkeld om investerings- en vestigingsrisico's in 'risicovolle landen' in te schatten. Als een bedrijf van plan is economische investeringen te doen in bijvoorbeeld Zimbabwe zal het in veel gevallen onder andere een beroep doen op *risk-rating-agencies* die de overheidskwaliteit in dergelijke landen als risicofactor onderzoeken. Ook hier is het duidelijk dat vooral de ontwikkelingslanden blootstaan aan dergelijke risico's.

De voorbeelden van governance rankings die hun hele meting of het onderdeel dat de bestuurskwaliteit beoordeelt, baseren op subjectieve indicatoren, zijn talrijk. Zo is bijvoorbeeld de component ‘Government effectiveness’ uit de bekendste governance ranking, de Worldwide

Governance Indicators van het World Bank Institute, exclusief gebaseerd op subjectieve indicatoren. Ondanks het feit dat de producent van deze ranking uit maar liefst 18 verschillende databronnen put voor haar meting, is geen enkele van de 49 gebruikte variabelen van objectieve aard. Ook de *'commercial business information providers*' zoals de Economist Intelligence Unit en de Political Risk Services Group opteren volledig voor subjectieve indicatoren in hun *expert opinion surveys*. Het International Institute for Management Development kiest in haar World Competitiveness Yearbook dan weer voor een mix van objectieve en subjectieve indicatoren, al hebben die laatste wel het overwicht.

Bij de keuze voor subjectieve indicatoren, moet men de belangrijke vraag stellen wie men zal bevragen. Dit gebeurt impliciet in de selectie van de respondenten. De vraag wiens percepties of opinies over de overheidswerking men in beschouwing neemt, is immers van groot belang voor de uitkomst van de meting. *'Where you stand depends on where you sit'*. Verschillende bevolkingsgroepen hebben vaak verschillende agenda's, verschillende belangen en dus ook verschillende opinies of percepties.

Er zijn in grote lijnen drie verschillende types van surveys die gebruikt worden in de governance rankings. Ten eerste zijn er de bevolkingssurveys. Hierbij wordt de perceptie van de man in de straat gevraagd en wordt er doorgaans geprobeerd een zo representatief mogelijk staal van de bevolking te bevragen. Aangezien men op deze manier probeert de publieke opinie weer te geven, minimaliseert men zo het risico op selectieve bias.

Ten tweede zijn er de bedrijfssurveys. In dergelijke bevragingen worden enkel bedrijfsleiders, managers en andere leden van het topkader binnen de economische wereld bevraagd over hun perceptie, opinie en ervaring.

Governance rankings hebben steeds vaker een economische inslag en maken dan ook vaak gebruik van dit soort surveys. Ze worden bovendien in eerste instantie geconsulteerd door economische instellingen, ondernemingen of bedrijfslui. In deze governance rankings worden de subjectieve indicatoren dan ook meestal ingevuld met surveys onder bedrijfslui. Aangezien deze groep geen afspiegeling van de bevolking kan genoemd worden, is er dus weinig sprake van representativiteit voor de bevolking van een land. Wat goede kwaliteit voor een inwoner van een land is, komt niet altijd overeen met een goede overheidswerking voor opererende bedrijven. Bedrijfslui en

managers zijn immers in algemene termen meer gebaat bij lage taksen en een minimale regulering, terwijl het publiek juist een eerlijk belastingsregime en gepaste regulering verwacht (Kauffman et al., 2007b).

Niet alle auteurs zijn het echter eens met deze laatste stelling. Kuhry (2004) stelt dat er een sterke samenhang bestaat tussen de subjectief ingeschatte kwaliteit van de overheid door de bedrijfswereld en het vertrouwen dat door de bevolking in de overheid wordt gesteld.

Men moet bovendien voor ogen houden dat de ook bedrijfswereld geen homogene categorie vormt. Daardoor zijn veel *firm surveys* geen goede afspiegeling van de economische wereld. De bedrijven die worden aangeschreven voor dergelijke surveys weerspiegelen vaak niet de economische realiteit in het land.

Een veel gehoorde kritiek is dat multinationals overrepresentatief zijn in de samples. Onder andere het World Competitiveness Yearbook zou vooral vertegenwoordigers van *multinationals* in haar survey opnemen (Van de Walle, 2006a). Aangezien het grootste deel van de respondenten uit de alumnilijst van de bijhorende *Business School* getrokken wordt, is die kritiek allicht wel gegrond. De producent zelf countert deze kritiek door te stellen dat dit onbewust gebeurt en dat er enkel gezocht wordt naar bedrijfsleiders die internationaal voldoende ervaring hebben om de governance vragen op een correcte manier te kaderen. Dat deze visie leidt tot een overrepresentatie van *multinationals* ten nadele van bijvoorbeeld KMO's is echter duidelijk.

De derde surveycategorie betreft de *expert opinion surveys*. In dit soort surveys worden er expertenpanels samengesteld die hun expertise combineren met aangereikte informatie om landen te quoteren op kenmerken van overheidskwaliteit. De experten kunnen ofwel *in-country* analisten ofwel *in-house* analisten betreffen. In het eerste geval bevinden de experten zich in het land dat ze moeten beoordelen. In het tweede geval gaat het om stafleden van de producerende organisatie die van 'uit de eigen (hoofd)zetel' -vanuit het buitenland- de indicatoren beoordelen.

Inherent aan dergelijke expertensurveys is dat ze weinig transparant zijn. De overwegingen van de experten zijn moeilijk na te gaan, laat staan dat de data reproduceerbaar zijn. Bovendien is er een groter risico op bias bij

experten, omwille van de aard van deze respondent en het doorgaans geringe aantal dat in het scoringsproces wordt ingezet.

Kauffman, Kraay en Zoido-Lobotàn, de auteurs van de Worldwide Governance Indicators, die in hun governance meting in grote mate gebruik maken van expertensurveys zien er geen kwaad in. Volgens hen bestaat er immers een grote eensgezindheid tussen de beoordelingen van analisten van *risk-rating-agencies* en de opiniepeilingen onder bedrijven of de bevolking. "*International analysts bring a global view while businesses and citizens bring local knowledge, and generally their views coincide* (Kauffman et al., 2000)."

Sommige governance rankings die zachte data gebruiken voor hun overheidsmeting, gaan resoluut voor één bepaald soort survey. Zo baseert de Economist Intelligence Unit zich in haar Country Risk Briefing -waarbij politieke risico's worden geanalyseerd, waaronder de 'government effectiveness'- uitsluitend op expertenscores. Dit zorgt voor een groot transparantieprobleem. Het is immers erg moeilijk om na te gaan hoe deze scores tot stand kwamen, waardoor ze zich moeizaam lenen tot leereffecten bij gerangschikte overheden.

Andere producenten trachten een gebalanceerde mix van de diverse surveytypes samen te brengen in hun meting. Het World Bank Institute onderstreept graag dat het zowel *'commercial business information providers*', als *'surveys of firms'* als *'surveys of households*' gebruikt in de samenstelling van haar governance ranking. Desalniettemin bekritiseren verschillende auteurs de producent van de Worldwide Governance Indicators voor *'misleading advertising*'. Het wekt immers verkeerdelijk de indruk dat bevolkingssurveys het grootste gewicht meekrijgen. Arndt gaat nog een stap verder en stelt dat het World Bank Institute dit bewust doet om de indicatoren extra aantrekkelijk te maken voor donororganisaties die op zoek zijn naar data die de levensstandaard van de armen willen verbeteren (Arndt, 2007).

De representativiteit van surveys voor de invulling van subjectieve indicatoren hangt niet enkel samen met de aard van de respondenten, ook het aantal respondenten speelt hierin een belangrijke rol. Zo baseert het International Institute for Management Development (IMD) zich in haar gerenommeerde World Competitiveness Yearbook wat betreft de zachte

data op een zelf afgenomen *Executive Opinion Survey* die in totaal 3.700 bedrijfsleiders en managers ondervraagt. Als men naar de uitsplitsing per land kijkt, oogt dit aantal al een stuk minder overtuigend. Zo komt men uit bij de vaststelling dat de Belgische scores voor 2007 gebaseerd zijn op een bevraging van niet meer dan 44 respondenten uit de economische wereld. Dit is toch wel een erg kleine sample om gewichtige conclusies over de overheidskwaliteit aan vast te knopen.

> 4.3. Conceptuele tekortkomingen in de meting

Ook op het vlak van de theoretische achtergrond en conceptuele validiteit, schieten zo goed als alle governance rankings tekort. De internationale rangschikkingen van publieke sectoren baseren zich doorgaans op indices, die de '*composite indicator'* zijn van een onderliggend raamwerk van indicatoren. De selectie van deze indicatoren en de keuze voor een aggregatiemethode is inherent normatief. De keuze voor bepaalde indicatoren is immers representatief voor een specifieke inhoudelijke invulling en tegelijkertijd een keuze 'tegen' andere variabelen.

Ook de weging die men aan elke component, subindicator of criteria toekent, is van normatieve aard. Een hoger gewicht betekent immers dat er meer belang aan de indicator gehecht wordt. De beweegredenen achter dergelijke methodologische overwegingen zijn dus duidelijk van politieke aard.

Dit zorgt ervoor dat zelfs voorstanders van het gebruik van '*composite indicators*' -ook wel geaggregeerde indicatoren, synthetische indicatoren, samengestelde indicatoren of indices genoemd- niet om het feit heen kunnen dat ze een mantel van onbetrouwbaarheid om zich heen dragen. Wegens de vaak onduidelijke en meestal oncontroleerbare onderliggende structuur van een geëvalueerd composiet fenomeen, is onzekerheid inherent aan de constructie van '*composite indicators'*. Het gebrek aan een standaard constructiemethode, de onenigheid onder experten betreffende het belang van de onderliggende indicatoren, en andere onopgeloste methodologische aspecten zijn illustratief voor deze onzekerheid (Saltelli, 2005).

In de lijn van deze bemerkingen ligt de frequent geopperde kritiek dat '*composite indicators*' kunnen misbruikt worden om een gewenst beleid te ondersteunen, zeker als het constructieproces niet transparant is en duidelijke statistische of conceptuele principes mist. Teneinde de onzekerheid ten aanzien van de metingen van publieke sectoren te milderen, zou er dus veel belang gehecht moeten worden door de producenten aan conceptuele precisering, theoretische onderbouw en normatieve verantwoording.

Maar daar knelt nu net het schoentje. De keuze voor een selectie indicatoren mist vaak elke onderbouwing. Ook de berekeningsmethode, waar de wegingsfactoren een belangrijke rol in spelen, komt steevast uit de spreekwoordelijke lucht gevallen. De gebruiker heeft er doorgaans het raden naar waarom de index op een bepaalde manier is geconstrueerd. Nochtans zouden de keuzes die moeten gemaakt worden, onderbouwd moeten worden met grondige analyses, omdat ze een grote invloed hebben op de betekenis van de uiteindelijke hoofdindicator.

Bovendien missen de geaggregeerde indicatoren vaak elke duiding. Of de producent nu landen quoteert op hun '*competitiveness*', '*governance*' of *'government effectiveness*', het is niet duidelijk wat nu eigenlijk gemeten wordt.

Als er al in een definitie of omschrijving wordt voorzien, verdrinkt die steevast in algemeenheid, onduidelijkheid en vaagheid. Dit is mogelijks een gevolg van het feit dat de structuur van de meeste rankings door de jaren heen significante wijzigingen heeft doorgemaakt.

De definitie die de auteurs van de Worldwide Governance Indicators voor governance voorzagen, is een passende illustratie van de vaagheid die voor dit soort definities wordt bewaard. Governance wordt daar gedefinieerd als *"the traditions and institutions that determine how authority is exercised in a particular country, the process by which governments are selected, held accountable, and replaced, the capacity of governments to manage resources efficiently and formulate, implement, and enforce sound policies and regulations, and the respect of citizens and the state for institutions that govern economic and social interactions among them"* (Kauffman et al., 2000). Rothstein en Teorell stellen terecht dat "*this definition is just about as broad as any definition of politics. More over it fails to distinguish between the content of specific policy programs on the one hand and governing procedures on the other*" (Rothstein & Teorell, 2005).

In het verlengde van deze discussie, is het bijzonder problematisch dat bijvoorbeeld in de Worldwide Governance Indicators 212 landen vergeleken worden op 'Government effectiveness', terwijl deze indicator voor verschillende landen op verschillende manieren wordt ingevuld. Het World Bank Institute stelt immers de scores van de verschillende landen vast op basis van de voor dat land beschikbare scores op 49 indicatoren uit 18

verschillende databronnen. Het feit dat geen enkel land in elk van die geselecteerde databronnen voorkomt, geeft aan dat veel landen vergeleken worden op basis van ongelijke indicatoren.

Als een internationale rangschikking van publieke sectoren de prestaties van de publieke sector wil meten en vergelijken, moet het ook aangeven welke opvatting aan '*performance*' wordt gegeven. Neemt men voor de prestaties van de overheid als uitgangspunt input, output, outcome, efficiency of effectiveness (Van de Walle, 2005)? Ook dit wordt echter niet systematisch nagekomen. In onze empirische analyse zal duidelijk geïllustreerd worden dat hier geen theoretische keuze in wordt gemaakt, wat een mix van de verschillende soorten prestatie-indicatoren in de meeste rankings oplevert.

Thomas (2007) bekritiseert de Worldwide Governance Indicators voor diens conceptuele zwakte. Ze verwijt het World Bank Institute een gebrek aan '*Construct validity*'. Voor de gebruiker van de WGI, lijkt de conceptuele en methodologische constructie immers nattevingerwerk. Dit is volgens de auteur één van de hoofdredenen waarom deze nochtans wijdverspreide en geprezen governance indicators niet bruikbaar zijn voor beleidsvorming.

Construct validity behoeft een cruciaal kenmerk. Er dient een duidelijke definitie te zijn voor het gemeten concept. Die definitie blijft zo goed als altijd achterwege. Bij de Worldwide Governance Indicators bestaat er wel een definitie van governance, maar die is zo vaag dat ze meer vragen opwerpt dan antwoorden te geven. De deelindicatoren, waar 'Government effectiveness' er één van is, ontbreken zelfs elke verdere duiding.

De auteurs counteren deze kritiek door te stellen dat er geen wetenschappelijke eensgezindheid bestaat over wat governance juist is. Zoals Kauffman, Kraay en Mastruzzi (Kauffman et al., 2007b) het verwoorden: "*Endlessly waiting for the articulation of a complete, coherent and consistent theory of governance before proceeding to measurement and action, while perhaps intellectually satisfying to a few, would be impractical to the point of irresponsibility*."

In onze ogen is dit een terechte repliek aan het adres van de criticasters die het gebrek aan conceptuele eensgezindheid vertalen in een eindeloos embargo op het verder meten en rangschikken van publieke sectoren. Maar het overtuigt ons niet wat betreft de instandhouding van het conceptuele

deficit. De auteurs moeten ook zonder internationale eensgezindheid stelling durven innemen en een conceptuele omschrijving voorzien die op wetenschappelijke basis is onderbouwd. Op die manier kan zowel de gebruiker als de onderzoeker begrijpen wat er gemeten wordt.

Op dit moment steken de producenten zich echter nog al te gemakkelijk weg achter het feit dat er geen één unaniem aanvaarde governance-theorie bestaat om conceptuele onduidelijkheid in hun eigen metingen te verantwoorden. Ook de keuze voor deze of gene onderliggende indicator moet verantwoord worden. Een governance-theorie zou de basis moeten vormen van elke indicatorenconstructie die onder een te meten geaggregeerde indicator zoals overheidskwaliteit of *'government effectiveness*' ligt. Idealiter gaat het hier om een theorie die eensgezindheid geniet.

De bestaande governance rankings werken echter vanuit een omgekeerde redenering. Aangezien een afgelijnde en duidelijke definitie van de gemeten indicator in de regel ontbreekt, is de betekenis ervan eigenlijk enkel af te leiden uit de onderliggende indicatoren. Op die manier is de omschrijving, het bereik van 'Government effectiveness' (zoals gemeten door het World Bank Institute) een rechtstreeks gevolg van de geselecteerde onderliggende indicatoren waarin het concept wordt geoperationaliseerd.

De kwaliteit van de overheid wordt in de gecoverde landen gemeten, maar niemand die kan aangeven wat die kwaliteit inhoudt. We verglijden op deze manier tot een situatie waarin "*the act of choosing a set of indicators and weights imposes a definition of 'quality*' (Usher & Savino, 2006)". Deze indicatoren worden bovendien meestal nog eens geselecteerd op basis van beschikbaarheid. De indicatoren die voorhanden zijn en in het plaatje van de meting passen, komen in aanmerking.

Een wetenschappelijk correcte conceptualisering zou daarentegen moeten voorzien in een *a priori* definitie van een te meten indicator (bijv. 'kwaliteit van het openbaar bestuur'), van waaruit dan een operationalisering in termen van basisvariabelen kan volgen. Een van de aanbevelingen voor de producenten luidt dan ook werk te maken van de conceptuele dimensie van hun governance rankings. Een governance-theorie die bepaalde accenten legt en vanuit een wetenschappelijke onderbouwing de conceptuele en

methodologische keuzes verantwoordt, draagt bij tot een betere en correctere interpretatie van wat gemeten wordt.

Niet alleen de inhoudelijke invulling en samenstelling verschillen tussen de verschillende governance rankings, ook het label varieert van de ene rangschikking tot de andere. In onze analyse doen we een onderzoek naar de rankings die de kwaliteit van het openbaar bestuur in verschillende landen meten en vergelijken. In sommige rangschikkingen spreekt men van 'Government effectiveness' (Cf. Worldwide Governance Indicators van het World Bank Institute (WBI)), in andere van 'Government efficiency' (Cf. World Competitiveness Yearbook van het International Institute for Management Development (IMD)). Maar ook 'Institutional (in)effectiveness' (Cf. Country Risk Briefing van de Economist Intelligence Unit (EIU)) zou eenzelfde lading moeten dekken. Als men er de waaier aan internationale rangschikkingen op naslaat, kan men enkel tot de conclusie komen dat de producenten niet overeenstemmen over het te meten concept, zowel qua benaming als qua invulling.

Een ander probleem dat conceptuele klaarheid verhindert, is het grote aantal indicatoren dat soms gebruikt wordt voor de samenstelling van een governance-meting. Opnieuw komen we hiervoor ter illustratie uit bij de Worldwide Governance Indicators. "De dimensie over 'Government effectiveness', die wij van naderbij bekijken in onze analyse, is een gebalanceerde indicator, maar het is uiteindelijk te moeilijk om in te schatten wat er eigenlijk gemeten wordt", stelt Van de Walle (2006a). De score komt tot stand na een aggregatie van 49 verschillende individuele indicatoren, die allemaal nog eens op een andere wijze tot stand kwamen. Er liggen immers 18 verschillende databronnen aan de oorsprong van deze individuele indicatoren en die hanteren niet allemaal dezelfde datagaringsmethode. Zo zitten er *expert opinion polls* tussen, bedrijfssurveys en ook bevolkingsbevragingen. Het wordt uiteindelijk bijzonder onoverzichtelijk wat de opeenhoping van dit grote aantal indicatoren nog van betekenis heeft.

> 4.4. Meetfouten: Vergeven en vergeten?

De rangschikkingen die landen op basis van de overheidskwaliteit vergelijken en zo een rangnummer aan elk land toekennen, ogen klaar en duidelijk, maar niets is minder waar. Door meetfouten expliciet mee in rekening te brengen –iets wat enkel Transparency International en het World Bank Institute sinds kort aandurven- vervalt een heel deel van de vergelijking in het niets. Doordat er relatief grote betrouwbaarheidsintervallen in het spel zijn, is het vaak moeilijk een uitspraak te doen over vooruitgang of achteruitgang in de tijd en zijn ook veel vergelijkingen tussen landen statistisch niets zeggend.

Veel producenten reppen echter met geen woord over *'measurement errors*'. Meetfouten zijn nochtans onvermijdbaar bij het meten van governance(-aspecten) en daarom niet onoverkomelijk (Kauffman & Kraay, 2007b). "Alle metingen van governance zijn immers uiteindelijk imperfecte *proxies* voor de governance-dimensie die men feitelijk wil meten", laten de auteurs van de Worldwide Governance Indicators optekenen in hun rapport.

Producenten van internationale rangschikkingen van publieke sectoren laten het nochtans zo goed als altijd na tot de berekening van een betrouwbaarheidsinterval over te gaan. Die nuanceren de ranking en de vergelijking tussen landen immers in te grote mate. Het World Bank Institute gaat er prat op als één van de enige internationale organisaties in haar governance rankings expliciet te wijzen op de meetfouten en voorziet ook een betrouwbaarheidsinterval (*confidence interval*) bij haar scores.

"Meetfouten", zo stellen de auteurs van de Worldwide Governance indicators, "zijn onvermijdelijk". Ze zijn een gevolg van codeerfouten en datafouten (Hood, 2007). Codeerfouten treden op als er een verschil bestaat tussen gecodeerde scores. De verantwoordelijkheid ligt hier bij de codeerder zelf. Wanneer twee onafhankelijke waarnemingen voor eenzelfde indicator twee verschillende scores opleveren, spreekt men van een codeerfout. De datafouten zijn fouten die gerelateerd zijn aan de databron(nen) die aan de basis liggen van de meting. Zo geven Kauffman, Kraay en Zoido-Lobotàn (Kaufmann et al., 1999) voor de Worldwide Governance Indicators toe dat "*many of the polls and surveys we use suffer from deficiencies such as poorly-worded questions about ill-defined and excessively broad concepts*".

Dat deze betrouwbaarheidsintervallen eigenlijk onontbeerlijk zijn, wijst een berekening van de auteurs van de Worldwide Governance Indicators zelf uit. In een van hun *working papers* geven Kauffman, Kraay en Mastruzzi toe dat minder dan 60% van de vergelijkingen tussen landen significant (op significantieniveau van 0.90) zijn als de betrouwbaarheidsintervallen in rekening worden gebracht. Tussen 1998 en 2006 heeft niet meer dan 30% van de landen een significante verandering ervaren in één van de zes governance-componenten (Kauffman et al., 2007a). Niet bepaald de conclusie die je voor ogen hebt als je blindelings de vele dataschommelingen in de tijdreeksen in beschouwing zou nemen.

De aanwezigheid van betrouwbaarheidsintervallen verhindert de mogelijkheid tot vergelijkingen over tijd en tussen landen niet. Integendeel, het geeft net aan welke vergelijkingen betekenisvol zijn en welke niet. Het World Bank Institute wil er naar eigen zeggen een tegengewicht mee vormen tegen -wat ze noemt- *'superficial elevator discussions'*. In dit soort benaderingen tracht men de rankings te herleiden tot kortzichtige uitspraken als 'land A presteert beter dan land B'.

De auteurs van de Worldwide Governance Indicators geven in hun naslagwerken meer dan eens aan niet voor een vergelijking op basis van de governance data in de strikte zin te zijn. Daarvoor is de foutenmarge te groot en de nuance in de vergelijking te ver weg. "*The imprecision indicates that 'running horse races' between countries to determine their precise rankings on governance is not useful*. (Kauffman et al., 2000)"

Ook Transparency International geeft expliciet aan dat er meetfouten achter de data schuilgaan. Deze producent geeft reeds lange tijd betrouwbaarheidsintervallen mee. Dat maakt de broosheid van de data duidelijk, maar werkt ook extra vertrouwen in de hand wat betreft de significante verschillen en vergelijkingen. Dit doet Galtung (2005) het volgende opmerken: "*The CPI* (Corruption Perceptions Index; de governance ranking van Transparency International, ML) *is not particularly precise. The CPI has the advantage, however, of being transparent and honest about this failing. […] The CPI does not claim to be more precise than it actually is*."

Naast de validiteit die in de conceptuele dimensie tot uiting komt, is ook de betrouwbaarheid van de data een belangrijk onderdeel van de kwaliteit van de internationale rangschikkingen van publieke sectoren. Hier geldt de

vraag: 'Worden de goede dingen gemeten?' De data moeten van aanvaardbare kwaliteit zijn. Ook hier geldt immers: *'Garbage in, garbage out*'. Met statistische analyses kan men geen slechte data rechtzetten (Van Dooren et al., 2006). Voorbeelden van slechte data zijn bijvoorbeeld statistieken uit een te kleine steekproef of op basis van een scheve respondententrekking.

Ten gevolge van enerzijds de brede betrouwbaarheidsintervallen bij de internationale rankings die meetfouten in rekening brengen en anderzijds het vermoeden dat de governance rankings die dit niet doen op zijn minst even grote betrouwbaarheidsintervallen zouden hebben, opteren sommige critici ervoor om het idee van rangschikkingen achter zich te laten en over te stappen op grote categoriseringen. Op die manier wordt er minder gewicht verbonden met kleine verschillen tussen landen, en eerder grote groepen gedifferentieerd van gelijk presterende overheden. Ook enkele producenten van governance rankings zijn dit idee genegen.

Freedom House is een van de producenten die ervoor kiest haar resultaten niet in rangschikkingen te gieten, maar de landen te clusteren volgens vergelijkbare score. '*Not free*' zijn de landen met de slechtste score (met een score tussen 5.5/7 en 7/7), '*Partly free*' geldt voor de landen in het middenstuk (tussen 3/7 en 5/7) en '*Free*' is voor de landen in het koppeloton gereserveerd (met een score tussen 1/7 en 2.5/7).

> 4.5. Onderbelichting van politiek-administratieve dimensie

De wetenschap heeft zich tot nog toe op twee vlakken gefocust. Enerzijds werd de conceptuele dimensie vanuit diverse invalshoeken geanalyseerd. In deze dimensie gingen auteurs op zoek naar de vraag: 'Meet de producent het goede?'. De conceptuele validiteit van de governance rankings en indicatoren werd onder de loep genomen en kreeg zo goed als altijd een onvoldoende. Anderzijds heeft de literatuur zich ook met de technische dimensie van de governance rankings ingelaten. Binnen deze dimensie stellen de auteurs eerder de vraag: 'Meet de producent de dingen goed?'. Hierbij wordt voornamelijk de statistische kant van de rankings doorgelicht.

Dit onderzoek wil aan deze twee dimensies nog een derde toevoegen. In een pijler, die we de politiek-administratieve dimensie dopen, bekijken we de producent van de internationale rangschikking van publieke sectoren. We gaan op zoek naar de doelstellingen en de beweegredenen achter de ranking. Daarnaast proberen we ook de aard van de producent en de beoogde doelgroep in kaart te brengen. Deze informatie biedt een meerwaarde voor de interpretatie van de governance data en moet leiden tot juistere conclusies uit internationale rangschikkingen van publieke sectoren.

We vragen ons onder meer af wie de producent is. Gaat het om een NGO, een private commerciële instelling of een overheidsgerelateerde internationale organisatie? Afhankelijk van de aard van de producent moeten de data immers vanuit een ander perspectief geïnterpreteerd worden. Daarmee verbonden is de doelstelling achter de rangschikking. Wat wil de producent van de governance ranking bewijzen met de data, waar wil ze naartoe? Ook deze objectieven hebben hun invloed op de samenstelling van de meting en de invulling van de data.

Niet alleen het beginpunt van de ranking –de producent- maar ook het eindpunt –de gebruiker- is van belang voor een goed begrip van de data. Door een onderzoek in te stellen naar de doelgroep van de governance data, willen we een betere kijk krijgen op het perspectief van waaruit de data moeten bekeken worden. Als de data bedrijfslui als doelgroep hebben, zal de meting immers anders geconstrueerd zijn dan bijvoorbeeld cijfers ten behoeve van milieuactivisten. Ook het antwoord op de vraag of de ranking een commercieel product betreft, dan wel een vrij toegankelijk

bewustmakingsinstrument, heeft zijn implicaties voor de conclusies die uit de data moeten getrokken worden. Antwoorden op de 'Wie?'-, 'Waarom?'- en 'Voor Wie?'-vragen moeten leiden tot een beter inzicht in de data. Het vormt in onze ogen een belangrijk onderdeel van de kwaliteit van de onderzochte governance ranking.

> 4.6. Onderhuidse bias

Veel internationale rangschikkingen van publieke sectoren zijn gestoeld op data die met een of andere vorm van vertekening kampen. Een deel van dit fenomeen is te verklaren door het feit dat governance niet eenduidig gedefinieerd is en moeilijk te meten valt. Doordat producenten zeer verschillende invullingen geven aan de meting en vergelijking van prestaties van de overheid, sluipt er vaak een bias in de meting. De invulling is immers niet alleen verschillend, er wordt ook vaak een fundamenteel andere invalshoek gebruikt.

Zo vertoont een belangrijk deel van de governance rankings een economische bias. De meting van de overheid wordt in deze gevallen opgevat als een beoordeling van de kwaliteit van de overheid 'ten gunste van bedrijven'. Dit uit zich voornamelijk in een *selection bias* bij subjectieve indicatoren. De scheeftrekking manifesteert zich dan in de samenstelling van de onderzoeksgroep. Zo bekritiseren Arndt en Oman (2006) en Knack (2006) de meting van de Worldwide Governance Indicators voor het feit dat ze *biased* zouden zijn in de richting van de *'business elite'*.

Ook de indicatorenselectie onderliggend aan de meting van de kwaliteit van de overheid vertoont soms vertekeningen. Zoals eerder beschreven, is het selectieproces van dimensies, componenten en indicatoren die samen de structuur van de meting uitmaken, inherent normatief. "*Measurement is as much about politics as about techniques*", luidt een veelzeggende boutade (Van Dooren et al., 2007). Politieke keuzes spelen een rol in de indicatorensamenstelling, het verwondert dus niet dat er in veel gevallen vertekening in de meting sluipt.

Zo gaat er een non-interventionistische opvatting schuil achter veel governance rankings van *private rating agencies*. Dergelijke metingen en rangschikkingen hebben dan ook als voornaamste doel het adviseren van bedrijven bij investeringsbeslissingen. Het mag daarom niet verbazen dat '*government effectiveness'* vertaald wordt als de mate waarin de overheid (g)een obstakel vormt voor bedrijven. Een minimum aan regulering, een vlot werkende administratie en een consistente economische koers zijn vanuit deze optiek belangrijke kenmerken van een goed presterende overheid.

Internationale rangschikkingen van publieke sectoren zoals die van het World Economic Forum (Global Competitiveness Index) en het International Institute for Management Development (World Competitiveness Yearbook) zijn duidelijk niet in de eerste plaats ontworpen om onderzoek of vergelijkende *Public Administration* te voeden. Hun afnemers zijn in eerste instantie bedrijven en investeerders, waardoor de overheid vanuit economische hoek benaderd wordt. Vrijhandel en minimale overheidsinterventie staan daarbij voorop. Voor *Public Administration*-doelen is dit uiteraard een vertekend en zeer onvolledig beeld (Van de Walle, 2006a).

Dit zet enkele critici ertoe aan de *good governance* uit de internationale rangschikkingen van publieke sectoren te vertalen als '*good-for-economic-development*'. Niet-economische gevolgen van overheidskwaliteit worden immers vaak toch genegeerd (La Porta et al., 1999). Ook de governance rankings die vanuit de World Bank worden aangestuurd worden als te economisch beschouwd. Drie van de vijf clusters die in de erg belangrijke Country Policy and Institutional Assessments worden gebruikt, zijn van strikt economische aard. Ook de Worldwide Governance Indicators vatten *good governance* in belangrijke mate op als de beleidslijnen die een goede marktwerking aansturen (Schmitz, 2007).

Ook rankings die in grote mate op expertenpanels berusten, worden door critici er van verdacht *biased* te zijn. Zo houden dergelijke surveys een groter risico in op culturele vertekening. Zeker bij *in-house* analisten is dit het geval. Het gaat hierbij vaak over nationale gezellen, die in grote mate eenzelfde achtergrond en visie op governance met elkaar delen.

Verder bestaan er ook grote twijfels over de 'objectiviteit' van governance-metingen en rankings die uitgaan van *Think tanks*. Het gedachtegoed van de *expert opinion* van de *Think tank* is immers mogelijks beïnvloed door de ideologie van de geldschieter(s) achter de denkgroep (Knack & Manning, 2000).

Van de Walle (2005) stelt dat de meeste in governance rankings gebruikte indicatoren opgesteld zijn met het Westerse democratische systeem als ideaal in het achterhoofd. Ook Norman Girvan, de Secretaris-generaal van de Association of Caribbean States, bekritiseert de onderhuidse assumptie in

de governance rankings dat het Westerse meerpartijensysteem het ideaal is (Girvan, 2002).

Vanuit deze optiek moet dan ook het belang ingeschat worden dat we toedichten aan de politiek-administratieve dimensie van de governance rankings. De aard van de dataverzamelaar, de doelstellingen verbonden aan de producent en de doelgroep van de governance ranking, spelen een belangrijke rol in de kleur van de ranking van publieke sectoren. Van Dooren, Thijs en Sterck (2006) geven aan dat het belangrijk is in te zien dat internationale organisaties niet enkel een wetenschappelijke, maar ook een diplomatieke structuur en cultuur hebben. Metingen en procedures worden heus niet enkel bepaald door wetenschappelijke behoefte.

Een extreem voorbeeld hiervan is de Freedom House. Deze organisatie werd in 1941 opgericht en verricht vanaf 1972 jaarlijkse governance-metingen die landen vergelijken op het vlak van hun democratische gehalte. De statuten van deze organisatie liegen er niet om: "*All united in the view that American Leadership in international affairs is essential to the cause of human rights and freedom*." De democratiseringsdoelstelling van de Freedom House heeft dus mogelijks specifiek het Amerikaanse model als voorbeeld. Dit leidt op zijn minst tot een bias in de richting van de Angelsaksische overheidsmodellen.

De bias die we in al de governance rankings terugvinden, vormen op zichzelf echter niet het grootste probleem. Zoals de uitdrukking het stelt: "*The only way to keep statistics out of politics, is to collect irrelevant data*" (De Neufville, 1975). Metingen en vergelijkingen van publieke sectoren worden sowieso gevormd door politieke keuzes en normatieve overwegingen. Het problematische aspect bij internationale rangschikkingen van publieke sectoren is echter dat de bias vaak niet aan de oppervlakte wordt gebracht, maar eerder onderhuids, haast verscholen, blijft.

Dit komt onder andere door het feit dat een aantal governance rankings (cf. Worldwide Governance Indicators, Corruption Perceptions Index) hun indicatoren uit verschillende onderliggende databronnen putten. Deze onderliggende databronnen ontberen vaak een aanvaardbare vorm van transparantie, niettemin zij mee verantwoordelijk zijn voor de selectie en normatieve betekenis van de indicatoren.

Het zijn bovendien vaak dezelfde databronnen -voornamelijk de expertenanalyses en bedrijfssurveys- die terugkomen in verschillende rankings. De producenten van deze internationale rangschikkingen van publieke sectoren staven de betrouwbaarheid van hun meting met het feit dat er grote correlaties bestaan tussen hun rangschikking en enkele andere governance rankings. Maar ze baseren zich in hun meting dan ook vaak op dezelfde *biased* data, wat uiteraard leidt tot een hoge correlatie.

> 4.7. Over-aggregatie

Internationale rangschikkingen van publieke sectoren ervaren ook vaak de nadelen van het aggregeren van indicatoren. Het gaat vaak zo ver dat we kunnen spreken van over-aggregatie.

Al eerder in dit rapport expliceerden we de *'composite indicator*', die veelvuldig gebruikt worden in governance rankings. Dit soort indicatoren, ook wel synthetische indicatoren of geaggregeerde indicatoren genoemd, aggregeren (sub-)variabelen door middel van een vastgelegd wegingsschema. Het vormt het eindpunt van een aggregatieproces van onderliggende indicatoren die allen een bijdrage leveren aan de totale meting door een aspect van de hoofdindicator te kwantificeren.

De internationale rangschikkingen van publieke sectoren die het huidige landschap van governance-metingen domineren, kampen echter met een soort '*hitparade*-fenomeen'. Ze willen ten alle prijzen indicatoren in een simpele en eenvoudig te begrijpen vorm gieten. Niet de meting, maar de ranking wordt het hoofddoel. Het verglijdt tot een *'horse race'* waarbij uitsluitend wordt gekeken naar welk land anderen de loef afsteekt.

De *composite indicators*, waarbij de meting van een groep basisvariabelen naar buiten wordt gebracht onder de vorm van één score heeft enkele aantrekkelijke voordelen (Cherchye et al., 2007). Deze voordelen zorgen er voor dat dit soort indicator in toenemende mate gebruikt wordt bij governance rankings. Ten eerste zorgt de *composite indicator* voor een samenvatting van de complexe en veelzijdige onderliggende informatie die de basisvariabelen aanbrengen. Het bezit de eigenschap om de veelheid aan multi-dimensionale informatie die op de bodem ligt op een duidelijke en begrijpelijke manier samen te vatten.

Ten tweede vergemakkelijkt de geaggregeerde indicator het vergelijken tussen landen onderling. Het levert een eenduidig vergelijkingspunt op in de plaats van meerdere heterogene vergelijkingen. Bovendien zorgt het ervoor dat landen die data voor een bepaalde indicator missen, toch kunnen vergeleken worden met andere landen.

Ten derde, en in de huidige mediamaatschappij misschien nog de belangrijkste factor, is de *'composite indicator*' een gemakkelijk, *trendy* en

toegankelijk communicatie-instrument. Net door het aanzicht en de kenmerken van een *hitparade* halen dergelijke indicatoren vlot de headlines van de kranten en vinden ze ook gemakkelijk ingang in persberichten en andere *quick communication*. Net deze populariteit strijkt de wetenschap echter tegen de haren. De voorstanders argumenteren dan weer dat geaggregeerde indicatoren door hun communicatiekracht voor een verhoogde bewustmaking van belangrijke beleidskwesties kunnen zorgen.

De *composite indicator* heeft naast de netgenoemde voordelen echter ook een aantal belangrijke minpunten. Voor de aggregatie moet immers een prijs betaald worden. De belangrijkste kritiek luidt dat door het aggregeren van de basisvariabelen veel waardevolle informatie wordt kwijtgespeeld. De aggregatie heeft immers tot doel de aanwezige informatie te reduceren, maar op die manier zouden essentiële aspecten wel eens verloren kunnen gaan. Over-aggregatie zorgt dus voor een minderwaardig en mogelijks vertekend beeld van de meting.

Een tweede nadeel is dat bij *composite indicators* de data afhankelijk worden van het gebruikte wegingsschema. De uitkomst in de vorm van een governance ranking zal dus variëren met de gekozen aggregatiemethode. Dit zorgt voor veel controverse; geaggregeerde indicatoren krijgen het label onzeker, onbetrouwbaar en onstabiel. Dit beeld wordt nog versterkt door het feit dat er geen eensgezindheid bestaat onder experten over het te gebruiken wegingsschema.

Vaak wordt daardoor geopteerd voor een vaste wegingsmethode. Binnen deze aggregatiemethode kiest men vervolgens in de meeste gevallen voor *equal weights*; aan elke subindicator wordt een even groot gewicht gehangen. Dit lijkt gebalanceerd en objectief de meest accurate aggregatiemethode, maar dat is enkel schijn. Het impliceert immers even goed een normatief oordeel; alle subvariabelen zijn even relevant in de meting van de indicator. Kortom; de subjectiviteit en onenigheid die met de wegingsmethoden gerelateerd zijn, voeden de discussie over het gebruik van *composite indicators*.

Bovendien bestaat een noodzakelijke stap voor het aggregatieproces er in om alle scores te herschalen. Ook deze methodologische fase zorgt voor kritiek. De normalisering zorgt er immers voor dat de volgorde van de

rangschikking van publieke sectoren kan veranderen, zonder dat de scores variëren.

Ondanks de lijst van tegenargumenten, lijkt de verleiding om complexe metingen samen te vatten in één enkele score niet te weerstaan. Het onderzoeksveld zou zich daarom moeten toeleggen op het uitdokteren van het meest accurate wegingsschema. Cherchye, Moesen, Rogge en Puyenbrouck (2007) deden in dit kader reeds een poging. Zij stellen een *Data Envelopment Analysis* voor. Het vaste wegingsschema wordt daarbij verlaten voor een variabele weging. Kort samengevat geeft men voor een land een grotere weging aan subvariabelen met een hogere score. De impliciete gedachte die hier achter zit, is dat hogere scores zouden wijzen op een nadrukkelijker beleid op dat vlak. Die nadruk zou een vertaling zijn van de prioriteiten van een overheid in dat land. Om deze reden verwijzen de auteurs ernaar met de term *'benefit-of-the-doubt'*-methode. De voordelen van een dergelijke aggregatiemethode zijn divers en spelen rechtstreeks in op de vermeende minpunten van de modale aggregatiemethode. Ten eerste behoeven ze op technisch vlak geen *ex ante* normalisering van de scores. De methode is immers onafhankelijk van de meeteenheid. Dit zorgt voor een verminderd risico op vertekende eindrangschikkingen.

Ten tweede verhoogt zo de transparantie, aangezien duidelijk wordt op welke basis de gewichten worden vastgesteld. Het gaat immers over flexibele wegingsfactoren die de prioriteiten per land vertegenwoordigen. Zo wordt de informatiekloof ten opzichte van de gebruiker voor een groot deel gedicht.

Ten derde is het een gemakkelijk(er) te interpreteren model. Met de waas van onzekerheid, onduidelijkheid en gepercipieerde afhankelijkheid die rond *composite indicators* hangt, is dit een welgekomen steun in de rug van de geaggregeerde indicator.

Tot slot kunnen landen ten vierde zich bij slechte scores niet langer verschuilen achter het argument van de wegingsfactoren. Deze zijn op zo een manier tot stand gekomen, dat de eigen prioriteiten er in zouden moeten weerspiegelen. Ze zouden zo voor elk land de best mogelijke score naar buiten moeten brengen.

> 4.8. Gebrek aan transparantie

Een volgende kwalitatief probleem in de context van internationale rangschikkingen van publieke sectoren dat we expliciet willen benoemen, betreft het gebrek aan transparantie. Het toenemende gebruik van, de stijgende interesse in en het groeiende belang van governance rankings onderstreept nochtans de noodzaak aan transparantie in de internationale vergelijkingen van overheden. De realiteit wijst echter uit dat er op het vlak van transparantie nog een hele weg is af te leggen.

Transparantie staat in dit kader voor doorzichtigheid, beschikbaarheid en onderbouwing. Er zijn in het kader van de governance rankings meerdere aspecten van transparantie. Ten eerste is er de transparantie van de data. Hierbij doelen we dan voornamelijk op de beschikbaarheid van de rankings, van de datascores die aan de basis van de rangschikking liggen. Maar niet enkel de indexscores zijn van belang, ook de scores op de onderliggende dimensies en de basisvariabelen hebben hun relevantie.

Governance rankings scoren op dit basisniveau van transparantie redelijk tot goed. Meer en meer doen producenten de moeite om de data tot op het basisniveau ter beschikking te stellen. De nuance ligt er hem nog wel in dat deze beschikbaarheid lang niet altijd kosteloos is. Veel producenten gebruiken de rankings als *teaser*; ze worden via persmededelingen de wereld ingestuurd en prijken op de websites. Maar de cruciale onderliggende individuele scores worden vaak enkel tegen betaling aangeboden. De aanzienlijke prijzen die producenten -en dan voornamelijk, maar lang niet alleen de commerciële adviesbureaus- vragen voor de toegang tot de individuele scores, wegen op de transparantie van de meting.

Vervolgens is er het transparantieaspect van de gebruikte methodologie. Zoals reeds eerder besproken in de paragraf over de conceptuele tekortkomingen (cf. 4.3) is er een groot gebrek aan verantwoording van normatieve keuzes bij de samenstelling van de meting. De kwaliteit van de overheid wordt gemeten aan de hand van subindicatoren en basisvariabelen, maar de overweging die gemaakt wordt bij het uitselecteren van deze indicatoren wordt haast zelden onderbouwd. Op die manier wekken de

producenten argwaan. Het lijkt erop alsof de subindicatoren louter geselecteerd werden op basis van beschikbaarheid.

Aangezien een gebrek aan conceptuele precisering van de meting eerder de regel dan uitzondering is, hebben de indicatoren deze functie overgenomen. De invulling van de meting (van de kwaliteit van de overheid), de betekenis van de index, komt zo tot stand aan de hand van de selectie van de onderliggende indicatoren. Daardoor is het erg belangrijk te weten te komen op basis van welke norm(en) deze selectie werd uitgevoerd.

Hetzelfde kan gezegd worden van de governance rankings die ervoor kiezen externe databronnen te gebruiken in de berekening van een nieuwe meting van de overheidskwaliteit. De Worldwide Governance Indicators zijn van deze aard. Ze stellen een meting op die volledig berust op data die uit externe databronnen werden aangeleverd. Er wordt echter niet goed gecommuniceerd op welke basis deze databronnen werden geselecteerd. Ook dit duidt op een gebrek aan transparantie.

Ook het veelvuldig gebruik van subjectieve indicatoren in internationale rangschikkingen van publieke sectoren, helpt de transparantie niet vooruit. Ten eerste zijn de *expert opinions* moeilijk reproduceerbaar en dus ook weinig transparant. Producenten trachten de laatste jaren rekening te houden met deze kritiek en geven in sommige gevallen de subjectieve analyses mee op welke basis de exportenscores tot stand kwamen.

Daarnaast is er ook een heel deel van de subjectieve indicatoren die ingevuld wordt aan de hand van bedrijfssurveys en bevolkingspolls. Het grote transparantieprobleem van deze bevragingen ligt in het feit dat er bijzonder weinig informatie beschikbaar is met betrekking tot de respondenten. Het is vaak onduidelijk op basis van hoeveel respondenten de score tot stand is gekomen. Dit is nochtans een belangrijke basis om de significantie en representativiteit van de meting te controleren. Over de samenstelling van de respondenten wordt meestal al helemaal niets vrijgegeven. Gaat het vooral om mannen of vrouwen, stedelingen of plattelandsbewoners, rijke of arme personen?

De vrijgave van dergelijke informatie zou nochtans een beter zicht op mogelijke scheeftrekkingen kunnen bieden. Ook de beschikbaarheid van de *questionnaire*, die in veel gevallen bijna als een geheim document wordt

achtergehouden, zou een nuttige bijdrage kunnen leveren aan een goed begrip van de data en de ranking. Tot slot geldt voor een goede survey, dat ook gegevens met betrekking tot de non-respons worden vrijgegeven. Dit is niet zomaar randinformatie, het vertelt immers iets meer over de inhoud van de respons. Dergelijke informatie ontbreekt echter volledig bij governance rankings.

Ten slotte komen we ook hier weer uit bij de politiek-administratieve component van de governance ranker. Duidelijk betreffende de aard van de producent, diens doelstellingen en de voornaamste (beoogde) gebruikers bieden een meerwaarde voor de interpretatie van de data, daar ze biases en normatieve keuzes blootleggen. Het grote gebrek aan transparantie in de governance rankings in hun geheel duidt nogmaals op het belang van het in kaart brengen van de politiek-administratieve kant van de data en dataverzamelaar.

Wat transparantie en beschikbaarheid betreft, woedt er echter ook een discussie over het nut van het publiekelijk maken van benchmarkingsresultaten. Er zijn auteurs die eerder weigerachtig staan tegenover volledige openbaarheid. Ze stellen dat het oorspronkelijke doel van de governance rankings erin bestaat leereffecten te creëren bij de verschillende overheden. Deze zouden wel eens kunnen geschaad worden, bij de bekendmaking van de resultaten. Vertrouwelijke omgang met de uitkomsten van dergelijke metingen en vergelijkingen zou volgens deze visie de bereidheid tot leren en veranderen vergroten (De Groot, 2004).

Dit is echter niet de *mainstream* gedachtegang onder de producenten van governance rankings. De meeste, zo niet alle, governance rankings meten de prestaties van de overheid om die in de openbaarheid te brengen. Dit zorgt voor een *'blaming and shaming*' onder de gerankte landen (Arndt, 2007). Achter deze openbaarmaking gaat de idee schuil dat met publieke middelen geleverde prestaties hoe dan ook zichtbaar gemaakt moeten worden.

> 4.9. Vergelijkbaarheidsproblemen

De kritische vraag die zich opwerpt, is of er zomaar vergelijkingen kunnen en mogen gemaakt worden tussen verschillende landen, systemen en overheden. Sommige critici bestrijden immers de idee om landen op het vlak van de overheidswerking te vergelijken. Zij argumenteren dat overheden niet te vergelijken zijn wegens de verschillende administratieve stelsels, de verschillende politieke culturen en de uiteenliggende beleidskeuzes die de basis uitmaken in de verschillende landen.

De vergelijkbaarheidsfactor is voor Eurostat een belangrijk aspect van kwaliteitsvolle statistieken. De comparabiliteit is af te meten aan het gebruik van uniforme concepten en definities, meetinstrumenten en procedures (Working Group 'Assessment of quality in statistics' - Eurostat, 2003).

Er zijn een aantal zaken op te werpen in het vergelijkbaarheidsdebat. Ten eerste moet de vraag gesteld worden of 'goed bestuur', 'een kwalitatieve overheid' of 'een efficiënt openbaar bestuur' wel in alle landen hetzelfde inhoudt. Vooreerst moet er bijvoorbeeld overeenstemming gevonden worden over enkele gerelateerde concepten. Dat dit geen sinecure is, werd reeds eerder aangetoond (cf. 1.2.2). Zo zouden de publieke sector (*public sector*), de administratie (*public administration*) en de overheid (*government*) duidelijker moeten afgelijnd worden. Zoals Rutgers het passend uitdrukt: "*this is not just a matter of finding the right translation*" (Rutgers, 2004). Het probleem binnen het *public administration*-onderzoek is dat er moeilijkheden bestaan in het vinden van eensgezindheid over deze definities en de operationalisering ervan. Bij vertalingen en vergelijkingen is het belangrijk in te zien dat het niet alleen om terminologische verschillen gaat, maar in feite om conceptueel onderscheid.

Ook de betekenis van '*performance*', goede prestaties of kwaliteit van de overheid is niet altijd transponeerbaar naar andere landen (Van de Walle, 2005). Enerzijds is er de vraag wat het uitgangspunt bij het meten van prestaties zou moeten zijn: input, output, outcome, efficiency of effectiveness? Anderzijds is er ook de vraag wat goed bestuur normatief inhoudt. Wat maakt een overheid tot een kwaliteitsvolle en goed presterende overheid? Deze vraag is in feite het startpunt van het kerntakendebat. De vraag wat de overheid op zich moet nemen en welke

prioriteiten ze moet stellen kent internationaal vooralsnog geen eenduidig antwoord.

Grindl (2007) stelt diezelfde vraag en concludeert dat *good governance* niet voor elk land hetzelfde is. Doordat elk land een andere invulling van goed bestuur voor ogen heeft, zijn er ook verschillende medicijnen nodig bij *bad governance*. Om dit euvel te overkomen, lanceert deze auteur een nieuw concept; *'good enough governance'*. Dit idee suggereert dat niet elk governancedeficit tegelijkertijd moet aangepakt worden en dat overheidsinstellingen een product van tijd en plaats zijn. "*Good enough governance, as a concept, means that interventions thought to contribute to the ends of economic and political development need to be questioned, prioritized, and made relevant to the conditions of individual countries*", aldus Grindl.

Een ander probleem dat de vergelijkbaarheid op internationale schaal in het gedrang brengt, is de samenstelling van de landensample. Zo wordt Transparency International vaak onder vuur genomen voor het feit dat ze een erg selectieve landensample hanteert (Galtung, 2005). Meer dan 90% van de wereldbevolking ontbreekt immers in de sample, wat het moeilijk maakt om op mondiaal vlak conclusies over corruptie te trekken.

Vooral slecht scorende landen ondervinden nadelen van onvolledige en soms *biased* landensamples. Ondanks het feit dat ze onderaan bengelen in de rankings van corruptie, overheidsefficiëntie en governance doen ze het mogelijk nog beter dan landen die niet in de meting zijn opgenomen. Het zijn immers vaak niet toevallig arme, onderontwikkelde landen die in de landenlijsten ontbreken om ideologische, praktische of opportunistische redenen.

Arme landen met minder ontwikkelde structuren en minder middelen voor datagaring, vallen vaak uit de boot omdat (harde) data gewoon niet beschikbaar zijn. Ofwel zijn de nationale data moeilijk te consolideren of er kan niet verzekerd worden dat de statistieken betrouwbaar zijn. Om dit praktische probleem te verhelpen wordt er in veel gevallen beroep gedaan op subjectieve indicatoren (Besançon, 2003) om zo bredere internationale vergelijkingen mogelijk te maken. Ook *composite indicators* helpen dit probleem op te lossen. Dit soort indicator zorgt er immers voor dat er kan

vergeleken worden met landen die ontbrekende waarden hebben voor enkele indicatoren (Cherchye et al., 2007).

Maar ook opportunisme speelt soms mee in het niet verschijnen van een land in een ranking. Het is niet ongebruikelijk dat landen, die bijvoorbeeld ondanks hun autoritair regime of dictatuur wel rijkom en/of grondstoffen bezitten, hun imago niet besmeurd willen zien door slechte resultaten op de internationale governanceladder (Orkin, 2002). Zij weigeren dan mee te werken of geven aan geen interesse te vertonen. Doordat ze de slecht scorende aspecten uit de meting houden of helemaal niet in de ranking opgenomen worden, hebben dergelijke landen geen last van *'naming and shaming*'. Op die manier levert de ranking echter een vertekend mondiaal beeld op en doet dit vragen rijzen over de comparabiliteit.

Een laatste obstakel voor de comparabiliteit tussen landen, betreft de meetfouten. De weinige producenten die het aandurven de meetfouten in rekening te brengen, moeten toegeven dat dit de comparabiliteitsfactor in grote mate vermindert. Het aantal significante vergelijkingen daalt met een veelvoud wanneer men meetfouten in rekening brengt. Dit is vermoedelijk de hoofdreden waarom veel producenten met geen woord over meetfouten reppen.

Kuhry (2004) geeft bovendien aan dat er in internationale rangschikkingen van publieke sectoren doorgaans een kleine groep *topperformers* is, naast een kleine groep slecht scorende landen. Maar daartussen zit veruit de grootste groep van landen erg dicht opeen. Hij noemt het de *'messy middle*'. De landen die tot dit grote peloton behoren zijn qua scores moeilijk van elkaar te onderscheiden, wat de vergelijkbaarheidsfactor verder naar beneden haalt.

Tot slot geeft Arndt (2007) treffend aan waarom omzichtigheid geboden is bij vergelijkingen tussen landen en over de jaren heen. Een verschillende landenscore of ranking op een internationale meting van publieke sectoren kan immers een gevolg zijn van vijf factoren. Ten eerste kan de samenstelling van de landensample van de meting gewijzigd zijn. Dit kan ervoor zorgen dat een zelfde prestatie leidt tot een veranderde relatieve score en rangschikking. Ten tweede kan de bronnensamenstelling wijzigingen ondergaan hebben. Doordat ineens andere indicatoren in de meting betrokken worden, kan dat op zich reeds een verandering in de score

of de plaats op de ladder met zich meebrengen. Hetzelfde geldt voor onderliggende indicatoren of bronnen die uit de meting worden geschrapt. Ten derde kan, door het frequent gebruik van subjectieve indicatoren in governance rankings, de perceptie gewijzigd zijn zonder dat een effectieve verandering heeft plaatsgevonden. Dit is het controversiële aspect van het gebruik van subjectieve indicatoren; het is niet altijd duidelijk wat de opinie van respondenten doet bewegen. Ten vierde is het door het relatieve aspect van de rangschikkingen mogelijk dat een vooruitgang of achteruitgang van andere landen de plaatsing van een ander land met een stabiel blijvende score wijzigt. Rankings geven immers enkel aan of een land het beter of slechter doet ten opzichte van de andere landen. Ten vijfde, kan de gewijzigde score ook effectief duiden op een echte verbetering of verslechtering van de prestaties. De moraal van het verhaal is dat de vergelijkbaarheid relatief is en dat er bijzonder voorzichtig met vergelijkingen en rangschikkingen moet omgegaan worden.

> 4.10. Conclusie: Meten, maar met verantwoordelijkheid voor producenten en gebruikers

In het voorgaande gedeelte hebben we duidelijk gemaakt dat de huidige internationale rangschikkingen van publieke sectoren te kampen hebben met een hele reeks kwaliteitsproblemen. De geïllustreerde problemen in het achterhoofd, moet er dus bijzonder omzichtig omgesprongen worden met internationale rangschikkingen van publieke sectoren zoals ze vandaag geproduceerd worden. De data hebben niet de kwalitatieve sterkte om zwaarwichtige beleidslessen te initiëren, noch laten ze toe veel betekenisvolle vergelijkingen te maken. Ze vormen echter een instrument voor overheden om relatieve zwaktes in eigen huis te detecteren en bieden de mogelijkheid enkele verschillende patronen tussen overheden op internationaal vlak te onderscheiden.

Moeten de kwaliteitsproblemen dan tot een *ranking-stop* leiden? Dit lijkt ons een brug te ver. De bewustwording van de kwalitatieve problemen van de governance metingen leidt tot nuance en werkt voorzichtigheid in de omgang met *governance data* in de hand. Ons standpunt luidt dan ook dat de problematiek geen verdere overheidsmetingen mag tegenhouden.

Het is integendeel de overtuiging van de auteurs dat het bijzonder nuttig is om de kwaliteit van de overheid te blijven meten. Onze stelling luidt echter dat de metingen en rangschikkingen van overheden niet blindelings mogen gebeuren. Zowel producenten als consumenten van internationale rangschikkingen van publieke sectoren moeten meer omzichtigheid aan de dag leggen. Enkel dan zullen de data een gepaste bijdrage kunnen leveren aan beleidsdiscussies.

De overhaaste conclusies die momenteel getrokken worden uit de governance rankings, betreffen feitelijk interpretatiefouten. Dit mag zowel de producenten als de gebruikers van de internationale governance rankings worden aangewreven. De producenten leveren immers niet altijd de juiste gebruiksaanwijzing bij hun producten. Zij redeneren dat het communiceren van teveel nuances, scheeftrekkingen en andere gebreken enkel afbreuk doet aan de geloofwaardigheid van hun data. Het gebrek aan transparantie dat hier een rechtstreeks gevolg van is, is een van de steeds wederkerende kritieken die in de internationale literatuur opduikt.

En in het uitzonderlijke geval dat de nuances of tekortkomingen wel goed worden gecommuniceerd met de data, houden gebruikers in politieke, ambtelijke en mediakringen er veelal te weinig rekening mee. Zo stellen onder andere de European Central Bank, het World Bank Institute en Transparency International in hun begeleidende rapporten duidelijk dat de gebreken op kwalitatief vlak om een zeer voorzichtige interpretatie van de empirische bevindingen vragen. Vanuit de hoek van de gebruikers worden de cijfers echter te snel als harde feiten aangenomen. Dat is de gepercipieerde kracht van data, ze lijken voor zich te spreken. Dat deze redenering niet met de realiteit strookt, illustreren we in dit onderzoek op duidelijke manier. Ook gebruikers moeten dus de hand in eigen boezem durven steken.

Internationale rangschikkingen van publieke sectoren *'are here to stay*'. Ondanks de imperfecties die ze vertonen, beantwoorden ze aan de toenemende vraag naar data over de kwaliteit van de overheid. De simpele en uitnodigende aanblik van de rankings draagt bij tot hun populariteit. Het is aan onderzoekers om de gebruikers te wijzen op de gebreken. De bewustwording van het kwaliteitsdeficit moet meer verantwoord gebruik van de data in de hand werken. Tegelijkertijd moeten producenten aangespoord worden om de metingen van publieke sectoren die op een kwalitatief hoger niveau tillen.

Tot slot mogen we ook de rol van de media in de verspreiding van governance data niet over het hoofd zien. Gelet op de macht van het woord die zij hanteren, is het ook noodzakelijk dat zij genuanceerd berichten over de governance metingen. Het is immers niet zelden het geval dat dagbladen of andere media ongenuanceerde boodschappen, selectieve bevindingen tot zelfs foutieve conclusies de wereld in sturen. Zo is het bijvoorbeeld treffend te moeten vaststellen dat, niettegenstaande de uitdrukkelijke waarschuwing van Transparency International over de onbetrouwbare verschuivingen in haar landenklassement, net deze positiewissels jaarlijks de *headlines* van de kranten halen (Ministère des Affaires étrangères et Européennes Française,).

5. Empirisch onderzoek

In hoofdstuk 4 maakten we een *round-up* van de kwaliteitsproblemen die de betrouwbaarheid en hanteerbaarheid van internationale rangschikkingen van publieke sectoren in meerdere of mindere mate in vraag doen stellen. Aan de hand van een literatuuroverzicht en onze eigen bevindingen kwamen we tot een oplijsting van een negental problematische punten waar de huidige governance rankings mee worstelen.

De synthese in hoofdstuk 4 van de kritieken die in de internationale literatuur aan bod komen, vormt als het ware een inleiding op dit hoofdstuk. Het geeft een aanzet tot onze empirische analyse van internationale rangschikkingen van publieke sectoren. Na het scheppen van een breedtezicht in het vorig hoofdstuk, trachten we in dit hoofdstuk een dieptebeeld te krijgen van internationale rangschikkingen van bestuur.

Meer bepaald focussen we ons op de governance rankings die op internationaal vlak een meting maken van (een aspect van) de kwaliteit van overheid. Het gaat daarbij om metingen van het openbaar bestuur in zijn geheel. We sluiten hierbij de rankings uit die een specifiek beleidsdomein meten. Internationale rangschikkingen op het vlak van onderwijs, gezondheidszorg of criminaliteit vallen dus buiten het bereik van dit onderzoek.

Het analyse-instrument voor onze empirische studie is de meta-fiche. De meta-fiche is een bundeling van kwaliteitscriteria -beschrijvend en beoordelend- waaraan we de governance rankings willen toetsen. Het eindproduct van deze analyse is een begeleidende fiche die de binnen- en buitenkant van elke ranking van publieke sectoren onder de loep neemt. Als achtergrondinstrument kan het een bijzonder nuttig instrument zijn dat de gebruiker vertrouwd moet maken met de governance ranking. Het kan dienst doen bij het beoordelen van de kwaliteit van de comparatieve data die jaarlijkse door de verschillende rangschikkingen van publieke sectoren de wereld worden ingestuurd.

Onze werkwijze voor deze analyse van internationale rangschikkingen van publieke sectoren zal in dit hoofdstuk uitgebreid uit de doeken gedaan worden. In de eerste paragraaf schetsen we het onderzoeksplan dat we

hebben doorlopen. Het zal op een bondige manier de doelstellingen en het proces van de analyse beschrijven.

De tweede paragraaf geeft een overzicht van de analyse-eenheden; de rangschikkingen van publieke sectoren. Voor de meta-analyse hebben we acht vooraanstaande governance rankings geselecteerd. In deze paragraaf geven we in eerste instantie uitleg bij deze selectie. Vervolgens geven we een korte beschrijving van elk van de acht internationale rangschikkingen van publieke besturen.

De derde paragraaf ten slotte geeft een beschrijving van de kwaliteitscriteria die we hebben uitgeselecteerd om de governance rankings te evalueren. We illustreren de kwaliteitskenmerken aan de hand van empirische bevindingen uit de individuele meta-fiches. De meta-fiches zelf zouden wegens hun grote reikwijdte, letterlijk uit de voegen van dit rapport barsten. Om de analyseresultaten desalniettemin niet verloren te doen gaan, achten we het interessant om de meta-fiches op een cd-rom bij het rapport bij te leveren. In een bundel bij dit rapport zal de analyse al op papier gemaakt worden van het onderdeel in de verschillende rankings dat een comparatieve meting uitvoert van de bestuursprestaties van de overheid.

De vierde paragraaf tot slot maakt een conclusie op van onze empirische analyse. Een tabel die aangeeft wat de kwalitatief positief ogende punten en de mankementen per ranking bloot legt, vormt het sluitstuk van het empirisch gedeelte in dit rapport. Deze plus- en minpunten zijn gebaseerd op de meta-fiches.

> 5.1. Onderzoeksplan

De empirische analyse bouwt verder op de kritische studies en onderzoeken die in het verleden door enkele internationale auteurs reeds werden gevoerd. We stelden aan de hand van de kennis verzameld in een literatuuronderzoek, een *listing* van kwaliteitspunten voor governance rankings samen. Deze criterialijst heeft geleid tot de meta-fiche.

De meta-fiche heeft als doel een beschrijvende en beoordelende analyse op te maken van de kwaliteiten van de governance rankings. We hebben daartoe de acht bekendste en meest gebruikte internationale rangschikkingen van bestuur geselecteerd. We namen daarvoor zowel historische kenmerken, als nieuwe evoluties in overweging. We hebben ons bovendien vooral op de laatste versie van de governance ranking geconcentreerd(meest recente publicatie voor januari 2008).

De fiche moet helpen gebruikers (en producenten) bewust te maken van de structuur, eigenschappen en kwaliteit van de grote governance rankings. Met dit doel voor ogen hebben we een blik achter de cijfers van de geselecteerde rankings geworpen. Het is immers het scoringsproces, de *machinerie* van de internationale rangschikking van publieke sectoren waarin we geïnteresseerd zijn.

De achtergrondinformatie die deze analyse oplevert, moet meegenomen worden bij de conclusies van de periodiek gepubliceerde governance data. Ze complementeren de naakte data die *an sich* weinig vertellen over de kwaliteit en betrouwbaarheid. De meta-fiche vormt een middel tegen overhaaste besluiten, ongenuanceerde vergelijkingen en niet door de data gesteunde beleidslessen. De meta-fiche legt daartoe de sterktes en zwaktes van de acht geselecteerd rankings bloot.

Een *ranking of the governance rankings* in de enge zin, wilden we echter ten alle prijzen vermijden. We willen immers niet de kortzichtigheid vertonen die we de producenten van governance rankings zelf verwijten. De rangschikkingen van publieke sectoren worden immers door zowel producenten als gebruikers en media te veel opgeworpen als *hitparades*. Waar onderliggende indicatoren een genuanceerd beeld scheppen van de kwaliteit van de overheid, wordt er te veel gefocust op de ene plaats op de ladder die een land aan het eind van de scoringsrit krijgt toegewezen. Dit

misbruik heeft ons ertoe aangezet wel positieve en negatieve punten in de verschillende rankings ten opzichten van elkaar af te zetten, maar geen kortzichtige en ongenuanceerde numerieke ranking op te maken van internationale rangschikkingen van publieke sectoren.

De totstandkoming van de empirische analyse met behulp van de meta-fiche werd reeds in hoofdstuk 3 (cf. 3.1.2) stap per stap beschreven. Hier volstaat het nog eens de stappen kort te overlopen. In eerste instantie hebben we getracht de grote lijnen van kritiek uit het internationaal wetenschappelijk onderzoek te destilleren. Dit leek ons het beste vertrekpunt om kwaliteitscriteria op te stellen die de kwaliteit van de governance rankings in beeld kunnen brengen.

De lijst van potentiële kwaliteitstoetsen voor de onderzochte internationale governance rankings vloeide middels een proefdraai op de *testcase* van de Worldwide Governance Indicators (World Bank Institute) over in de gestandaardiseerde meta-fiche die een analyse moest opmaken van de acht geselecteerde internationale rangschikkingen van publieke sectoren.

De invulling van de meta-fiches vormde het zwaartepunt van dit empirisch onderzoek. Zowel primaire als secundaire bronnen werden ingezet ter evaluatie van de verschillende kwaliteitsaspecten van de respectievelijke rankings. De hoofdbron wordt steevast gevormd door de ranking zelf, gecombineerd met de beschrijvende rapporten en begeleidende technische en methodologische documenten die de producenten voorzien. Deze *de jure* informatie werd aangevuld met kritische studies van onderzoekers over specifieke governance rankings, theoretische en methodologische inzichten en persoonlijke communicatie met producenten van de rankings.

> 5.2. Selectie internationale governance rankings

Omdat het hele landschap van governance rankings te uitgebreid is, was het voor ons noodzakelijk een selectie te maken voor onze empirische toetsing. We hebben daarbij geopteerd voor een selectie van acht governance rankings. Van alle bestaande internationale rangschikkingen van publieke sectoren, hebben we getracht de meest vooraanstaande, bekendste, breedste, frequentst gebruikte en veelvuldig geciteerde rankings eruit te lichten.

Zo kwamen we tot een lijst van de volgende acht governance rankings; The Worldwide Governance indicators (World Bank Institute), The World Competitiveness Yearbook (International Institute for Management Development), The Global Competitiveness Index (World Economic Forum), The Corruption Perceptions Index (Transparency International), The Country Risk Briefing (Economist Intelligence Unit), The Country Policy and Institutional Assessments (World Bank), The International Country Risk Guide (Political Risk Services Group), The Public Sector Performance (European Central Bank).

Deze selectie doorstaat de toets van verschillende andere onderzoeksselecties. Zo maakte Hood onlangs in een soortgelijke oefening, *'rating the rankings'* (2007) genaamd, een selectie van 11 internationale rangschikkingen van publieke sectoren. Drie daarvan hebben betrekking op criminaliteitsstatistieken, nog eens drie behandelen onderwijsindicatoren en twee andere betreffen rankings in het kader van de gezondheid. Deze acht rankings hebben een sectorieel perspectief en behandelen beleid. Omwille van die reden worden ze in onze studie buiten beschouwing gelaten, hier worden immers enkel rankings van bestuur in beschouwing genomen. De overige drie governance rankings, zijn rangschikkingen van bestuur. Het gaat om de Worldwide Governance Indicators, de Corruption Perceptions Index en de Global Competitiveness Index. Dit trio zit ook in onze selectie.

Volgens de World Bank, vervolgens, bestaan er ongeveer 140 governance rankings op dit moment. De meest gebruikte zijn volgens hen; de International Country Risk Guide (PRS Group), de Corruption Perceptions Index (Transparency International), de Country Policy and Institutional Assessments (World Bank), de Worldwide Governance Indicators (World Bank Institute) en de Freedom House-ratings (Freedom House). Buiten deze

laatste ranking, die het democratisch gehalte van een land meet en zo eerder systeemkenmerken dan bestuurskwaliteiten in kaart brengt, hebben ze allemaal een plaats gekregen in onze selectie. Deze twee voorbeelden illustreren dat onze selectie gesteund wordt door andere internationale onderzoeken, analyses en rapporten vanwege wetenschappers en betrokkenen.

Aan de lijst van de World Bank voegden we nog de twee economisch getinte internationale rangschikkingen van publieke sectoren toe die in zo goed als alle governance-discussies aan bod komen; nl. de Global Competitiveness Index (World Economic Forum) en het World Competitiveness Yearbook (International Institute for Management Development). Daarbij plaatsten we nog de Country Risk Briefing (Economist Intelligence Unit), die vanuit het gerenommeerde Britse magazine The Economist wordt gepubliceerd en vooral in Angelsaksische landen veel aanzien geniet. Tot slot kon ook de studie van de European Central Bank, de Public Sector Performance, niet ontbreken. Deze studie deed bij publicatie heel wat stof opwaaien.

De selectie van acht internationale rangschikkingen is een weloverwogen en gebalanceerde sample, die de belangrijkste rankings op het vlak van governance in haar rangen telt. Zoals zal blijken, zitten er voorbeelden bij met een economische inslag, net als rankings die er eerder een politieke visie op nahouden. Zowel rankings met Angelsaksische perspectieven, als overheidsmetingen met een Europees getinte invalshoek behoren tot de selectie. Als deze eerste in onze selectie toch nog het overwicht hebben, is dat omdat Europa op het vlak van internationale rangschikkingen van publieke sectoren wat achtergebleven is op de Angelsaksische organisaties met een liberalistische, non-interventionalistische kijk op de overheid.

> *5.2.1. De Worldwide Governance Indicators*

De Worldwide Governance Indicators zijn zonder twijfel de bekendste en meest geciteerde governance ranking van het laatste decennium. Sinds 1996 voert het World Bank Institute op periodieke basis governance-metingen uit. Tot 2002 gebeurt dit elke twee jaar, sindsdien volgende publicaties elkaar jaarlijks op. Met hun diepgang en breedte als belangrijkste troef hebben ze een belangrijke rol gespeeld in het internationale debat over governance.

Governance wordt door de auteurs gedefinieerd als 'the traditions and institutions by which authority in a country is exercised: the process by which governments are selected, monitored and replaced, the capacity of the government to effectively formulate and implement sound policies and the respect of citizens and the state for the institutions that govern economic and social interactions among them' (Kauffman et al., 2007a). Dit is zoals eerder gesteld een vage omschrijving, het bakent niet echt een specifiek terrein af (Rothstein & Teorell, 2005). Een gebrek dat wel meer governance rankings vertonen.

Het gebrek aan '*construct validity*', een goed omschreven governance-concept, verhindert beleidsmakers om er concrete en beleidsrelevante lessen uit te trekken (Thomas, 2007). Sterker nog; Arndt (Arndt, 2007) ging ten rade bij een staflid van het World Bank Institute over de conceptuele gebreken. Deze gaf in het interview toe dat deze vaagheid de onkunde van de ontwikkelingsorganisatie verhult om overheden van Derde Wereldlanden oplossingen aan te reiken in het ontwikkelingsproces. "*It shifts the focus from what the donors should do, which has become an increasingly embarrassing question since we clearly do not know, towards the question what governments should do*", klonk het.

De Worldwide Governance Indicators bevatten verschillende maatstaven die iets zeggen over de kwaliteit van het overheidsapparaat. Het World Bank Institute verzamelt gegevens rond zes thema's, de zogenaamde Governance Indicators; Voice and Accountability, Political Stability, Regulatory Quality, Rule of Law, Control of Corruption en Government Effectiveness (Kauffman et al., 2007a).

In tegenstelling tot andere governance rankings, maken de auteurs van de Worldwide Governance Indicators -Kauffman, Kraay en Mastruzzi, en in een vroeger stadium ook Loido-Zobotàn- er een punt van deze zes governance indicators niet verder te aggregeren en samen te smelten tot één superscore op het vlak van governance. De zes *governance indicators* staan op zichzelf en zijn allemaal de basis van een afzonderlijke ranking.

De meting van de Worldwide Governance Indicators is gebaseerd op een secundaire analyse van aangebrachte data. De producenten gebruiken indicatoren en data uit een groot aantal externe databronnen, ze voeren zelf geen meting uit. Het idee achter dit meetsysteem wordt door de

auteurs als volgt omschreven: "*individual data sources provide a noisy 'signal' of the broader concept governance*" (Kauffman & Kraay, 2007a). De aggregatie van deze 'ruis' moet een goede benadering van (een van de zes dimensies van) governance opleveren.

Enkel subjectieve indicatoren worden opgenomen in de meting van de Worldwide Governance Indicators. Het gaat hierbij om percepties van experten, burgers en vertegenwoordigers van het (internationale) bedrijfsleven. Een eerste voordeel van het gebruik van zogenaamde zachte data (*soft data*) is dat er geen of minder *'time lag*' optreedt tussen beleid en gevolgen. Ten tweede kunnen er op deze manier ook landen opgenomen worden waarbij harde data ontbreken. De potentiële nadelen zijn echter divers en werden in hoofdstuk 4 al grondig besproken. Het risico bestaat dat het imago gemeten wordt in plaats van de reële prestaties, er is een grotere kans op biases en subjectieve indicatoren zijn niet bevorderlijk voor de transparantie.

Het laatste rapport van de Worldwide Governance Indicators maakt melding van 33 verschillende onderliggende databronnen die gehanteerd worden bij de meting van de Worldwide Governance Indicators. Deze zijn afkomstig van niet minder dan 30 verschillende organisaties. Hiertussen zitten zowel commerciële *risk-rating agencies*, als multilaterale organisaties, Think tanks en NGO's. Alles samen levert dit 310 individuele variabelen op die uiteindelijk geaggregeerd worden in een van de zes *governance indicators*.

Dit enorme aantal indicatoren heeft zijn voor- en zijn nadelen. Een eerste pluspunt is dat op die manier een bredere *country coverage* kan bekomen worden. In totaal worden immers 212 landen opgenomen in de ranking, het meeste van alle governance rankings. Dit betekent niet dat al deze landen een score voor iedere individuele variabele ter beschikking hebben. Elk land dat in minimum drie van de gebruikte databronnen (per governance indicator) opduikt, wordt opgenomen in de ranking.

Een tweede voordeel is dat meer indicatoren en een diversiteit aan externe bronnen ervoor zorgen dat de producent met behulp van statistische modellen de meetfouten kan reduceren (Bertelsmann, 2005). Het World Bank Institute opteert immers voor een *'unonbserved-components model*' bij de aggregatie. Dit systeem bepaalt de wegingsfactoren op basis van de relevantie en representativiteit van de onderliggende databronnen. Hoe

meer de data (gemiddeld) samenhangen met data uit andere bronnen, hoe 'juister' ze worden ingeschat en hoe groter het gewicht dat ze krijgen in de totale meting.

De meetfouten in de Worldwide Governance Indicators blijven significant groot, maar ze worden er tenminste in rekening gebracht. Het World Bank Institute is, samen met Transparency International (cf. 5.2.4), de enige producent van governance rankings, die expliciet vermeldt om rekening te houden met de meetfouten in de data. Meetfouten zijn volgens de producent van deze ranking onvermijdelijk. "*Any proxy for governance is by definition an imperfect measure of broader concepts of governance. There are no silver bullets*", aldus de auteurs (Kauffman & Kraay, 2007a; Kauffman & Kraay, 2007a; Kauffman & Kraay, 2007b).

Meer nog, de Worldwide Governance Indicators worden begeleid door betrouwbaarheidsintervallen. Deze hebben tot doel betekenisvolle verschillen en vergelijkingen te onderscheiden van onbetrouwbare en ongenuanceerde analyses. Het is de verdienste van deze producent dat ze gebruikers er nadrukkelijk op wijst dat enkel scores, waarbij de betrouwbaarheidsintervallen niet overlappen, op een significante wijze met elkaar mogen vergeleken worden.

Een eerste nadeel van het enorme aantal gebruikte indicatoren is, dat het als gebruiker bijna onmogelijk is in te schatten wat er nu eigenlijk gemeten wordt. De conceptuele precisie heeft te lijden onder het aggregatieproces (Knack & Mannig, 2000). De auteurs spreken zelf treffend van een '*general snapshot*' (Kauffman et al., 2006), te vaag om er beleidslessen uit te trekken.

De afwezigheid van een onderliggende governance-theorie plaatst nog meer vraagtekens bij het conceptuele kader dat aan de basis ligt van de meting. Zowel de keuze voor de databronnen, als de daaropvolgende selectie van indicatoren wordt niet onderbouwd. De zes governancecomponenten worden niet omschreven, de normatieve betekenis moet integraal afgeleid worden uit de indicatoren die er aan ten grondslag liggen.

Een tweede nadeel is dat het noodgedwongen aggregatieproces als vereiste heeft dat de individuele scores genormaliseerd worden. Het gemiddelde wordt op 0 vastgelegd met een standaarddeviatie van 1. De scores kunnen

hierdoor enkel relatief ingeschat worden. Een betere score betekent enkel een beter dan gemiddelde prestatie, en zegt niets over de vraag of een land er in absolute waarde is op vooruitgegaan of niet.

Een derde nadeel van het grote aantal indicatoren en de diversiteit van bronnen is het gebrek aan transparantie dat eruit voortvloeit. Men moet immers steeds teruggrijpen naar de onderliggende databronnen om de methodologische en technische kenmerken van de variabelen in te schatten. Sommige basisindicatoren zijn bovendien moeilijk toegankelijk of enkel beschikbaar tegen betaling van een aanzienlijke prijs (Arndt & Oman, 2006).

Tot voor kort liet het World Bank Institute zelfs volledig na de individuele scores van de landen kenbaar te maken. Hier kwam echter felle kritiek op. Dit deed de transparantie immers geen goed. Bovendien kregen de overheden zo niet de noodzakelijke informatie aangereikt over de reden van hun (slechte) score. Pas in 2005 hebben de auteurs gehoor gegeven aan deze kritiek en stelde het World Bank Institute het grootste deel van de onderliggende data beschikbaar van de gebruikers. Deze data worden echter wel enkel in genormaliseerde vorm en als gemiddelde per databron weergegeven, waardoor slechts een deel van de initiële informatieve waarde overblijft. Enkele bronnen, zoals de Country Policy and Institutional Assessments van de World Bank, van de African Development Bank en van de Asian Development Bank geven op verzoek van de producenten (en de opgenomen landen) zijn niet toegankelijk tot de basisscores.

Het vierde minpunt betreft de vergelijkbaarheid. Zoals aangegeven, kan het gebruik van een dusdanig groot aantal indicatoren uit verschillende databronnen ervoor zorgen dat meer landen ingang vinden in de ranking, maar deze medaille heeft ook een keerzijde. Landen die data ontbreken op een aantal indicatoren worden wel opgenomen in de meting, maar kunnen nog moeilijk op een betekenisvolle manier vergeleken worden met andere landen. De scores zijn immers op een andere manier tot stand gekomen, er wordt in feite iets anders gemeten. De auteurs erkennen dit probleem, maar vinden het belangrijker meer landen te kunnen opnemen dan een perfecte comparabiliteit te garanderen.

Het groot aantal gebruikte externe databronnen moet bovendien enigszins genuanceerd worden. De 33 databronnen die aan de basis liggen van de meting van de Worldwide Governance Indicators zijn immers niet allemaal

even innovatief. Integendeel, er bestaat een grote interdependentie tussen de verschillende bronnen. Veel bronnen gebruiken soortgelijke of veelal dezelfde basisdata om hun dataset of ranking op te bouwen. Het gaat bovendien vaak over *private sector surveys*, die eenzelfde economische bias vertonen. In plaats van dit te compenseren, versterkt de aggregatiemethode van de Worldwide Governance Indicators deze vertekening nog verder. Databronnen die het dichtst aanleunen bij de resultaten van de andere gebruikte bronnen, krijgen immers een grotere weging toebedeeld. Als er in veel gevallen echter steeds dezelfde data gebruikt worden, is het niet onlogisch dat deze bronnen een grotere samenhang met de gemiddelde governance-databron kunnen voorleggen.

De Worldwide Governance Indicators worden ook bekritiseerd omwille van hun gepercipieerde bias. Diverse auteurs stellen dat de data te lijden hebben onder een *private sector bias*. De auteurs beweren dat vooral bevolkingssurveys aan de basis van de meting liggen, maar de feiten bewijzen het tegendeel. De *expert opinions* en de *firm surveys* hebben samen een gewicht van meer dan 60% in de totale meting. Vooral deze categorieën vertonen bovendien een grote vertekening in de richting van de economische elite.

Ook de bruikbaarheid van de Worldwide Governance Indicators als beleidsinformatie wordt hier en daar ter discussie gesteld. Niet alleen is het moeilijk in te schatten wat de meting juist inhoudt en zijn er vraagtekens te plaatsen bij vergelijkbaarheid over de jaren heen en tussen landen onderling. Er is ook de vraag welke kwaliteiten *good governance* bijbrengt. Zo hebben veel auteurs vragen bij de leereffecten die overheden uit de Worldwide Governance Indicators kunnen halen.

Het debat wordt ook binnen het World Bank Institute gevoerd. Er gaan stemmen op om meer in de richting van *action-worthy indicators* te gaan in plaats van de *outcome indicators* die nu het grootste deel van de ranking uitmaken. Dit soort indicatoren moet beter inspelen op beleidsproblemen door tastbare en beleidsbeïnvloedbare aspecten van de overheid te onderzoeken. Vooralsnog moeten zelfs de auteurs van de Worldwide Governance Indicators toegeven dat de data weinig beleidsondersteunend gewicht met zich meedragen: "*the data are a blunt tool to use in formulating policy advice. To move forward, we need better data and better tools*" (Kauffman et al., 2000).

Dit heeft de Verenigde Staten er niet van weerhouden om de Worldwide Governance Indicators te gebruiken in hun Millennium Challenge Account. Met dit programma lanceerde president Bush het plan om miljarden dollars aan ontwikkelingshulp op basis van de data van het World Bank Institute te gaan verdelen. Dit tegen het nadrukkelijke advies in van de producent zelf (Arndt & Oman, 2006).

Het onderdeel dat wij in de meta-fiche van naderbij hebben bekeken, is de component/indicator 'Government Effectiveness'. Deze dimensie leunt het dichtst aan bij een meting van de kwaliteit van het openbaar bestuur. De meting staat voor een beoordeling van de effectiviteit van het overheidsbestuur. Boyle (2007) vertaalt de indicator als '*a measurement of the competence of the bureaucracy of public service delivery*'.

> 5.2.2. Het World Competitiveness Yearbook

Het World Competitiveness Yearbook van het International Institute for Management Development maakt sinds 1989 op jaarlijkse basis een rangschikking van landen op, gebaseerd op competitiviteit. Competitiviteit wordt door de producent gedefinieerd als 'de mate waarin een land het klimaat weet te scheppen en te onderhouden dat economische groei voor de bedrijven en welvaartsverhoging voor de bevolking met zich meebrengt'.

De meting van de overheidskwaliteit is in deze studie dus te vertalen als de zoektocht naar overheden die een (gepercipieerd) gunstig klimaat scheppen voor ondernemingen. Een minimum aan regulering, lage taksen en een vlotwerkende administratie, zijn hierbij de subjectieve prioriteiten. Een non-interventionistische overheid die vrijhandel nastreeft, vormt het ideaalbeeld.

De doelgroep van deze data bestaat dan ook bijna integraal uit investeerders en bedrijfsleiders die er een adviserend instrument in zien ten behulpe van hun financiële beslissingen in het buitenland. De ranking is dan ook minder geschikt voor vergelijkingen van publieke sectoren of beleidsonderzoek (Van de Walle, 2006a). Het is wel interessant om te zien hoe bedrijfslui de competitiviteit en kwaliteit van een (land) overheid inschatten.

Het International Institute for Management Development gebruikt voor haar meting zowel subjectieve (1/3) als objectieve (2/3) indicatoren. De harde data komen uit verschillende internationale bronnen van organisaties zoals de Organisation for Economic Co-operation and Development, de World Bank, de World Trade Organisation, de United Nations Educational, de Scientific and Cultural Organization, het International Monetary Fund en de 50 partnerinstellingen die met het Institute verbonden zijn. De subjectieve data komen dan weer van eigen hand. Ze worden verzameld in de Executive Opinion Survey, die periodiek onder 3.700 bedrijfsleiders en managers uit de onderzochte landen wordt gehouden.

Een kritiek aan het adres van deze survey betreft het geringe aantal respondenten. Zo worden de zachte data voor België berekend op basis van de opinie van slechts 44 bedrijfsleiders, wat erg weinig is. Het gaat bovendien enkel om ondervragingen binnen de economische wereld, wat een bias oplevert in de richting van de *business elite*. Zeker als men weet dat de aangezochte bedrijven betreffen ook nog eens overwegend *multinationals* zijn.

Het International Institute for Management Development kiest er met 246 variabelen voor om een erg groot aantal indicatoren te gebruiken in haar meting. Op die manier kiest ze er tegelijkertijd voor om slechts een beperkt aantal (geïndustrialiseerde) landen in de ranking op te nemen. Voor een groot deel van de wereld zijn immers geen vergelijkbare harde data beschikbaar.

Dit lijdt tot een van de probleempunten van het World Competitiveness Yearbook. Het covert slechts een klein deel van de wereld, van een mondiale rangschikking is geen sprake. Bovendien maken de 55 landen die wel in de meting worden betrokken, praktisch allemaal deel uit van het rijke Westen.

Het International Institute for Management Development is een van de weinige producenten die ook oog heeft gehad voor regio's in haar ranking. In 2003, 2004 en 2005 werden onder andere Zhejiang, Bavaria, Catatonie, Ile-De-France, Schotland, Rhon-Alps, Maharashtra, Lambardije en Sao Paolo opgenomen. Vlaanderen kreeg geen plaats in de rangschikking. Dit initiatief bloedde echter snel dood.

De internationale rangschikking van publieke sectoren van het International Institute for Management Development onderscheidt 4 grote pijlers van competitiviteit. Economic Performance, Government Efficiency, Business Efficiency en Infrastructure zijn de vier geïdentificeerde dimensies van de totale meting. Elk van deze vier factoren wordt verder onderverdeeld in vijf subfactoren.

De 246 gebruikte indicatoren worden op een gestructureerde manier geaggregeerd tot één score, de competitiviteitsindex van elk land. Er wordt geopteerd voor een vaste weging, waarbij de factoren elk voor 0,25 worden meegerekend en de onderliggende subfactoren (onafhankelijk van het aantal variabelen dat ze vertegenwoordigen) voor 0,05. De totaalscore vormt de basis voor de uiteindelijke landenrangschikking. Daar waar de ranking en de index jaarlijks met grote mediabelangstelling worden bekendgemaakt, zijn de individuele scores enkel op commerciële basis te raadplegen. Toegang tot deze basisdata kost ongeveer 3.500 euro. Er worden geen betrouwbaarheidsintervallen meegegeven, wat het risico op ongenuanceerde vergelijkingen vergroot.

Ook hier geldt de opmerking dat de scores en totaalranking enkel relatieve prestaties weergeven. Doordat de basisdata genormaliseerd werden voor het aggregatieproces, betekent een betere score enkel dat het land vooruitgang maakt ten opzichte van de andere landen. Of dit ook in absolute waarde een effectief hogere competitiviteit met zich heeft meegebracht, is niet bekend.

Het onderdeel van het World Competitiveness Yearbook dat ons het meest interesseert, de component die de kwaliteit van het overheidsbestuur meet, is de 'Government Efficiency'. Het betreft een van de vier pijlers van de competitiviteitsmeting. Er wordt gepeild in welke mate het overheidsbestuur een gunstig ondernemingsklimaat weet te creëren. De meting van deze component is overwegend subjectief getint.

Meer dan 60% van de 72 onderliggende indicatoren vertegenwoordigen de opinie van bedrijfsleiders in elk van de gequoteerde landen. Deze onderliggende indicatoren worden verder onderverdeeld in subcategorieën Public Finance, Fiscal Policy, Institutional Framework, Business Legislation en Societal Framework. De sterke economische en fiscale inslag van de meeste indicatoren wijst op de selectieve invulling van de meting van de

kwaliteit van het overheidsbestuur. In feite wordt er gemeten in welke mate de overheid (administratie) als een obstakel of risico voor investeerders wordt gepercipieerd.

> *5.2.3. De Global Competitiveness Index*

De Global Competitiveness Index, krijgt haar beslag in het Global Competitiveness Report, dat jaarlijks door het World Economic Forum wordt uitgebracht. Deze comparatieve competitiviteitsmeting wordt in basisvorm reeds sinds 1979 uitgevoerd. Voor 2001 werkte het World Economic Forum samen met het International Institute for Management Development, erna gingen beide organisaties hun eigen weg. Sindsdien maken ze alle twee jaarlijks een internationale competitiviteitsrangschikking op. Niettegenstaande het feit dat ze vaak in een adem worden vernoemd, leggen ze ietwat verschillende accenten en kiezen ze op meetvlak voor een erg verschillende aanpak.

Met haar competitiviteitsmeting doet de ranking van het World Economic Forum een beschouwing van de *'business operating environment*'. Hiertoe worden de instellingen, het beleid en de omgevingsfactoren getoetst op hun bijdrage aan het algemene productiviteitsniveau van een land.

Tegenhanger International Institute for Management Development stelt in haar meting van de competitiviteit van landen objectieve indicatoren boven subjectieve variabelen te verkiezen. "Zachte data dienen enkel ter aanvulling van de harde data", zo luidt het. Wat moeilijk met objectieve indicatoren te meten is, kwantificeert ze aan de hand van bevragingen. Het World Economic Forum daarentegen gaat echter voornamelijk uit van subjectieve indicatoren om de competitiviteit vast te stellen.

Doordat er minder harde data aan de meting te pas komen, kunnen er meer landen in de ranking opgenomen worden. De laatste rangschikking (2007) telde 131 landen, meer als het dubbele van de tegenhanger van het International Institute for Management Development. Op die manier wordt een groot stuk van de wereld door de ranking gecoverd. Samen zijn de opgenomen landen goed voor 98% van het *World's Gross Domestic Product*.

De methodologie achter de competitiviteitsmeting veranderde sterk in de voorbije jaren. In 2005 werden slechts 36 indicatoren gemeten om tot de

geaggregeerde score te komen. Een jaar later steeg dit aantal tot 96 indicatoren, die onderverdeeld werden in 9 pijlers. De laatste rangschikking van 2007 splitste de ene marktpijler verder op in 4 nieuwe pijlers, en telde in totaal 116 indicatoren. Deze instabiele samenstelling maakt comparabiliteit over de jaren heen moeilijk, zo niet onmogelijk.

De weging van de verschillende pijlers is variabel. Dit wil zeggen dat het gewicht per land wordt bepaald. De weging is gebaseerd op de economische ontwikkelingsfase waarin elk land zich bevindt. De idee achter deze variabele weging is dat overheden andere accenten leggen op basis van het ontwikkelingsstadium waarin het land zich bevindt (gemeten op basis van het GDP/capita). Ondanks het feit dat een variabele weging volgens veel auteurs de meest correcte manier is om een aggregatie uit te voeren, draagt dit niet bij tot de comparabiliteit. De verschillende landenscores zijn immers op een verschillende manier tot stand gekomen, waardoor appelen met peren worden vergeleken.

De ranking beroept zich voor het grootste deel op resultaten van de Global Competitiveness Survey, die het World Economic Forum op periodieke basis afneemt bij 11.000 *'top management business executives'*. Het gaat om subjectieve bevragingen die weinig representatief zijn. De respondentensamples per land zijn immers klein. Zo wordt de Belgische score afgeleid uit slechts een 40-tal opinies. Het objectieve indicatorengedeelte is afkomstig van internationale organisaties zoals het International Monetary Fund, de World Bank en de United Nations.

De basisdata zijn ook bij de World Competitiveness Index niet vrij te raadplegen. De ranking wordt met veel tromgeroffel gepubliceerd en is op de website gemakkelijk in te kijken. Maar de onderliggende data zijn enkel tegen betaling beschikbaar, al vormt de lagere prijs een kleinere drempel dan bij veel andere governance rankings.

Het World Economic Forum stelt in haar rapport dat de ranking 'als doel heeft om overheden te helpen bij het identificeren van hindernissen voor economische groei en hen bij te staan in de formulering van een aangepaster (economisch) beleid'. De Batselier (2007) merkte op het MOVI-Colloquium "Vlaanderen 'anders' bekeken" echter terecht op dat de rangschikking van het World Economic Forum geen goed instrument vormt voor beleidslessen. Minder dan een derde van de indicatoren kan -en dan

nog in beperkte mate- beïnvloed worden door overheidsoptreden. Het is dus uiterst moeilijk om via gerichte beleidsmaatregelen een significante scoreverhoging te bekomen.

Het onderdeel van de meting dat vanuit ons oogpunt relevant is in de zoektocht naar metingen van de kwaliteit van de overheid, is de 'Government Efficiency'-factor. Ook hier valt op dat de rol van de overheid vrij eenzijdig wordt ingevuld. De indicatoren geven een beeld weer van een overheid die het bedrijfsleven zo weinig mogelijk hindert. De respondenten (bedrijfsleiders) en het doelpubliek (bedrijfsleiders en investeerders) in acht genomen, mag dit wederom niet verbazen.

> *5.2.4. De Corruption Perceptions Index*

De Corruption Perceptions Index is verruit de meest bekende en meest frequent gebruikte overheidsranking op het vlak van corruptie. Aangezien het ontbreken van corruptie als een belangrijke kwaliteitsdimensie van een overheid wordt beschouwd, is deze ranking van Transparency International een van de belangrijkste internationale rangschikkingen van publieke sectoren.

Transparency International stelt sinds 1995 jaarlijks een ranking op op basis van de gepercipieerde corruptie in de publieke sector. De laatste ranking telde maar liefst 180 landen. De producent baseert zich daarbij volledig op externe databronnen. Corruptie wordt hierbij gedefinieerd als 'het misbruik van publieke macht voor persoonlijk of privaat gewin'.

In totaal selecteerde Transparency International 14 externe databronnen, die door 12 verschillende internationale organisaties en instellingen werden samengesteld. De voorwaarde voor de inclusie van een externe databron is dat de bron de laatste twee jaar een meting moet gedaan hebben van het algemene corruptiepeil in het publieke en politieke domein. De meting van (waargenomen) corruptie mag daarbij overigens niet te veel doorkruist worden door indicatoren en vragen over andere zaken.

Bovendien luidt de tweede voorwaarde dat de databronnen tot een landenrangschikking moeten geleid hebben. Deze laatste voorwaarde komt er feitelijk op neer dat elke bron in elk van de landen eenzelfde methodologie moet gehanteerd hebben. Dit maakt een zekere mate van

comparabiliteit mogelijk. Transparency International draagt er zorg voor dat de inclusie van databronnen in de meting volgens een transparant proces verloopt. In tegenstelling tot vele andere governance rankings, geeft de Corruption Perceptions Index een gedetailleerde en weloverwogen verantwoording van de bronnenselectie. Ook de indicatorenkeuze en verdere omschrijving van wat er door de variabele precies gemeten wordt, komt ruimschoots aan bod in het begeleidend rapport bij de Corruption Perceptions Index. Dit bevordert de transparantie van de ranking.

Galtung (2005) stelde in zijn evaluatie van de Corruption Perceptions Index dat de ranking een grote *private sector bias* vertoont. Hij stelde het feit aan de kaak dat zo goed als alle onderliggende bronnen van private en commerciële instellingen komen. Dit soort bron vertoont een vertekening in de richting van de *private sector*-visie, en de respondenten ervan zijn overwegend mannelijk, hoog opgeleid en eerder welstellend. Transparency International heeft hier echter de laatste jaren aan gewerkt en heeft nieuwe bronnen in de meting opgenomen. Hierdoor komen vandaag reeds 5 van de 12 bronnen uit internationale en multilaterale organisaties die niet van een *private sector*-bias verdacht kunnen worden.

De corruptiemeting maakt enkel gebruik van subjectieve indicatoren. De visie van de auteurs is immers dat '*international surveys on perceptions of corruption serve as the most credible means of compiling a ranking of nations*'. De validiteit van deze stelling kan moeilijk nagetrokken worden, aangezien er geen harde data bestaan voor de reële omvang van corruptie in elk van de landen (Van de Walle et al., 2004). Zowel expertenanalyses, als bevragingen van zakenlui worden betrokken in de meting. Er zijn geen bevolkingspolls bij de onderliggende databronnen.

In totaal liggen dus 12 verschillende surveys aan de basis van de Corruption Perceptions Index. De ranking van het International Institute for Management Development en die van de Political and Economic Risk Consultancy worden van twee verschillende jaren gebruikt, omdat er een andere expertenevaluatie aan de basis ligt van beide jaarrapporten. In totaal komt de producent dus aan 14 databronnen. Geaggregeerd levert dit een rangschikking van landen op, die gebaseerd is op de gepercipieerde mate van corruptie bij ambtenaren en politici. Er wordt door de onderliggende databronnen immers geen onderscheid gemaakt tussen administratieve corruptie en politieke corruptie.

Transparency International is samen met het World Bank Institute een van de enige producenten van governance rankings die waarschuwt voor ongenuanceerde conclusies en betekenisloze vergelijkingen. Ze brengt de meetfouten in rekening en stelt voor elke score betrouwbaarheidsintervallen op. Enkel landen waarvan deze *'confidence intervals'* niet overlappen kunnen op significante wijze van elkaar onderscheiden worden. De grootte van de betrouwbaarheidsintervallen hangt samen met het aantal gebruikte bronnen voor een land. In de ranking zitten immers landen die een score verkregen op basis van 3 bronnen, het minimum voor inclusie. Daartegenover staat een land als Indonesië dat in 11 verschillende databronnen is terug te vinden. Hoe meer databronnen aan de basis van een score liggen, hoe accurater de score en hoe kleiner het betrouwbaarheidsinterval.

Bovendien raadt de producent af om vergelijkingen te maken op basis van de ranking, aangezien een veranderde positie op de internationale ladder even goed een gevolg kan zijn van een veranderde methodologie of andere samplesamenstelling dan van een verslechterde of verbeterde prestatie op vlak van corruptie. Door de veelvuldige fundamentele wijzigingen van de methodologie en de jaarlijks veranderende landensamenstelling is het gevaarlijk vergelijkingen te trekken. Zelfs Transparency International zelf waarschuwt hiervoor. Het is dan ook opmerkelijk dat de Amerikaanse regering in haal Millennium Challenge Account veel gewicht toekent aan de Corruption Perceptions Index. De ranking van Derde Wereldlanden geldt zelfs als het *in-or-out-criterium* voor het ontvangen van ontwikkelingssteun. Transparency International heeft zich al meermaals gedistantieerd van dergelijk foutief gebruik van haar data (Arndt & Oman, 2006).

Als internationale NGO heeft Transparency International er altijd voor geijverd haar ranking zeer toegankelijk te maken. De onderliggende databronnen worden kenbaar gemaakt, net als de totaalscore per land en het daarmee verbonden betrouwbaarheidsinterval. De basisdata worden echter niet vrijgegeven, daarvoor moet de gebruiker zich wenden tot de onderliggende databron zelf.

Aangezien de geaggregeerde corruptie-indicator van Transparency International zich in zijn geheel uitspreekt over de gepercipieerde corruptie in de publieke sector en binnen het overheidsapparaat, bekijken we de globale index als indicator van een goed of slecht overheidsbestuur.

> 5.2.5. *De Country Risk Briefing*

De Country Risk Briefing van de Economist Intelligence Unit is een van de meest gebruikte governance rankings komende van een commercieel *risk-rating agency*. De producent maakt deel uit van The Economist Group, waaronder ook het bekende Britse zakenblad *The Economist* valt. Sinds 1997 produceert ze een waaier aan *country intelligence*, waaronder ook de Country Risk Briefing.

Deze internationale rangschikking van publieke sectoren meet een aantal risicofactoren voor het bedrijfsklimaat van een land. De belangrijkste doelstelling van deze ranking is immers bankiers, institutionele investeerders en bedrijfsleiders te adviseren bij investeringsbeslissingen in het buitenland. De uitkomst is een internationale rangschikking van 150 landen op het vlak van hun *'overall business operating risk'*.

Om tot de totaalscore te komen, beoordelen *in-house* analisten (vanuit de hoofdzetel in Londen) en *in-country* experten (in de onderzochte landen) van de Economist Intelligence Unit de landen op tien risicodomeinen: Security risk, Political stability risk, Government effectiveness risk, Legal & regulatory risk, Macroeconomic risk, Foreign trade & payments risk, Financial risk, Tax policy risk, Labour market risk en Infrastructure risk. De landen krijgen na aggregatie van de individuele indicatoren een indexscore (0-100), een plaats in de rangschikking (1-150) en een risico-categorisering (A-E).

Het feit dat alle scores tot stand komen op basis van een experten-evaluatie, maakt het mogelijk een volledige score voor elk land te bekomen. Ze doen dit op basis van een analysemodel, dat economische en politieke aspecten combineert. Er zullen met andere woorden in de regel geen *missing data* zijn.

Zoals vaak het geval is bij internationale rangschikkingen van publieke sectoren opgesteld door private adviesbureaus, laat ook de Country Risk Briefing het afweten op conceptueel en in mindere mate ook op transparantievlak. Het conceptuele kader is geheel gehuld in mist; er wordt nergens verantwoording afgelegd voor de selectie van dimensies en indicatoren. Dit belemmert ook de transparantie van de data. De indicatoren zijn bovendien moeilijk toegankelijk en voor de toegang tot de

individuele data moet betaald worden. De producent vraagt 5.500 pond voor een algemene toegang tot de individuele landenscores per jaar. Bovendien is het uiterst moeilijk voor gebruikers om in te schatten op welke basis de experten een situatie als risicovol hebben ingeschat.

De component die een vergelijking mogelijk maakt op het vlak van de overheidskwaliteit, betreft het 'Government Effectiveness risk'. Het gaat om een van de tien risicodomeinen die de producent identificeert in zijn risicoanalyse. De meting omvat 7 onderliggende indicatoren: Policy Formulation, Quality of Bureaucracy, Excessive Bureaucracy/Red-tape, Vested Interests/Cronyism, Corruption, Accountability of Public Officials en Human Rights. Ook deze dimensie wordt volledig op basis van subjectieve indicatoren gemeten. Experten van de Economist Intelligence Unit beoordelen deze indicatoren op een schaal van 1 tot en met 7 en doen daarbij beroep op extern beschikbare informatie en contactpersonen in de onderzochte landen.

> *5.2.6. De Country Policy and Institutional Assessments*

De Country Policy and Institutional Assessments zijn als internationale meting van publieke sectoren vooral gekend vanwege hun rol in de ontwikkelingswereld. De producent, de World Bank, beroept zich immers op deze ranking bij de allocatiebeslissingen van ontwikkelingssteun en goedkope leningen (IDA-middelen).

Sinds 1977 doet de World Bank in haar Country Policy and Institutional Assessments een meting van 'de kwaliteit van het beleid en het institutionele kader van een land'. Er wordt voornamelijk ingeschat in welke mate de instellingen en het beleid van ontwikkelingslanden doelstellingen als armoedevermindering, duurzame groei en een efficiënte aanwending van ontwikkelingssteun kunnen waarmaken.

De hoofdbetrachting is immers in te schatten welke progressie ontwikkelingslanden boeken en of de ontwikkelingssteun die aan de Derde Wereld wordt besteed zijn vruchten afwerpt. Critici stellen echter dat het gaat om een verdoken verplichting tot conformering aan de World Bank-koers: neoliberale economische beleidsmaatregelen, sterke instellingen (die onder andere de *property rights* moeten beschermen) en een

bedrijfsvriendelijk ondernemingsklimaat zouden op deze manier impliciet worden gepromoot.

Een andere kritiek aan het adres van de Country Policy and Institutional Assessments, die in het verlengde hiervan ligt, betreft de economische invalshoek van deze ranking. Schmitz (2007) stelt dat er veel te veel de nadruk ligt op economische indicatoren in de meting. "Op die manier wordt een goed beleid gelijk gesteld met een beleidskoers die een goede werking van de markt verzekert", zo luidt zijn kritiek. Er moet gezegd worden dat de World Bank deze kritiek niet in de wind heeft geslagen. Het heeft door de jaren heen haar methodologie bijgewerkt van een nadrukkelijke sociaaleconomische focus naar het opnemen van meer governance-aspecten in de meting.

De gehanteerde meting van vandaag is gebaseerd op vier grote clusters: Economic Management, Structural Policies, Policies for Social Inclusion/Equity en Public Sector Management and Institutions. De vier pijlers worden verder onderverdeeld in 16 subindicatoren. Al deze indicatoren worden gemeten aan de hand van experten-evaluaties. Stafleden, waaronder zowel experten als economici, van de World Bank maken een analyse van de situatie in de verschillende landen. Deze analyse leidt dan weer tot een subjectieve score. Samen vormen deze scores de basis voor een ranking van 136 landen. Deze indicatoren worden aangevuld met de indicator Portfolio Performance.

De Country Policy and Institutional Assessments zijn voor bepaalde landen van vitaal belang, aangezien de verdeling van de middelen van de International Development Association (IDA) ermee verbonden is. Onder het '*aid targetting*'-principe speelt de score die in de ranking bekomen wordt een cruciale rol in het ontvangen (en de hoeveelheid) van IDA-steun.

Dit was een van de redenen dat er enkele jaren geleden een golf van kritiek opkwam tegen het uitblijven van een verhoogde transparantie in de Country Policy and Institutional Assessments. De World Bank koos er immers gedurende lange tijd voor om de individuele scores niet vrij te geven. Maar vooral ontwikkelingsexperten hadden grote problemen met deze koers. Veel critici vonden het immers immoreel om miljarden dollars ontwikkelingsgeld te verdelen op basis van scores die de het daglicht niet haalden. Ook de donorlanden die financiële hulp ter beschikking stellen via het International

Development Association drongen aan op meer transparantie en beschikbaarheid van de data.

Na een externe review, besloot de World Bank in de zomer van 2006 te starten met het openbaar maken van de Country Policy and Institutional Assessments-data. Enkel de landen die in aanmerking komen voor IDA-steun zien hun score publiekelijk bekend gemaakt worden. Voor de andere landen geldt deze nieuwe regel (nog) niet.

De World Bank tracht met deze ranking overheden in de Derde Wereld bij te staan in hun ontwikkeling. Daarom kiest het ook bewust voor zoveel mogelijk '*actionable*' indicatoren. Deze indicatoren worden zo opgesteld dat overheden er via beleidsplannen dadelijk op kunnen inspelen. Het *governance*-aspect in de meting is duidelijk aanwezig. De World Bank probeert daarbij vooral de sociale en structurele dimensies te benadrukken.

Het onderdeel dat het dichtst aanleunt bij een meting van de overheidskwaliteit, is de subcomponent Quality of Public Administration. Experten beoordelen dit governance-aspect door vier indicatoren te beoordelen; Policy Coordination and Responsiveness, Service Delivery and Operational Efficiency, Merit and Ethics en Pay Adequacy and Management of the Wage Bill.

> *5.2.7. De International Country Risk Guide*

De International Country Risk Guide, is samen met de producten van the Economist Intelligence Unit, een van de belangrijkste bronnen van *'country risks*'. Vooral internationale investeerders maken in grote mate gebruik van deze internationale rangschikking van publieke sectoren.

Sinds 1984 voert de Political Risk Services Group op maandelijkse basis risicoanalyses uit op financieel, economisch en politiek vlak. De directe aanleiding voor het ontstaan van deze governance ranking was de financiële crisis na de val van de Iraanse Sjah begin jaren 1980. De economische voorspellingsmodellen waren duidelijk tekortgeschoten in het voorspellen van deze crisis, waardoor nieuwe -meer politiek getinte- risicomodellen werden opgestart.

De Political Risk Services (PRS) Group maakt een risico-evaluatie van 140 landen. Het doet daartoe een meting van drie risicodimensies. De Financial Risk Index, de Economical Risk Index en de Political Risk Index omvatten samen 22 indicatoren. De eerste twee indices worden gemeten op basis van harde data en objectieve criteria. De politieke component wordt daarentegen samengesteld op basis van subjectieve interpretaties. Het zijn experten uit de International Country Risk Staff die een beoordeling op een vierpuntenschaal maken van alle subjectieve criteria.

De scores op deze drie dimensies worden op het eind geaggregeerd aan de hand van een vast wegingsschema. Het politieke gedeelte krijgt een weging van 0,50, de twee andere dimensies tellen voor 0,25 mee in de eindafrekening. De onderliggende basisindicatoren krijgen een weging mee die afhangt van een toegekend aantal risicopunten. Het eindresultaat is een *overall country score* (0-100) en in plaats van een uni-dimensionale ranking, een categorisering volgens *'very low risk*' (0-49.5), 'moderate risk' (50-79.5) of *'very high risk*' (80-100).

Het ontbreekt expertenscores vaak aan transparantie, aangezien de evaluaties van dit soort respondenten veelal moeilijk in te schatten zijn en niet reproduceerbaar blijken. De PRS Group tracht hier iets aan te doen en levert bij de individuele scores ook de informatie en de data op basis waarvan de risicoscores tot stand kwamen. Bovendien worden ook de subjectieve interpretaties van deze informatie en data door de experten ter beschikking gesteld. Op het vlak van transparantie mag dit initiatief een voorbeeld voor andere organisaties zijn.

Een negatief punt in de beschikbaarheid en transparantie van de International Country Risk Guide is dat de datatoegang wel erg duur is. Er wordt maar liefst 8.000 dollar aangerekend voor toegang tot de individuele risicoscores. Naast de negatieve impact op de transparantie, bevordert dit ook de leereffecten niet. Veel landen zullen immers omwille van deze hoge prijs aarzelen om deze informatie binnen te halen.

De subcomponent 'Bureaucratic Quality' geeft een score voor de kwaliteit van het overheidsbestuur. Ze maakt deel uit van de Political Risk Index en wordt dus gemeten op basis van expertenopinies. De indicator meet de institutionele sterkte en de kwaliteit van de ambtenarij, doet een inschatting van de macht en expertise die ambtenaren hebben, en

onderzoekt ten slotte in welke mate politieke veranderingen leiden tot (drastisch) oponthoud in overheidsdiensten of in de beleidsuitvoering. Een lage score (en een klein risico) wijst op een autonome administratie, die onafhankelijk is van politieke druk en over een duurzaam mechanisme beschikt van rekrutering en opleiding.

> *5.2.8. De Public Sector Performance*

Enkele jaren geleden maakte ook de European Central Bank een internationale rangschikking van publieke sectoren op. Het ging om een studie die in een meting en rangschikking voorzag van publieke sectoren van voornamelijk Europese landen. De daaraan verbonden studie *'Public Sector Efficiency: an International Comparison*' (Alfonso et al., 2005) wekte veel media-aandacht omdat het om een innovatieve meting ging van de prestaties en efficiëntie van de publieke sector. De studie vloeide voort uit de Lissabon-strategie die de EU-lidstaten in 2000 onderschreven.

De vergelijkingsoefening tussen 23 geïndustrialiseerde landen moest de Europese landen aan de hand van *best practices* helpen om de prestaties van hun eigen publieke sector te verbeteren. De kwaliteit van de publieke sector wordt immers als een belangrijke factor in het realiseren van de Lissabon-agenda gezien: de EU tot meest competitieve economie ter wereld maken.

De Public Sector Performance is een alomvattende rangschikking, die naast de effectiviteit ook de efficiëntie van de overheid nagaat. De effectiviteit van de publieke sector wordt gedefinieerd als 'de effecten die de overheid voortbrengt'. Bovendien combineert de meting bestuurs- met beleidselementen. De meting steunt op 7 subcomponenten en in totaal 15 indicatoren. Een eerste deel, het *'opportunity indicators*'-gedeelte, neemt de administratie, het onderwijs, de gezondheidszorg en de infrastructuurskwaliteit in rekening. Tot het tweede deel behoren de *'Musgravian indicators'*, waarbij allocatie, herverdeling en stabiliteit worden nagegaan. Waar deze tweede dimensie eerder (macro-)economisch getint is, peilt het eerste gedeelte vooral naar de '*quality of life*'.

Deze internationale rangschikking van publieke sectoren is anders dan de anderen die we in deze sectie hebben besproken in de zin dat het om een

eenmalige oefening ging. De meting werd uitgevoerd in 2003 en verzamelde comparatieve data voor 1990 en 2000. Er is verder geen opvolging voorzien.

De European Central Bank beroept zich in haar meting zowel op objectieve als op subjectieve indicatoren om tot de Public Sector Performance te komen. De 4 onderliggende indicatoren van het bestuursaspect (Administrative performance) zijn van subjectieve aard. De Executive Opinion Survey van het World Economic Forum doet dienst voor drie van de vier indicatoren, de grootte van de schaduweconomie wordt overgenomen van een subjectief onderzoek dat Schneider (2002) uitvoerde. Ook hier zijn de percepties dus grotendeels vertekend in de richting van de economische topklasse. De overige 11 indicatoren komen uit officiële statistieken van diverse internationale organisaties. Onder andere bronnen van de World Bank, de Organisation for Economic Co-operation and Development en het World Economic Forum liggen aan de basis van de harde data.

De Public Sector Performance kreeg bij haar publicatie veel aandacht onder media, beleidsmakers en onderzoekers. Maar ook deze governance ranking bleef niet gespaard van kritiek. Critici plaatsten flink wat vraagtekens bij de onderliggende structuur en de keuze van de indicatoren van de meting.

De vraag die telkens terugkwam, was: "Meet deze index werkelijk de effectiviteit en de efficiëntie van de publieke sector?" (Van Roosbroek & Van Dooren, 2006). De keuze voor de indicatoren wordt nauwelijks verantwoord, net zoals de gewichten. Waarom maakt de kwaliteit van de communicatie en transportinfrastructuur 14% uit van de index die de effectiviteit van de publieke sector meet, en de werkloosheid nog geen 5%? Wat is er belangrijker: de economische prestaties van een land, of de verdeling van de welvaart? Of de stabiliteit van de economische groei? Of onderwijs? In wezen zijn dit allemaal politieke vragen, waar niet één juist antwoord op bestaat. Het gemak waarmee deze kwestie door de European Central Bank-studie genegeerd wordt, verbaast dan ook.

Bovendien wijst het beperkt aantal onderliggende indicatoren ook op een selectief beeld van de overheid. Een goed presterende overheid wordt opgevat als een overheid die voorzieningen moet treffen op het vlak van onderwijs, gezondheidszorg, infrastructuur en hoort bovendien over een goede administratieve werking te beschikken. Daarnaast moet ze zich de taken toe-eigenen in een goede sociale verdeling te voorzien, economische

stabiliteit te verzorgen en goede economische prestaties neer te zetten. Waarom deze aspecten in de meting van de overheidseffectiviteit werden opgenomen en andere niet, wordt ook nergens verantwoord. "*The selection from a vast pool of variables reflects our best attempt*", is de gemakkelijke uitleg van de auteurs (Alfonso et al., 2005).

De ranking neemt ook geen meetfouten in rekening en laat dus na betrouwbaarheidsintervallen bij de data voegen. Het feit dat veel landen zich in het zogenaamde *'messy middle*' bevinden, met scores die erg dicht bij elkaar liggen, vergroot het risico op foutieve vergelijkingen.

Het gedeelte van de index dat het overheidsbestuur evalueert, is de component *'Administrative performance'*. Deze 'effectiviteit van de administratie' is één van de zeven componenten van de 'Public Sector Performance'. Ze is samengesteld uit vier variabelen. Een optelling van de prestaties op het vlak van corruptie (survey), red tape (survey), kwaliteit van het juridische systeem (survey) en de schaduweconomie (schatting) levert de European Central Bank een score voor de effectiviteit van de administratie op. Naast de inschatting van de schaduweconomie, die slechts een klein onderdeel vormt van de bezigheden van een klassieke administratie, hebben we te maken met drie indicatoren die tot stand komen door inschattingen van experten of *insiders*. De 'effectiviteit van de administratie' meet dus eerder de percepties van bevoorrechte getuigen, in plaats van een feitelijke situatie. Vandaar de stelling dat de studie van de ECB met een mindere score voor ons land vooral het imagoprobleem van de Belgische administratie bevestigt (Van de Walle, 2006a). De producent beweert niettemin een objectief prestatieniveau van de verschillende publieke sectoren vast te leggen.

> 5.3. Analyse

In deze paragraaf zullen we de meta-fiche voorstellen. Het gaat om een gestandaardiseerd analyse-instrument waarin we de verschillende internationale rangschikkingen van publieke sectoren op hun kwaliteit hebben geëvalueerd. De meta-fiche kan gezien worden als een uitgebreide steekkaart die evaluerende met beschrijvende elementen combineert. Aan de hand van 33 kwaliteitscriteria hebben we de acht governance rankings uit onze selectie geanalyseerd, beschreven en beoordeeld. De 33 kwaliteitsindicatoren zijn onderverdeeld in 14 kwaliteitscomponenten. Deze werden gegroepeerd in 3 kwaliteitsdimensies.

Ten eerste is er de conceptuele dimensie. In deze dimensie hebben we proberen vast te stellen wat de producent in haar ranking tracht te meten en te vergelijken. We zijn ook nagegaan of de meting valide is: 'Wordt er gemeten wat er beweerd wordt te meten?' Ook vanuit welke overheidsomschrijving de overheidsmeting is opgebouwd, komt hier aan bod. Tot slot beschrijven we voor elke ranking de aard van de indicatoren; individueel of samengesteld, subjectief of objectief. De indicatoren vormen immers het zwaartepunt van de rankings.

Ten tweede hebben we de technische dimensie van de metingen beoordeeld. Dit onderdeel legt de nadruk op de statistisch-technische kant van de meting. We zijn voor de internationale rankings onder andere nagegaan hoeveel observaties in de meting vertegenwoordigd zijn. Ook de oorsprong van de data is een belangrijk element van de kwaliteitsvaststelling van de ranking. Daarnaast hebben we ook de analyse-eenheid en het *time-frame* geanalyseerd. De databeschikbaarheid en transparantie zijn eveneens van groot belang in een kwaliteitsvolle comparatieve oefening en werden verder onderzocht. We hebben voorts de duurzaamheid van de meting en de betrouwbaarheid van de vergelijkingen ingeschat. Hiervoor hebben we wijzigingen onder de loep genomen in het meetmodel, de wegingsschema's en de landensample. Aangezien de meeste governance rankings voor een groot deel of soms zelfs volledig op subjectieve indicatoren steunen, hebben we ook de survey-eigenschappen die aan de basis liggen van de zachte data in beschouwing genomen.

De derde dimensie betreft de politiek-administratieve dimensie. Dit onderdeel moet een antwoord bieden op de vragen: 'Wie meet?', 'Waarom

wordt er gemeten?', en 'Welke doelgroep heeft de producent met zijn meting voor ogen?'. Dat deze aspecten niet triviaal zijn, hebben we reeds in voorgaande hoofdstukken geïllustreerd. Ze spelen een belangrijke rol in het perspectief van waaruit de data van de verschillende governance rankings moeten benaderd worden. Als een internationale rangschikking van publieke sectoren wordt opgesteld vanuit commerciële overwegingen zal die er een andere benadering op nahouden dan een governance ranking opgemaakt door een NGO die een bepaalde problematiek aan de kaak wil stellen. Ook de eindgebruiker is van belang. Managers uit *multinationals* zullen vanuit investeringsoverwegingen naar andere dingen op zoek gaan in internationaal comparatief onderzoek van publieke sectoren dan bijvoorbeeld beleidsmakers van Derde Wereldlanden die advies zoeken over hoe ze verdere ontwikkeling vanuit de overheid kunnen stimuleren.

In wat volgt, geven we een overzicht van de 14 kwaliteitscomponenten die we voor elk van de acht governance rankings hebben onderzocht. Enerzijds wordt voor elke component beschreven wat we ermee in kaart hebben willen brengen, welke kwaliteitscriteria ze clusteren en hoe we deze criteria hebben ingevuld. Anderzijds illustreren we ook elk kwaliteitscriterium aan de hand van opmerkelijke of illustratieve fragmenten uit de meta-fiches. Daar waar de vorige paragraaf (cf. 5.2) naast een beschrijving van de geselecteerde internationale rangschikkingen van publieke sectoren, in feite een horizontale lezing maakten van onze meta-fiche, houdt deze pararaaf een verticale lezing van het analyse-instrument in.

> *5.3.1. Conceptuele dimensie*

'Kiezen is verliezen', luidt het spreekwoord. Bij het opstellen van een index of een rangschikking, moet het onderzoeksconcept gepreciseerd en afgelijnd worden. Impliciet gebeurt dit in feite aan de hand van de selectie van de indicatoren. Het antwoord op de vraag welke indicatoren in de meting worden opgenomen, is van groot belang in de invulling van de meting. Producenten kunnen niet om het feit heen dat er keuzes moeten gemaakt worden. Keuzes maken veronderstelt echter normatieve beoordeling, politieke overweging en selectieve voorkeur. Deze keuzes zijn vaak ingegeven door de aard van de producent, het doel van de studie of simpelweg de (on)beschikbaarheid van data. In de Conceptuele dimensie hebben we getracht zicht te bieden op de beoogde betekenis van de

meting, de werkelijke betekenis en de keuzes die gemaakt werden bij de selectie van de indicatoren. Samengevat deed deze pijler navraag naar 'wat er gemeten wordt'.

In eerste instantie hebben we de omschrijving bekeken die de producent aan de meting meegeeft. Hoe heeft de auteur van de governance ranking het concept gedefinieerd? Op welke manier heeft de producent getracht de prestaties van de publieke sector te meten?

In tweede instantie zijn we de validiteit van de metingen nagegaan. Wordt er wel degelijk gemeten wat de producent beweert te meten? Aan de hand van enkele basiseigenschappen van de meting van publieke sectoren, hebben we getracht in te schatten in welke mate de omschrijving van de producent overeenkomt met de reële betekenis van de meting.

Een normatief element dat een belangrijk licht op de governance meting werpt, is de opvatting die de producent er over de publieke sector op nahoudt. De invalshoek van waaruit de auteurs van de internationale rangschikking van publieke sectoren naar de overheid kijken, speelt een bepalende rol in de manier waarop de resultaten van de meting moeten worden geïnterpreteerd. Het vormt het derde item dat we onder de conceptuele dimensie onderzocht hebben.

De omschrijving van het gemeten concept door de producent -hetzij van de efficiëntie van de overheid, hetzij van de effectiviteit van de overheid- is één zaak, maar de feitelijke normatieve invulling van de meting is een andere zaak. De inhoudelijke betekenis van de meting is echter paradoxaal genoeg af te meten aan de indicatoren die eraan ten grondslag liggen. Een theoretische totstandkoming van de meting van overheidskwaliteit blijkt een utopie, waardoor het de indicatoren zijn die in grote mate betekenis verlenen aan de meting. Het is dan ook vitaal voor een goed begrip van de governance rankings, een blik te werpen op de selectie en aard van de onderliggende indicatoren. Met het oog hierop, hebben we telkens aangegeven in welke mate gebruik gemaakt wordt van individuele of samengestelde indicatoren.

Hierbij aansluitend is een erg belangrijk aspect in de invulling van een overheidsmeting ook of men gebruik maakt van objectieve of subjectieve indicatoren. Daar waar de eerste een vaststelling zijn van een officiële

statistiek, zijn de tweede gebaseerd op een inschatting, een perceptie of een opinie. Zonder ons hier te willen uitspreken tegen de een of de ander, is het belangrijk om voor ogen te houden dat ze niet met elkaar mogen gelijk gesteld worden.

> *5.3.1.1. Validiteit*

Een eerste component die we onder de conceptuele paraplu onderzocht hebben, is de validiteit. Samengevat komt de validiteitsvraag neer op de vraag: 'Worden de goede dingen gemeten?' Validiteit heeft echter vele dimensies, wij hebben voornamelijk de conceptuele validiteit en de constructievaliditeit onderzocht.

De *'construct validity'* houdt volgens Dill en Soo (2005) in dat de diverse prestatie-indicatoren die geselecteerd werden om de kwaliteit van de overheid te meten, ook daadwerkelijk relevante factoren of dimensies betreffen van een goed presterende overheid. Aangezien er geen unanimiteit bestaat over de invulling van een meting van de overheidskwaliteit, de efficiëntie van de administratie of governance, is het echter onmogelijk de verschillende indicatorenselecties af te zetten tegenover een standaardformule. Wel hebben we de mate van selectiviteit trachten aan te geven of bepaalde vertekeningen geïdentificeerd.

De *'conceptual validity'* sluit rechtstreeks aan bij de keuze van de indicatoren en het meetmodel. De vraag wordt gesteld naar de onderbouwing van deze keuzes. Is de selectie van de indicatoren met andere woorden theoretisch te verantwoorden? Ligt er een (governance-)theorie of een grondige analyse aan de basis van de structuur van de meting, de selectie van de indicatoren en de wegingsfactoren (Kuha & Gaskell, 2005)?

Aangezien de metingen van publieke sectoren steevast gestoeld zijn op indices is het noodzakelijk dat de producent zijn keuzes voor het model en de invulling van indicatoren verantwoordt. Deze keuzes hebben immers een grote invloed op de uiteindelijke indicator.

De rode draad in de analyse van de validiteit van de acht geselecteerde governance rankings, is een fundamenteel gebrek aan onderbouwing van de conceptuele keuzes. Indicatorenselecties berusten veeleer op

beschikbaarheid dan op theoretische gronden en ook over de motieven achter de wegingsschema's laten de producenten de gebruiker en de onderzoeker steevast op hun honger zitten.

De meting van de Worldwide Governance Indicators door het World Bank Institute vormt hier een treffende illustratie van. Deze internationale rangschikking van publieke sectoren doet een meting van het governance-gehalte in 212 landen. De producent identificeert zes dimensies van governance: Voice and Accountability, Political Stability and Absence of Violence, Government Effectiveness, Regulatory Quality, Rule of Law en Control of Corruption. Waar deze dimensies vandaan komen en waarom deze clusters werden verkozen boven andere, wordt echter niet door de producent verklaard. De auteurs komen niet verder dan te stellen dat *'these indicators are defined to correspond to what the authors consider to be 'fundamental governance concepts*' (Kauffman et al., 2007a).

Het World Bank Institute gebruikt maar liefst 310 indicatoren in haar governance-meting. Deze worden via een wegingsschema geaggregeerd tot de zes Worldwide Governance Indicators. Achter de aggregatiemethode van deze governance ranking gaat dan weer wel een expliciete visie schuil. De auteurs (Kauffman & Kraay, 2007a) gebruiken het *'unobserved components model'* om de veelheid aan onderliggende indicatoren te aggregeren tot de zes Governance-pijlers. Het World Bank Institute doet beroep op externe databronnen om de eigen ranking samen te stellen. Het idee is dat de governance scores uit elke onderliggende primaire databron een '*unobserved*' signaal van '*true governance*' in dat land meten. Hoe hoger de correlatie van deze score met de gemiddelde scores uit andere bronnen, hoe beter ze overeenstemt met een meting van *'true governance*' en hoe groter het gewicht van die databron bijgevolg moet zijn.

Er zijn ook rankings met een overzichtelijkere meting. Het geringe aantal indicatoren dat de European Central Bank bij de meting van de Public Sector Performance gebruikt, maakt deze governance ranking -positief bekeken- een stuk helderder dan de anderen. Deze governance ranking van 23 geïndustrialiseerde landen (waaronder België) meet de effectiviteit van de overheid af op basis van 15 indicatoren, die onderverdeeld zijn in zeven subcomponenten. Enerzijds is er een gedeelte dat *'opportunity indicators*' meet; hier worden enkele overheidstaken beoordeeld die de levenskwaliteit van haar burgers moeten verhogen. Anderzijds is er een gedeelte dat de

'Musgravian indicators' meet; dit aspect peilt naar de prestaties op economisch gebied; meer bepaald de allocatie, de herverdeling en de stabiliteit in elk land.

Voor de helderheid van de meting ten gevolge van het gering aantal indicatoren, moet echter wel een prijs betaald worden. Het zorgt voor een selectief beeld van de overheid. Het beperkt aantal indicatoren doet twijfels rijzen over de betrouwbaarheid van de meting. Voorts scoort deze governance ranking ook slecht op vlak van conceptuele validiteit. De keuze voor de onderliggende componenten wordt niet onderbouwd. Waarom bepaalde aspecten wel en andere dan weer niet in de meting worden opgenomen, is erg onduidelijk. Tot slot krijgt ook het gehanteerde wegingsschema nergens een degelijke verantwoording.

> *5.3.1.2. Opvatting publieke sector*

Kiezen is niet alleen verliezen, het is ook aangeven wat belangrijk wordt geacht. De rankings zijn opgebouwd op basis van een groep variabelen en subindicatoren die door de producent werden geselecteerd. Kiezen voor indicator a, b en c en niet voor indicator x, y en z impliceert een politieke keuze, een normatieve overweging. De ideologie, de theorie of de idee die achter deze keuze schuilgaat, is een belangrijk element in de totstandkoming van de governance meting.

De invalshoek van de meting wordt mede bepaald door de opvatting van de publieke sector die de producent als uitgangspunt neemt. Een effectieve of efficiënte overheid hangt immers sterk af van de gekleurde vraag wat een overheid tot een goed presterende overheid maakt.

Een letterlijke definitie of omschrijving van 'de goede overheid' door de ogen van de producent van de ranking is meestal niet terug te vinden. Aan de hand van de geselecteerde indicatoren, de wegingsschema's en de evaluatiecriteria kan er wel een bepaald beeld opgeroepen worden van wat een kwaliteitsvolle en goedpresterende overheid in een bepaalde internationale rangschikking van publieke sectoren inhoudt.

Simpel gesteld kan men bij een internationale rangschikking van publieke sectoren, die landen vergelijkt op het vlak van 'Government effectiveness' (cf. Worldwide Governance Indicators, World Bank Institute), aan de hand

van de indicatorenselectie en de scoringsmodaliteiten inschatten hoe de publieke sector wordt opgevat.

Op het vlak van de indicatorenselectie, zien we vaak een selectieve en enge opvatting van de overheid. Niet zelden wordt een effectieve of efficiënte overheid geoperationaliseerd in overwegend economisch getinte indicatoren. In dat geval wordt 'overheidseffectiviteit' opgevat als een overheid die zich voluit inzet voor een bedrijfsvriendelijk klimaat, een minimum aan regulering oplegt aan economische actoren, lage taksen instelt en over een administratie beschikt die *red tape* beperkt. Over bijvoorbeeld sociale herverdeling, gender-aspecten of milieubeleid zal in dergelijke metingen zo goed als niets terug te vinden zijn.

Ook de scoringsmodaliteit verraadt een bepaalde visie op de publieke sector in de verschillende overheidsmetingen. Een treffend voorbeeld is de indicator in verband met geldende arbeidsrechten die in verschillende comparatieve overheidsmetingen wordt opgenomen. In de governance rankings die duidelijk een adviserende rol voor ogen hebben voor bedrijfsleiders, managers of (internationale) investeerders, wordt deze indicator eerder negatief opgevat. Hoe meer arbeidsrechten er in het onderzocht land heersen, hoe meer de bedrijven er gebonden worden. Arbeidsrechten worden *in casu* opgevat als een belemmerende factor ten aanzien van het economische verkeer. Of dit een correcte evaluatie is of niet, is hier niet de kwestie. Het gaat om een normatief beeld ten aanzien van de overheid en diens taken. Deze opvatting van de publieke sector bloot leggen, draagt bij tot een beter begrip van de governance ranking en de daarbij horende data.

De analyse van de acht geselecteerde governance rankings toont een overwegend economisch geïnspireerde opvatting van de overheid in de meeste rankings. Hierbij wordt een goede overheid opgevat als een overheid die niet in de weg loopt van de bedrijven, het economische verkeer vlot doet verlopen en lage ondernemingsbelastingen instelt.

Het mag niet verbazen dat internationale rangschikkingen van publieke sectoren zoals de World Competitiveness Index (World Economic Forum), het Global Competitiveness Yearbook (International Institute for Management Development) en de Country Risk Briefing (Economist Intelligence Unit) vooral onderzoeken in welke mate de overheid zorgt voor

een goed ondernemingsklimaat en de stabiliteit voor gevestigde internationale bedrijven en financiële investeerders weet te garanderen. Het gaat hier immers over commerciële risico-inschattingen die vooral door bedrijven worden geconsulteerd. Economische agenten hechten vooral belang aan een *Laissez-faire* overheid met een minimum aan regulering en taksen. Dit weerspiegelt zich vooral in de selectieve (economische) indicatorenselecties.

De Country Policy and Institutional Assessments van de World Bank, die ingezet worden bij het ontwikkelingsproces in de Derde Wereld, is in een andere hoek te situeren, nl. die van de ontwikkelingsleer. Een goede overheid wordt in deze optiek opgevat als een overheid die er in slaagt een beleid te voeren en een institutioneel kader uit te bouwen dat duurzame groei en armoedebestrijding nastreeft en efficiënt omgaat met de ontwikkelingsgelden. Op deze manier wordt gekeken wat de *return on investment* is voor de toegekende ontwikkelingshulp: wordt de financiële steun efficiënt aangewend en wordt er ontwikkeling gerealiseerd?

> *5.3.1.3. Aard van indicator*

In eerste instantie hebben we voor de rankings onderzocht of ze gestoeld zijn op een individuele indicator, dan wel een samengestelde indicator of index. Een samengestelde of geaggregeerde indicator herbergt een grotere complexiteit dan een individuele indicator, aangezien er keuzes voor onderliggende variabelen moeten gemaakt worden en er ook normatieve wegingsschema's dienen opgesteld te worden. Dit maakt het vaak moeilijker om de uiteindelijke score te interpreteren. Het bemoeilijkt eveneens de leereffecten voor overheden, aangezien het vaak moeilijk in te schatten is hoe hervormingen op bepaalde domeinen (op specifieke indicatoren) zich zullen vertalen in de geaggregeerde index.

Daar tegenover staat wel dat samengestelde indicatoren een belangrijke vereenvoudiging maken van de bundel informatie die er aan ten grondslag ligt. Ze kunnen uit een diversiteit aan onderliggende informatie toch de *bottom line* vatten. Dit zorgt ervoor dat de samengestelde indicatoren interessanter zijn voor beleidsmakers en media. Door diverse indicatoren op te nemen in de meting van een complex fenomeen, worden bovendien ook de meetfouten afgevlakt waardoor kleinere betrouwbaarheidsintervallen ontstaan.

Er zijn dus zowel argumenten voor als tegen het gebruik van geaggregeerde indicatoren op te werpen. Saltelli stelt zelfs dat het bijna ondenkbaar is dat men er in het debat over het gebruik van *'composite indicators*' ooit zal uitgeraken (Saltelli, 2005). Aangezien veel aspecten van de kwaliteit van de overheid echter niet rechtstreeks kunnen gemeten worden, komt er zo goed als altijd een *'composite indicator*' aan te pas in de governance-metingen. Dit verplaatst het inhoudelijk zwaartepunt van de rankings naar de onderliggende indicatoren.

Ook binnen deze basisvariabelen is er verdere differentiatie mogelijk. Enerzijds kan ook hier het onderscheid gemaakt worden tussen individuele en geaggregeerde indicatoren. Hier gelden dezelfde voor- en nadelen als hoger beschreven. Anderzijds is er het onderscheid tussen objectieve en subjectieve indicatoren.

Aan de ene kant zijn er de objectieve indicatoren, die harde data voortbrengen. Aan de andere kant zijn er de subjectieve indicatoren - perceptiemetingen en *opinion polls*- die aan de basis van de zachte data liggen. Het betreft een belangrijk onderscheid, waarbij de eerste feitelijkheden registreren en de tweede de visie of inschatting van respondenten over een bepaald fenomeen weerspiegelen. Percepties mogen immers niet verkeerdelijk als feiten verkocht worden of omgekeerd.

Aangezien er slechts een beperkt aantal officiële statistieken en harde data beschikbaar zijn ter indicatie van een goed presterende overheid, zien veel producenten van internationale rangschikkingen van publieke sectoren zich haast genoodzaakt een beroep te doen op surveys om bepaalde aspecten van de overheidswerking te meten. Objectieve data op *crosscountry* basis zoeken, is op het governance-vlak bijna onbegonnen werk. De onbeschikbaarheid van harde data is met andere woorden een directe aanleiding voor het veelvuldig gebruik van subjectieve indicatoren in governance rankings.

Op zich stelt dit geen probleem. In tegendeel; subjectieve indicatoren zijn complementair aan objectieve indicatoren. Waar harde data kwantiteit meten, voegen de zachte data een kwaliteitsdimensie toe. Bovendien registreren de objectieve indicatoren eerder de *de jure* situatie, terwijl de subjectieve indicatoren beter geschikt zijn om de *de facto* toestand te vatten. Het is een misvatting dat het gebruik van objectieve indicatoren de

meting ook objectiever en accurater maakt. De selectie van objectieve indicatoren is immers gebaseerd op een subjectieve beoordeling. Welke harde data wel en welke niet op te nemen, is een keuze die ook normatief gekleurd is. Niettemin moet men bij subjectieve indicatoren meer op zijn hoede zijn voor biases, vertekende beeldvorming. Onder andere om deze bezorgdheid tot uiting te laten komen, lijkt het ons belangrijk aan te geven in welke mate de geselecteerde rangschikkingen van publieke sectoren een beroep doen op objectieve, dan wel subjectieve indicatoren.

De analyse van de acht geselecteerde internationale rangschikkingen van publieke sectoren wijst uit dat zo goed als alle rankings gestoeld zijn op een index. Het gaat daarbij om een geaggregeerde indicator (*composite indicator*) die samengesteld is op basis van een reeks onderliggende indicatoren die een aspect van de hoofdindicator meten.

Op de Worldwide Governance Indicators (World Bank Institute) na worden alle governance metingen geaggregeerd tot één score. Dit betekent dat de nationale overheden elk een totaalscore krijgen toebedeeld die als basis geldt voor een uni-dimensionale landenrangschikking. De Worldwide Governance Indicators daarentegen leiden tot zes *country scores* per land, één voor elk van de verschillende governance-dimensies. De auteurs van deze governance indicators maken de expliciete keuze om geen *overall score* op te maken, de verschillende dimensies zijn niet verder samen te smelten.

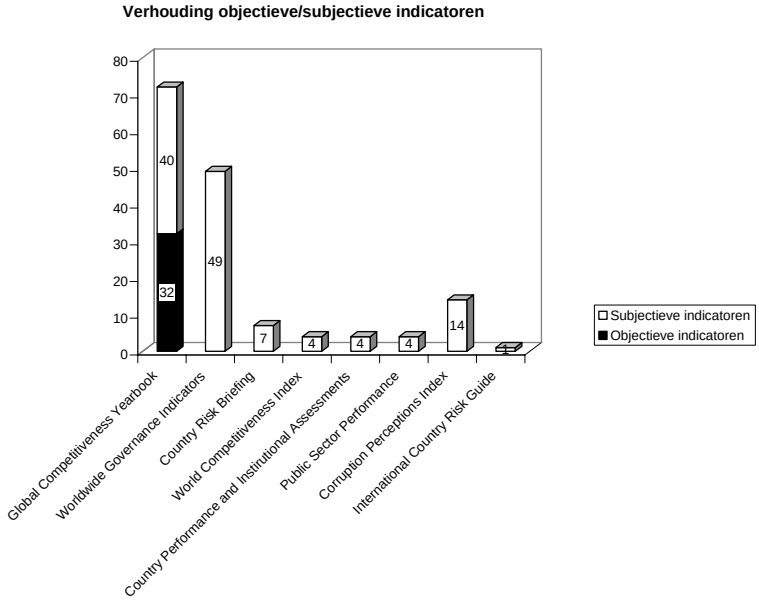
Het onderdeel dat een score toekent aan de werking van het overheidsapparaat bestaat ook in zo goed als alle gevallen uit een geaggregeerde indicator. Het is immers moeilijk om in één indicator - objectief of subjectief- het hele concept van een effectieve overheid te vatten. Enkel de governance ranking van de Political Risk Services Group, de International Country Risk Guide, bevat een indicator 'Bureaucratic Quality', die niet is samengesteld op basis van onderliggende indicatoren en gewoon als item deel uitmaakt van een expertenquotering.

In de meeste gevallen is de meting van de overheidskwaliteit dus geen finaliteit op zich, maar maakt ze deel uit van de inschatting van een hoger liggend concept. Bij het Global Competitiveness Yearbook (International Institute for Management Development) en de World Competitiveness Index (World Economic Forum) vormt de overheidskwaliteit een onderdeel van de

competitiviteit van een land. Bij de International Country Risk Guide (PRS Group) maakt ze deel uit van een *overall risk rating* van elk land. Voor het World Bank Institute is de berekening van de Government Effectiveness een einddoel op zich; het is een van de zes governance-pijlers die gemeten worden.

Daarnaast is het ook overduidelijk dat de overheidskwaliteit in overwegende -haast exclusieve- mate gemeten wordt via subjectieve indicatoren (zie figuur 3). Enkel het International Institute for Management Development gebruikt ook objectieve indicatoren in haar samenstelling van de indicator Government Efficiency. 32 van de 72 gebruikte indicatoren komen uit officiële statistieken, het merendeel van de data blijft dus ook in deze ranking subjectief. De andere zeven governance rankings baseren zich bij de inschatting van de overheidskwaliteit volledig op subjectieve indicatoren, hetzij aan de hand van surveys, hetzij op basis van experten-analyses.

Figuur 3: Verhouding objectieve/subjectieve indicatoren



> *5.3.2. Technische dimensie*

De tweede dimensie die we in de meta-fiches onderzocht hebben, is de technische dimensie. Dit gedeelte behelst onder andere de statistische kwaliteit, de databeschikbaarheid en de betrouwbaarheid van de cijfers. Ook de analyse-eenheid, de dataoorsprong en de survey-eigenschappen komen er in aan bod.

In een zin tracht deze dimensie een antwoord te bieden op de vraag 'wordt er goed gemeten?'. Er wordt hierbij hoofdzakelijk gekeken naar de statistisch-technische kant van de meting, naar de manier waarop de data tot stand zijn gekomen, hoe de data gepresenteerd worden en welke betrouwbaarheid ze vertonen.

Onder de technische koepel hebben we in eerste instantie de statistische kwaliteit van de governance ranking en diens data in beschouwing genomen. Binnen deze component hebben we eerst bekeken hoeveel observaties aan de basis van de rangschikking liggen. Daarnaast hebben we aangegeven of de producent van de governance ranking een betrouwbaarheidsinterval voor de data berekende. Dit is immers noodzakelijk om zicht te krijgen op de meetfouten in de databron en een voorwaarde om betrouwbare en betekenisvolle vergelijkingen in de ranking te kunnen maken. Vervolgens hebben we de revisiemechanismen onderzocht die op de betrouwbaarheid van de data en het dataverwerkend proces moeten toezien. Ook de aanwezigheid van een officiële, onafhankelijke en externe instelling die voor een statistische revisie moet zorgen, werd gecheckt.

In tweede instantie hebben we onderzoek gedaan naar de oorsprong van de data. De data die de individuele variabelen voeden en mee aan de basis van de index liggen, kunnen immers door de producent zelf verzameld zijn of overgenomen zijn uit externe databronnen.

Ten derde hebben we geprobeerd de analyse-eenheid van de geselecteerde internationale rangschikkingen van publieke sectoren in beeld te brengen. Vooreerst werd nagegaan hoeveel landen in de ranking worden opgenomen, de zogenaamde *country coverage*. Daarnaast hebben we ook onderzocht of naast landen ook regio's in de ranking een plaats hebben verworven.

Onder de component 'time frame' geven we mee hoe ver de periodieke governance meting al teruggaat. Naast het ontstaansjaar is ook de periodiciteit van de internationale rangschikking van publieke sectoren van belang voor het creëren van tijdsreeksen. Ten slotte hebben we nog getracht na te gaan hoe groot de tijdspanne is tussen datagaring en de publicatie van de verschillende databronnen.

Ook de beschikbaarheid van de data en de transparantie van de methodologie van de governance ranking belangen ons in deze studie aan. De bruikbaarheid van de internationale rangschikkingen van publieke sectoren wordt immers in belangrijke mate beïnvloed door de mate waarin de data instrumenteel zijn voor beleidsvorming. Minimumvoorwaarden daarvoor zijn dat de data goed toegankelijk moeten zijn en dat er voldoende duidelijkheid en openheid is op het vlak van de conceptuele keuzes in het meetproces, het dataverzamelingsproces en het dataverwerkingsproces. We hebben in dit kader in de meta-fiches aangegeven of de producent de mogelijkheid heeft voorzien om de data af te laden in een dataverwerkend programma. Dit maakt het in eerste instantie voor onderzoekers, maar ook voor beleidsmakers en andere gebruikers, mogelijk om eigen analyses van de data te maken of het dataverwerkingsproces na te gaan op fouten. Tot slot zijn we ook nagegaan in welke mate de producent in de samenstelling van haar internationale rangschikking van publieke sectoren oog heeft gehad voor haar cruciale en -vanuit bestuurskundig oogpunt- primaire functie van het creëren van leereffecten. Overheden die een slechte positie op de governance ranking(s) bekleden, zouden aan de hand van de ranking moeten kunnen onderscheiden wat geleid heeft tot de relatief slechtere score. Bovendien zouden de indicatoren dusdanig geconstrueerd moeten zijn dat er ook uit af te leiden is hoe elk land concreet vordering kan maken in de rangschikking.

We hebben op basis van de beschikbare informatie verder nog een inschatting gedaan van de duurzaamheid/robuustheid van de internationale rangschikking van publieke sectoren. Hierin hebben we aangeven hoe de wegingsfactoren voor de indices werden samengesteld. Gerelateerd daarmee is de vraag of de wegingsfactoren en het meetmodel standvastig zijn gebleven door de jaren heen. In veel gevallen zijn er echter grote methodologische wijzigingen doorgevoerd in een zoektocht naar een verbeterde meting van publieke sectoren. Dat dit de comparabiliteit over de jaren heen niet bevorderd heeft, hoeft geen betoog. We hebben voorts ook

de berekeningsmethode binnen de verschillende rankings geduid. Daarnaast hebben we een overzicht opgemaakt van de evolutie die de landensample heeft doorgemaakt gedurende de laatste jaren. Tot slot hebben we nog getracht zicht te krijgen op de consistentie van de ranking. Hier zijn we enerzijds nagegaan in welke mate de auteurs van de meting een controlesysteem invoerden om binnen een ranking consistente scores voort te brengen. Vooral bij experten-analyses is er gevaar op inconsistentie, aangezien verschillende experten mogelijks ook op een verschillende manier scores toekennen. Dit kan bijvoorbeeld verholpen worden door regionale en mondiale *reviews* in te lasten. Anderzijds hebben we hier trachten in te schatten in welke mate methodologische en samplesamenstellingswijzigingen de consistentie over de jaren heen in het gedrang zouden kunnen gebracht hebben.

Aangezien uit de analyse van de indicatoren die aan de basis van de geselecteerde governance rankings liggen, blijkt dat er in grote mate gebruik wordt gemaakt van subjectieve indicatoren, hebben we onder de titel 'survey-eigenschappen' ook de manier beschreven waarop deze zachte data werden ingevuld. Zo hebben we onder meer de cruciale vraag gesteld naar de aard van de respondent. Als men voor bepaalde metingen terugvalt op surveys, is het niet onbelangrijk aan te geven wie er bevraagd werd. Hier komen we dan ook uit bij de vertekening die gerelateerd is aan de aard van de verkozen respondent. Bij een respondentengroep die exclusief bestaat uit bedrijfsleiders en managers moet men bijvoorbeeld op zijn hoede zijn voor een economische vertekening in de data. Ook de grootte van de samples bij de surveys -het aantal respondenten of experten dat aan de basis ligt van de scores- wordt per rangschikking geëvalueerd. Bovendien zijn we nagegaan of de producent de selectie van een bepaalde respondentengroep verantwoordt. Tot slot hebben we nog onderzocht of er vage, dan wel concrete vragen werden voorgelegd in de surveys.

De analyse van de acht geselecteerde internationale rangschikkingen van publieke sectoren wijst uit dat het de meeste governance rankings aan technisch-statistische kwaliteit ontbreekt. Er wordt in het algemeen te weinig rekening gehouden met meetfouten die in het datagarings- en verwerkingsproces zijn geslopen. Hieruit volgt dat de berekening van betrouwbaarheidsintervallen eerder de uitzondering dan de regel zijn. Deze dienen nochtans aan te geven welke verschillen wel en welke niet significant zijn.

Daarnaast rijzen er nog andere problemen in verband met de comparabiliteit binnen de rankings. Veranderingen in de methodologie en in de landensample zijn schering en inslag, waardoor scores niet over de jaren heen te vergelijken zijn.

Voorts vertoont de analyse ook een groot gebrek aan transparantie en databeschikbaarheid. Hoewel er de laatste jaren een evolutie in positieve zin te merken is op het vlak van het beschikbaar stellen van data en methodologie, is er nog steeds een groot gemis aan verantwoording van de keuzes en theorieën die de modellen onderbouwen.

Bovendien komt ook het gebrek aan regionale inclusie, de beperkte mate van leermogelijkheden en de onderhuidse bias in deze empirische analyse boven drijven. In wat volgt geven we een omschrijving van de componenten van een kwalitatieve ranking, en illustreren we deze ook aan de hand van analyseresultaten uit de meta-fiches.

> *5.3.2.1. Statistische kwaliteit*

Een belangrijk aspect van een kwaliteitsvolle internationale rangschikking van publieke sectoren is de statistische kwaliteit van de data die aan de oorsprong van de rankings ligt. Ten eerste hebben we in dit kader aangegeven hoeveel observaties of datapunten aan de basis van de ranking liggen. Dit is in feite de vraag naar het aantal componenten, indicatoren en factoren dat gemeten wordt per land om de *overall ranking* samen te stellen.

Daarnaast hebben we onderzocht of de producent al dan niet voorbij is gegaan aan de meetfouten die onvermijdelijk met een meting van de overheidskwaliteit samenhangen. De impliciete of expliciete erkenning van de aanwezigheid van meetfouten door de producent, zou zich in de praktijk idealiter moeten vertalen in de berekening en toevoeging van betrouwbaarheidsintervallen. Deze intervallen geven aan welke veranderingen en vergelijkingen statistisch significant zijn en welke net zo goed toegeschreven kunnen worden aan het toeval. Hun aanwezigheid is dus net een indicatie van betrouwbaarheid, niet van onnauwkeurigheid. Een alarmerende vaststelling is dat bijzonder weinig producenten van governance rankings betrouwbaarheidsintervallen voor hun data hebben

berekend. Impliciet geeft de producent zo aan dat elke afwijkende score ook duidt op een effectief verschil in de realiteit, dat er geen meetfouten of vertekeningen in de data zitten. Of dit nu *wishful thinking* van de producent is, onwetendheid betreft of in de marketingstrategie past, het zorgt voor ongenuanceerde, onverantwoorde en vaak zelfs ronduit foutieve vergelijkingen bij gebruikers.

Omdat de gebruiker van governance data volledig op de producent moet vertrouwen bij het interpreteren van de data en statistieken aan de basis van de rankings, wordt vanuit diverse hoeken de idee geopperd om een officieel en onafhankelijk orgaan te betrekken bij de controle op de gebruikte data en opgemaakte statistieken (Bevan & Hood, 2006; Majone, 1996; Thomson, 2005b). Onder het kwaliteitsaspect '(statistische) revisie' zijn we in dit kader tot slot nagegaan of er interne controlemechanismen door de producenten van de rankings werden ingesteld om de kwaliteit van de data en de ranking te garanderen.

De analyse van de acht geselecteerde governance rankings wijst op een gebrekkige statistische kwaliteit. Er wordt al bij al door de producenten weinig gehoor gegeven aan de kritiek op de betrouwbaarheid van de data en rankings.

Het aantal observaties in de governance rankings is een indicator voor de omvang van de meting. De analyse wijst uit dat het aantal datapunten sterk varieert binnen de acht internationale rangschikkingen van publieke sectoren. Belangrijk is wel aan te geven dat landen in een internationale rangschikking van publieke sectoren niet voor elke opgenomen indicator een score dienen te hebben. Het product van het aantal indicatoren met het aantal gecoverde landen is dus in veel gevallen een overschatting van het aantal datapunten. Deze formule heeft echter negatieve effecten op de comparabiliteit tussen de landen onderling. Het zorgt er immers voor dat landen wel tegenover elkaar gezet worden op een rangschikking, maar op basis van een score die op een andere manier is tot stand gekomen. Het voordeel is dan weer dat landen die een score ontbreken om welke reden dan ook, toch kunnen opgenomen worden in de ranking, wat de *country coverage* en representativiteit verhoogt.

De Worldwide Governance Indicators zijn waarschijnlijk de omvangrijkste governance ranking die er vandaag bestaat. Met haar 310 indicatoren voor

212 landen heeft het World Bank Institute een erg brede meting van het governance-gehalte tot stand gebracht. In totaal baseert de ranking zich op net geen 11.000 datapunten.

Dan is bijvoorbeeld de Public Sector Performance een kleinschaligere comparatieve overheidsmeting. Er werden slechts 23 geïndustrialiseerde landen in de oefening van de European Central Bank opgenomen, waarvoor uiteindelijk 15 onderliggende indicatoren werden gemeten.

Op het vlak van de omgang met meetfouten is er nog veel progressie te maken. Zoals eerder gesteld zijn Transparency International en het World Bank Institute de enige producenten die in hun ranking expliciet meetfouten erkennen door betrouwbaarheidsintervallen te berekenen. Het is een verdienste op zich dat deze rankings een betrouwbaarheidsinterval aan de score toevoegen. Ze laten toe een onderscheid te maken tussen de verschillen die op toeval berusten en de verschillen die wijzen op een significant verschillende situatie.

Van de Corruption Perceptions Index wordt in de literatuur gesteld dat ze een grote onbetrouwbaarheid herbergt. De betrouwbaarheidsintervallen zijn met andere woorden relatief groot, de comparabiliteit binnen de ranking daardoor een stuk minder. Een voorbeeld kan dit illustreren. De Kaapverdische Eilanden staan met een score van 4.9 op een 49e plaats. De relatief grote betrouwbaarheidsmarge (3.4-5.5) illustreert echter de onzekerheid van deze positie. De eilandengroep zou in feite net zo goed op de 37e plaats (met een score van 5.5) kunnen staan, of in het slechtste geval terecht komen op een 79e plaats (met een score van 3.4). Aangezien de Corruption Perceptions Index wordt berekend op basis van surveyvragen uit externe databronnen en niet elk land in al die databronnen voorkomt, is het betrouwbaarheidsinterval voor elk land vooral afhankelijk van het aantal databronnen waarin het is opgenomen. Hoe groter dat aantal, hoe groter de geschatte datanauwkeurigheid en hoe kleiner het betrouwbaarheidsinterval.

De betrouwbaarheidsmarges bij de Worldwide Governance Indicators (World Bank Institute) zijn een stuk kleiner. Het laatste rapport (Kauffman et al., 2007a) vermeldt een gemiddelde meetfout van 0.21 voor de genormaliseerde governance scores. Doordat deze governance rankings

gebruik maakt van 310 individuele variabelen uit 33 verschillende databronnen, kunnen de meetfouten enigszins beperkt worden.

> *5.3.2.2. Data-oorsprong*

Een tweede aspect dat we binnen de technische dimensie voor elke governance ranking uit onze selectie zijn nagegaan, is de oorsprong van de data. Sommige governance rankings doen immers gedeeltelijk of zelfs volledig beroep op externe databronnen om een nieuwe ranking te produceren. Daarbij zijn er aan de ene kant een aantal internationale rangschikkingen van publieke sectoren die letterlijk indicatoren en scores overnemen uit andere databronnen. Aan de andere kant zijn er de Expert Assessments, die in strikte zin eigen data produceren, maar daar in feite nadrukkelijk een beroep voor doen op de informatie die andere governance rankings voortbrengen.

De idee achter het compileren van bestaande data, is dat data over eenzelfde fenomeen, gemeten door verschillende onafhankelijke instanties, samen nog beter de realiteit zouden moeten benaderen. Het nadeel van een dergelijke werkwijze is dat er een grote afhankelijkheid bestaat tussen de verschillende governance rankings, wat eveneens betekent dat de gevestigde visies, bestaande data, en, belangrijker nog, de biases versterkt worden. De individuele datasets, en dan vooral die van commerciële *risk ratings agencies*, maken daardoor gecorreleerde *errors* in hun metingen van publieke sectoren en zijn dus minder informatief dan aangenomen (Arndt & Oman, 2006).

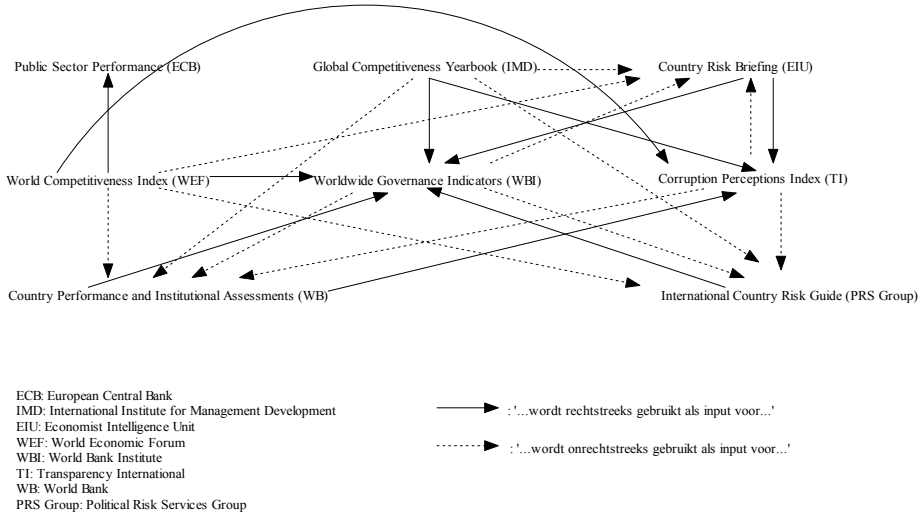
Uit onze analyse komt bijvoorbeeld naar voren dat de Worldwide Governance Indicators voor de samenstelling van hun meting exclusief beroep doen op externe databronnen. Om zoveel mogelijk de vertekeningen van de individuele databronnen tegen te gaan, doet het World Bank Institute een beroep op maar liefst 33 verschillende databronnen. Dit moet een correctere en betrouwbaardere meting van governance opleveren.

Ook Transparency International (in haar Corruption Perceptions Index) en de European Central Bank (in haar Public Sector Performance) doen een beroep op externe bronnen voor de zachte data die in hun respectievelijke rankings worden gebruikt. De World Bank (Country Performance and Institutional Assessments), de PRS Group (Country Risk Briefing) en de Economist

Intelligence Unit stellen hun ranking samen op basis van experten-scores. In de begeleidende papers wordt echter expliciet vermeld dat er daarbij gekeken wordt naar andere internationale metingen van publieke sectoren. Het World Economic Forum (World Competitiveness Index) en het International Institute for Management Development (Global Competitiveness Yearbook) baseren zich elk op een zelf opgestelde *Executive Opinion Survey* onder bedrijfsleiders.

In onderstaande figuur geven we een overzicht van de interdependentie bij de governance rankings uit onze analyse. Enerzijds zijn er een aantal rankings die (gedeeltelijk) beroep doen op de indicatoren van andere rankings om een eigen internationale rangschikking van publieke sectoren op te bouwen. Anderzijds zijn er de *commercial risk rating agencies* die zich bij hun evaluaties nadrukkelijk laten leiden door enkele bestaande gerenommeerde governance rankings. Het spinnenweb dat deze oefening oplevert geeft aan dat er in grote mate beroep gedaan wordt op bestaande data uit aanverwante databronnen om nieuwe rankings op te bouwen (zie figuur 4). Dat sommige producenten van internationale rangschikkingen van publieke sectoren de validiteit en betrouwbaarheid van hun ranking illustreren aan de hand van de hoge correlatie van de eigen scores met die uit andere governance rankings, is dan ook veel te gemakkelijk.

Figuur 4: Interdependentie governance rankings



> *5.3.2.3. Analyse-eenheid*

Vervolgens hebben we een blik gericht op de analyse-eenheid. Vanuit het Vlaamse perspectief is dit een zeer belangrijk aspect van de internationale rangschikkingen van publieke sectoren. Vlaanderen heeft, als regio met een zekere verworven autonomie, op vlak van benchmarking dezelfde noden als nationale overheden. Het Vlaamse beleidsniveau streeft er immers evenzeer naar om zich internationaal te positioneren en de eigen overheidsprestaties af te wegen ten opzichte van andere regio's of naties.

Het is vanuit dat standpunt echter ontnuchterend vast te stellen dat er in al de rankings aan regio's wordt voorbijgegaan. Op enkele kortstondige en uiterst beperkte testcases na worden in de verschillende overheidsrangschikkingen uit onze analyse enkel en alleen landen opgenomen. Hiervoor worden door de producenten van internationale rangschikkingen diverse redenen naar voor geschoven. Ze vinden het vaak een te arbeidsintensief proces om vergelijkbare data voor regio's te vinden. Ook de commerciële balans voor regionale opname zou negatief uitvallen, zo lieten enkele producenten van governance rankings ons tijdens gesprekken weten. Bovendien zouden zich ook enkele nieuwe moeilijke keuzes opdringen: Vergelijking met welke regio's? Wat te doen met de vele ontbrekende (harde) data op regionaal niveau?

Ook de grootte van de landensample hebben we in dit onderdeel meegenomen. De *country coverage* doet immers uitspraak over de geografische omvang van de meting. Sommige governance rankings zijn eerder beperkt in regionale spreiding, terwijl andere rankings er een expliciete doelstelling van hebben gemaakt zoveel mogelijk landen in hun meting op te nemen.

Een kritiek die een aantal governance rankings te beurt valt, is dat ze zich in hun ranking op een te selectief deel van de wereld richten. Sommige rankings focussen zich bewust of onbewust op bepaalde regio's van de wereld en laten daarbij andere 'wereld-delen' buiten beschouwing. Andere internationale rangschikkingen houden dan weer geen rekening met regionale spreiding, maar beperken zich wel tot een beperkt landenaantal. Nog andere governance rankings hebben misschien in omvang wel een grote landensample, maar vertegenwoordigt deze sample slechts een beperkt

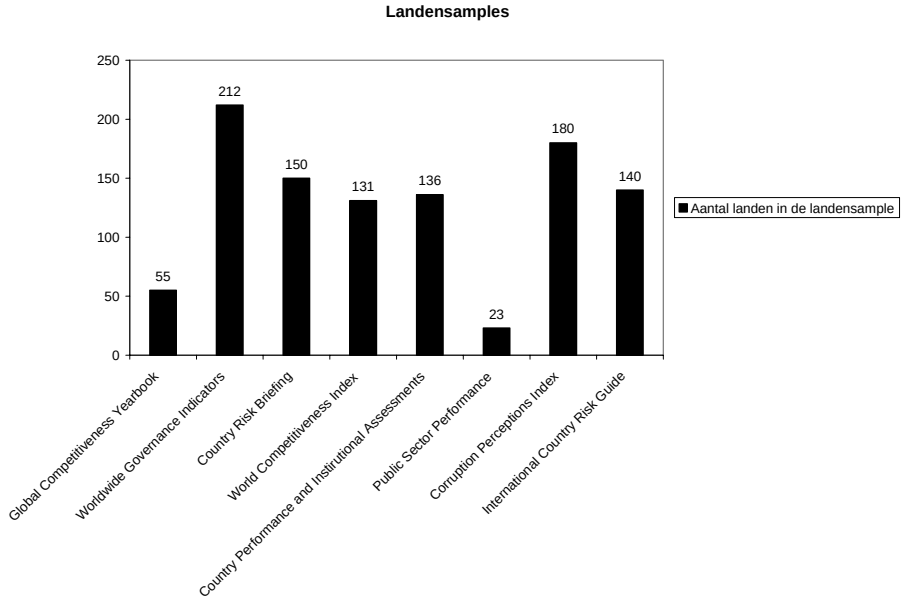
aandeel van de wereldbevolking (inwoners) of van de wereldrijkdom (mondiale GDP).

De analyseresultaten van de acht internationale rangschikkingen van publieke sectoren illustreren deze diversiteit. Enkele governance rankings zijn eerder beperkt in omvang, zoals de Public Sector Performance (European Central Bank) met haar 23 (Westerse) landen. Ook het International Institute for Management Development opteert er in haar World Competitiveness Yearbook voor minder landen in de sample op te nemen. Deze ranking maakt in grote mate gebruik van harde data, die voor slechts een beperkt aantal landen beschikbaar en betrouwbaar zijn. Enkel landen waarvoor voldoende officiële statistieken voor handen zijn, kunnen daardoor deel uitmaken van de ranking.

Aan de andere kant staan de internationale rangschikkingen van publieke sectoren die een omvangrijke landensample vertonen. Zo nemen de Worldwide Governance Indicators 212 landen op in de ranking en ook de Corruption Perceptions Index komt dicht bij een *worldwide coverage*. De Global Competitiveness Index van het World Economic Forum wordt berekend voor 131 landen, maar vertegenwoordigt daarmee slechts 2% van het World GDP.

Onderstaande figuur (zie figuur 5) geeft een overzicht van het aantal landen dat in elk van de onderzochte governance rankings is opgenomen. Zoals beschreven bestaat er een grote diversiteit op het vlak van de grootte van de landensamples. De meeste rangschikkingen van publieke sectoren trachten echter zo veel mogelijkheid overheden in de meting te betrekken.

Figuur 5: Landensamples



België is in zeven van de acht governance rankings opgenomen. Enkel de Country Performance and Institutional Assessments van de World Bank hebben ons land niet in de landensample zitten. Deze ranking is echter in het leven geroepen om de financiële steun aan ontwikkelingslanden mee te verdelen. De sample beperkt zich daardoor bijna uitsluitend tot de Derde Wereld.

Wat de regionale component betreft, wordt in onze analyse het grote gebrek aan regionale inclusie in de governance rankings geïllustreerd. Geen enkele van de acht internationale rangschikkingen van publieke sectoren heeft in haar recentste versie regio's opgenomen.

Enkel het International Institute for Management Development heeft in het verleden geëxperimenteerd met het opnemen van regio's in de ranking. In 2003, 2004 en 2005 werden verschillende regio's opgenomen in het World Competitiveness Yearbook. Het ging om Zhejiang, Bavaria, Catatonie, Ile-De-France, Schotland, Rhon-Alps, Maharashtra, Lambardije en Sao Paolo. Vlaanderen hoorde niet bij de uitverkoren regio's.

Doordat de auteurs het proces van regionale datagaring te energie- en tijdrovend ervoeren, werd aan dit experiment geen gevolg gegeven. In interviews met analisten van het International Institute for Management Development blijkt dat er te weinig capaciteit was binnen de organisatie om naast de nationale ook regionale data te verzamelen. Bovendien merkten ze op dat de kosten niet opwogen tegen de (financiële) baten en de interesse.

Als gevolg daarvan heeft deze organisatie in 2007 beslist om geen regio's meer in haar ranking op te nemen. Regionale studies worden wel nog uitgevoerd, maar op individuele basis en op kosten van de opdrachtgever. Deze regionale datagaring staat wel steeds los van de gekende algemene ranking. Ile-De-France en enkele regio's in Maleisië hebben in het korte verleden dergelijke regionale studies laten opmaken. Momenteel gebeurt dit voor Galicië. Vlaanderen is in onderhandeling geweest met de producent achter het World Competitiveness Yearbook, maar dit heeft onder andere wegens de hoge vraagprijs verder geen gevolg gekregen.

Zoals eerder vermeld is er een initiatief geweest binnen het Steunpunt Innovatie om de Belgische data in het Global Competitiveness Index (World Economic Forum) te herberekenen voor Vlaanderen (Bowen et al., 2007; De Witte & Moesen, 2007). Dit was echter een eenmalig initiatief en stuitte op de grenzen van de regionale datagaring. Governance data op het niveau van Vlaanderen zijn op dit moment niet in grote mate verzameld, zeker niet op vlak van de overheidskwaliteit. Bovendien geven de onderzoekers aan dat het monnikenwerk was om de hele herberekening voor Vlaanderen te doen. Als je weet dat het aantal indicatoren van de Global Competitiveness Index sindsdien gestegen is van 36 tot 116, is de terughoudendheid van onderzoekers en beleidsmakers om continuïteit op dit vlak te verzorgen snel verklaard.

> *5.3.2.4. Timeframe*

Het laatste aspect onder de statistische kwaliteit dat we zijn nagegaan, betreft het tijdskader, het *timeframe*. Onder deze component schuilt onder meer de vraag sinds wanneer de meting wordt uitgevoerd. Hoe ver de data teruggaan, geeft aan in welke mate er tijdsreeksen kunnen opgesteld worden voor de onderzochte landen. Dergelijke tijdsreeksen vormen theoretisch gezien het ideale instrument om evoluties waar te nemen.

Ook de periodiciteit hebben we in dit onderdeel getracht te duiden. De meeste rankings worden op jaarlijkse basis geproduceerd, maar er zijn ook internationale rangschikkingen van publieke sectoren die slechts om de drie jaar verschijnen. Aan de andere kant van het spectrum zijn er dan weer governance rankings -meestal *risk assessments*- die bijvoorbeeld op dagelijkse basis hun data bijwerken opdat de meting als een *early warning system* voor investeerders kan dienen.

Ten slotte hebben we beschreven hoe groot de *time lag* is tussen de vaststelling van de meting en de publicatie van de data. Hoe groter deze kloof, hoe groter ook het risico dat de cijfers reeds achterhaald werden door de realiteit. Ook voor beleidsmakers is dit belangrijk. Het geeft hen aan hoe snel bepaalde beleidsmaatregelen of hervormingen zich kunnen manifesteren in een wijziging in de *governance scores* en verschuivingen in de ranking kunnen bewerkstelligen.

De analyse laat zien dat de Country Performance and Institutional Assessments de oudste governance ranking vormen uit de selectie. Deze ranking wordt reeds sinds 1977 opgemaakt en wordt ingezet bij de verdeling van de ontwikkelingshulp aan Derde Wereldlanden. De meeste andere governance rankings zijn ontstaan tussen eind jaren '80 en midden jaren '90.

De meeste rankings voorzien in een jaarlijkse update van de data. De producenten van risico-inschattingen, die hun data meestal via een website communiceren, trachten op maandelijks (PRS Group) of soms zelfs dagelijkse basis (Economist Intelligence Unit) de data te verversen.

De Public Sector Performance van de European Central Bank vormt een uitzondering op dit vlak. Het ging immers om een eenmalige oefening uit 2003 die in een score voorzag voor het jaar 1990 en 2000. Op deze manier kon een prestatievergelijking opgemaakt worden aan de start van het nieuwe millennium. Bovendien moest de ranking helpen om de landen te identificeren die progressie maakten ten opzichte van de situatie een decennium voorheen.

> *5.3.2.5. Databeschikbaarheid*

Een andere (technische) eigenschap van een kwaliteitsvolle internationale rangschikking van publieke sectoren is een goede databeschikbaarheid. De data met betrekking tot de kwaliteit van het overheidsapparaat moeten er minstens ten dele op gericht zijn beleidsmakers bij te staan in hun constante ijver naar een beter presterende overheid.

Vanuit ons oogpunt is het immers een voorname functie voor de internationale rangschikkingen van publieke sectoren om als beleidsinformatie te fungeren in hervormingsdebatten. Deze functie kan echter pas benut worden indien de data op een toegankelijke manier aan de gebruikers ervan gepresenteerd worden. Zowel op vormelijk vlak, als op inhoudelijk en praktisch vlak moeten de data aan bepaalde voorwaarden voldoen.

Een eerste aspect dat we in verband met de databeschikbaarheid in elk van de acht cases hebben getoetst, is de transparantie en beschikbaarheid van data en indicatoren. In eerste instantie lijkt het ons van cruciaal belang, dat overheden immers weten op basis van welke indicatoren ze zullen beoordeeld worden. Niet alleen de weergave van deze variabelen is daarbij van belang, ook de scoringscriteria zouden moeten bekendgemaakt worden.

Vervolgens zou er ook klaarheid omtrent de methodologie moeten zijn, waarbij in ideale omstandigheden een theoretisch onderbouwing aan te pas komt. Gormley en Weimer (1999) stellen in hun onderzoek naar de kwaliteit van *Organizational Report Cards* dat de manier waarop ze gepresenteerd worden en een goede uitleg verschaffen ('*accessibility*'), een belangrijke bijdrage levert aan de kwaliteit van deze instrumenten. Deze stelling geldt ook voor internationale rangschikkingen van publieke sectoren.

In laatste instantie zouden ook de data, onderliggend aan de ranking, op toegankelijke wijze beschikbaar gemaakt moeten worden van de gebruiker. Producenten die een aanzienlijke prijs plakken op datatoegang, verhinderen dit echter. Ook dit vormt een aspect van de transparantie van de overheidsmetingen.

Het tweede onderzocht aspect sluit rechtstreeks bij dit laatste aan. We zijn immers voor elke governance ranking de toegangsvoorwaarden nagegaan. De

meeste rankings zijn van commerciële aard en worden als product op de markt aangeboden. Niet zelden worden hoge toegangsprijzen vastgesteld om zicht te krijgen op de individuele datascores uit de meting. De governance rankings die hun *overall ranking*, de basisscores en de methodologie kosteloos ter beschikking stellen van alle gebruikers zijn eerder een uitzondering dan de regel.

Een derde aspect dat we zijn nagegaan, is van praktische aard. We geven mee of de data aflaadbaar zijn. Indien de data in een *format* beschikbaar zijn waarin ze rechtstreeks als input in een dataverwerkend programma kunnen ingevoerd worden, verhoogt dit de bruikbaarheid voor onderzoekers en beleidsmakers. Dit vergroot immers de mogelijkheid om eigen analyses te maken met de onderzoeksdata en laat toe om de statistieken uit de rankings te *dubble checken*.

Naast de vormelijke dimensie van beschikbaarheid, is ook de inhoudelijke of praktische dimensie van beschikbaarheid relevant voor ons onderzoek naar de kwaliteit van governance rankings. De indicatoren, de methodologie en de data mogen dan nog vormelijk zo mooi voor de dag komen, als de indicatoren geen oog hebben voor leereffecten, zijn ze van weinig beleidsmatige relevantie. Daarom hebben we de verschillende rankings geëvalueerd op hun intrinsieke 'leermogelijkheid'. Om nogmaals de parallel te trekken met het onderzoek van Gormley en Weimer (1999) naar de kwaliteit van *Organizational Report Cards*, spreken deze auteurs van een vereiste van '*functionality*'. De functionaliteit of relevantie wordt vertaald als de vraag: 'Is de *report card* ontworpen op een manier die de gerankte entiteiten aanzet tot verbetering [...]?'

In onze analyse vertalen we dit door na te gaan in welke mate de governance ranking en de onderliggende indicatoren beleidsmakers en hervormers kunnen bijstaan bij het identificeren van kwaliteitsmankementen in de overheid. Niet alleen de vaststelling van deze gebreken zou echter in een goede ranking vervat moeten zitten, er moet ook een aanknopingspunt zijn voor verbetering (Kuha & Gaskell, 2005). Te veel governance rankings doen via hun rankings aan *'naming and shaming'*, maar laten overheden met een slecht rapport verweesd achter met de vraag hoe ze zich kunnen verbeteren.

Leermogelijkheden worden vergemakkelijkt door transparante data en indicatoren. In eerste instantie moet er immers duidelijkheid zijn over slechte scores en hoe deze tot stand kwamen. Voorts is de '*actionability*' van de indicatoren van belang voor de lessen die overheden uit de data moeten kunnen trekken. '*Actionable indicators*' (ook wel *'action-worthy indicators*' genoemd) zijn zo opgesteld dat overheden er gemakkelijk en rechtstreeks op kunnen inspelen om zo hun score op te krikken en hun nationale situatie te verbeteren (Kauffman & Kraay, 2007b). Subjectieve indicatoren hebben een negatieve impact op de leermogelijkheid. Het is immers moeilijker in te schatten hoe bepaalde maatregelen of koersveranderingen de perceptie van bepaalde respondenten zullen beïnvloeden, dan hun effect op officiële statistieken te voorspellen.

De analyse van onze selectie van acht governance rankings geeft duidelijk aan dat er op het vlak van databeschikbaarheid nog veel werk aan de winkel is. In eerste instantie op vlak van de transparantie. Enerzijds is er het indicatorenniveau, waar werd nagegaan in welke mate het voor de gerangschikte landen en gebruikers duidelijk is op welke basis de scores tot stand komen. De hoofdvraag luidt: Worden de indicatoren uit de meting vrijgegeven? Een *country score* zegt op zichzelf niets als er niet geweten is hoe deze tot stand is gekomen. De meeste producenten zorgen ervoor dat de indicatoren beschikbaar zijn voor de gebruikers, al gebeurt dit niet altijd gratis.

Op het vlak van de databeschikbaarheid doen de geanalyseerde governance rankings het een stuk minder goed. Een aantal rankings geeft slechts een beperkt aandeel van de onderliggende datascores vrij (cf. Worldwide Governance Indicators) of geeft de basisscores enkel in genormaliseerde vorm weer. In meerdere gevallen worden de basisdata enkel tegen betaling vrijgegeven. Door het feit dat er meer informatie in de basisdata vervat zit dan in de uiteindelijke uni-dimensionale ranking, heeft elke beperking op de raadpleging van deze data een negatieve invloed op de transparantie van de meting.

De transparantie van de rangschikkingen van publieke sectoren is al langer een probleem. Er werd in het verleden dan ook meermaals druk uitgeoefend vanuit de hoek van de ontwikkelingsorganisaties, onderzoekers en *practicioners* om transparantere metingen te verkrijgen. Overheden die een slecht rapport krijgen, moeten op zijn minst meekrijgen voor welke

concrete indicatoren ze achterop hinken. Organisaties zoals de World Bank in haar Country Policy and Institutional Assessments en het World Bank Institute in de bekende Worldwide Governance Indicators hebben deze kritiek deels ter harte genomen en hebben de laatste jaren vooruitgang geboekt op het vlak van de transparantie.

Er is de transparantie op het vlak van de indicatorenbeschikbaarheid, op het vlak van de (basis)data, maar daarnaast is er ook de transparantie op methodologisch vlak. Op dit vlak schieten alle governance rankings uit de analyse tekort. Het raamwerk van indicatoren die onder de hoofdindicator liggen, gecombineerd met de wegingsfactoren komen haast altijd als het ware 'uit de lucht gevallen'. Er is geen theoretisch kader of welomschreven analyse voorzien dat verantwoording biedt voor de geselecteerde onderliggende databronnen, dimensies en indicatoren. Ook het wegingsschema, dat impliciet aangeeft welke aspecten belangrijker geacht worden dan andere, wordt haast nooit onderbouwd. Dit is een groot gebrek. Landen worden op het vlak van de overheidswerking op een bepaalde manier gemeten, maar de overheden en de gebruikers krijgen geen uitleg over de keuzes binnen deze meting.

De Public Sector Performance, de Worldwide Governance Indicators, de Country Performance and Institutional Assessments en de Corruption Perceptions Index bieden hun basisdata gratis online aan. Zowel de rangschikking als de individuele datascores per land is op die manier vrij toegankelijk.

De andere vier internationale rangschikkingen van publieke sectoren worden geproduceerd door commerciële instellingen. De *commercial risk rating agencies* de Economist Intelligence Unit en de Political Risk Services Group leveren de duurste rankings af. De eerste biedt toegang tot haar Country Risk Briefing aan tegen de prijs van £5.500, de laatste vraagt voor haar International Country Risk Guide $8.000 per jaar om de basisscores te raadplegen.

Een kritische analyse wijst uit dat de leermogelijkheden, de potentie voor leereffecten, in de meeste governance rankings uiterst beperkt zijn. Ten eerste vormt de beperkte transparantie en databeschikbaarheid die we zonet besproken en geïllustreerd hebben een obstakel voor het creëren van leereffecten.

Ten tweede belemmert ook het gebrek aan comparabiliteit over de jaren heen de leermogelijkheden. Het is immers vaak onmogelijk uit de governance rankings af te leiden of een land er in absolute termen op is vooruitgegaan of niet. De data doen in de meeste gevallen enkel een uitspraak over de vraag of een land het er in vergelijking met de andere landen beter van af heeft gebracht of niet. Bovendien zijn deze onderlinge vergelijkingen ook niet altijd betrouwbaar.

Ten derde worden er in grote mate subjectieve indicatoren gebruikt, waarvan de beïnvloeding door de overheid moeilijker in te schatten is. De perceptie van de bevolking, bedrijfsleiders of andere deelgroepen loopt niet altijd parallel met de reële omstandigheden. Zo is er in veel gevallen een grote *time lag* tussen het nemen van beleidsmaatregelen en de impact die deze hebben op de perceptie van bevoorrechte getuigen. Bij experten-analyses is er het nadeel dat de gebruiker vaak niet weet op welke basis de selecte groep van experten tot een bepaalde evaluatie en score is gekomen.

De Political Risk Services Group tracht in te spelen op dit probleem en geeft de informatie en de data vrij, op basis waarvan de experten de *risk scores* toekennen. Ook de situatieanalyses die de experten tot een bepaalde perceptie hebben aangezet, worden openbaar gemaakt. Een negatieve illustratie vinden we dan weer terug bij de Country Risk Briefing van de Economist Intelligence Unit. Experten geven in deze ranking een beoordeling van de risicosituatie in de verschillende landen, die dan uiteindelijk de basis vormt voor de governance ranking. Er is echter bijzonder weinig informatie beschikbaar over het waarom van de scores. Bovendien is het merendeel van de *raters in-house expert*, wat er op neerkomt dat ze 'vanuit hun bureau' een perceptie moeten vormen in plaats van vanuit het land zelf de situatie in te schatten.

Een vierde factor die de leereffecten bemoeilijkt, is het grote gebrek aan zogenaamde *actionable indicators* in de verschillende metingen. Dit soort indicatoren richt zich specifiek op de meting van aspecten die onder de controle van de beleidsmakers vallen. De overheid zou er met andere woorden direct op moeten kunnen inspelen. Op die manier heeft de overheid de sleutel zelf in handen om negatieve scores op te krikken. De Country Policy and Institutional Assessments (World Bank), die ingezet worden bij de allocatie van ontwikkelingshulp, is in grote mate opgebouwd uit dergelijke *actionable indicators*. Deze meting moet immers niet allen

helpen de financiële steun aan Derde Wereldlanden te verdelen, ze moet ook een aanknopingspunt bieden voor deze landen over hoe het ontwikkelingsproces (en dus de score) verder te verbeteren.

In ieder geval is het belangrijk aan te geven dat de basisdata de overheden interessante en relevante informatie over hun prestaties kunnen bieden. De media staren zich echter te vaak blind op de ranking *an sich*. Deze bundelt alle informatie, maar heeft op zich enkel een indicatieve waarde. De positieve en negatieve punten van een publieke sector zijn enkel uit de basisdata af te leiden.

> *5.3.2.6. Duurzaamheid*

Een volgende component onder de technische dimensie betreft de duurzaamheid van de ranking. De duurzaamheid, standvastigheid of robuustheid van een internationale rangschikking van publieke sectoren hangt samen met diens consistentie van de meting over de jaren heen. Er zijn verschillende aspecten die we onder deze noemer zijn nagegaan.

Ten eerste hebben we de wegingsfactoren in kaart gebracht. De robuustheid van het model is in grote mate afhankelijk van de gekozen wegingsfactoren. Sommige modellen zijn sterk afhankelijk van het geselecteerde wegingsmodel en zullen totaal andere resultaten voortbrengen bij een ander wegingsschema. Andere rankings zijn dan weer minder gevoelig voor de wegingsfactoren en zouden tot ongeveer gelijkaardige resultaten komen bij een differentiële weging. De manier waarop de weging van de verschillende componenten en indicatoren is vastgesteld en de wijzigingen die dit schema heeft doorgemaakt, hebben we onder de noemer 'wegingsfactoren' trachten samen te vatten.

Vervolgens hebben we ook de dynamiek van de samplegrootte onder de loep genomen. De evolutie van de '*country coverage*' werpt een licht op de standvastigheid van de landensample. Als er grote verschillen over de jaren heen waar te nemen zijn, zal het moeilijk zijn om vergelijkingen te maken ten opzichte van voorgaande jaren. Scorewijzigingen kunnen in dergelijke gevallen immers net zo goed het gevolg zijn van de toevoeging of schrapping van landen in de sample dan dat ze wijzen op een effectieve verandering.

Onder het kwaliteitscriterium 'model', hebben we aangegeven of het meetmodel, de manier waarop de totaalscores tot stand komen, door de jaren heen veel wijzigingen heeft doorstaan. Hier loert hetzelfde gevaar als bij de samplegrootte. Als het model grote veranderingen heeft doorgemaakt doordat er nieuwe dimensies, componenten of indicatoren zijn toegevoegd of dat er een aanpassing aan het wegingsschema is doorgevoerd, zullen vergelijkingen ten opzichte van voorgaande jaren nutteloos zijn. Doordat er in een dergelijk geval in feite iets nieuws gemeten wordt, is er geen betekenisvolle vergelijking meer mogelijk met scores uit het verleden.

Onder de noemer 'berekeningsmethode' hebben we getracht het scoringsmodel kort samen te vatten. Het is immers belangrijk de weg te kennen, die werd afgelegd tussen het vaststellen van bepaalde onderliggende fenomenen en de berekening van de totaalscore die verbonden is met de opmaak van de ranking. Hier hebben we dus onderzocht hoe bepaalde variabelen worden gemeten, hoe subjectieve indicatoren worden ingevuld en hoe de onderliggende scores worden geaggregeerd.

Onder het criterium 'consistentie' zijn we nagegaan hoe de consistentie binnen de ranking wordt gegarandeerd. Vooral bij subjectieve evaluaties door expertenpanels, is het van belang de verschillende scores op elkaar af te stemmen. Ook de consistentie over jaren heen, potentieel aangetast door veranderingen in de methodologie of landensamenstelling, wordt hier geëvalueerd.

De analyse van de acht governance rankings, die allen werken met indices of geaggregeerde indicatoren wijst uit dat er weinig aandacht besteed wordt aan de modaliteit en verantwoording van de wegingsmethode. De meeste rankings hanteren een *fixed* wegingsschema, waarbij een indicator voor elk land met eenzelfde gewicht wordt meegerekend in de totaalscore. Dit is de meest simpele aggregatiemethode, maar daarom niet de objectiefste. In de meeste gevallen trekt men een gemiddelde van de onderliggende componenten en/of indicatoren om tot de index te komen.

Het World Bank Institute opteert voor een vaste weging, maar stapt af van de gemiddeldetrekking. In haar Worldwide Governance Indicators gebruikt ze het *'Unobserverd Components Model'* waarbij de weging van de onderliggende databronnen (en indicatoren) afhangt van de 'precisie' van de

data. Hoe hoger de correlatie van de databron met de gemiddelde governance datascores, hoe groter het gewicht. De Worldwide Governance Indicators vormen hiermee ook de uitzondering op de regel dat de gewichten zelden tot nooit theoretisch onderbouwd worden.

Dit is immers een van de grote kwaliteitsgebreken die uit de analyse naar voor zijn gekomen. De keuze voor bepaalde wegingsfactoren, die een erg belangrijke rol in de aggregatie van de scores en de verhoudingen binnen de rankings spelen, worden haast nooit beargumenteerd. De producenten van governance rankings maken politieke en normatieve keuzes met de vastlegging van de wegingsfactoren, maar onderbouwen hun gronden niet.

Het World Economic Forum is de enige die in haar ranking een variabele weging gebruikt. Bepaalde indicatoren wegen meer door dan andere in bepaalde landen, afhankelijk van de ontwikkelingsfase waarin het land zich bevindt (uitgedrukt in GDP/capita). Voor landen in ontwikkelingsfase liggen de prioriteiten anders dan bijvoorbeeld in geïndustrialiseerde landen. Dit heeft het voordeel dat er tegemoet wordt gekomen aan de kritiek dat landen niet allemaal op dezelfde aspecten kunnen beoordeeld worden. Het nadeel is echter het comparabiliteitsverlies. Het wordt moeilijker en minder betrouwbaar om landen met elkaar te vergelijken op basis van scores die op een andere manier tot stand zijn gekomen.

Wat de samplesamenstelling betreft is het een algemene vaststelling dat de rankings doorgaans trachten zoveel mogelijk landen in de ranking op te nemen. De meeste rankings hebben de laatste tien jaar dan ook een lichte of forsere stijging gekend van het aantal landen in de sample. Zo is Transparency International er de laatste acht jaar in geslaagd tot een verdubbeling van het aantal opgenomen landen te komen.

Enkel de Public Sector Performance, die vooral in Europa veel aanzien genoot, is bij haar initiële 23 landen gebleven. Het ging dan ook om een eenmalige oefening, niet bedoeld ter continuatie.

De consistentie in veel governance rankings komt in het gedrang door de constante toevoeging van landen en de vele methodologische wijzigingen die over de jaren heen werden doorgevoerd. De experten-analyses trachten de interne consistentie van hun rankings nog wel te vrijwaren door regionale en onderwerpsrevisies uit te voeren op de landenscores.

> *5.3.2.7. Survey-eigenschappen*

Zowel objectieve, als subjectieve indicatoren worden ingezet om publieke sectoren te meten. Aangezien bepaalde governance-aspecten moeilijk te vatten zijn in officiële statistieken en zeker op internationaal-vergelijkbare schaal moeilijk vindbaar zijn, worden er vaak surveys gebruikt om dergelijke variabelen te meten.

Aangezien ook binnen de subjectieve indicatoren nog grote diversiteit bestaat, achten we het interessant om dieper in te gaan op de eigenschappen van de surveys die aan de basis van deze zachte data liggen. Ten eerste speelt de aard van de respondent een cruciale rol in de invulling van de subjectieve indicatoren. In grote lijnen zijn er drie categorieën te onderscheiden.

Ten eerste zijn er *expert opinion polls.* Hierbij doet men een beroep op experten, die op basis van de beschikbare informatie, externe databronnen en een correspondentennetwerk een inschatting maken van de situatie in een land. Deze experten zijn soms te situeren in academische, wetenschappelijke of economische kringen, maar komen nog het vaakst uit de staf van de organisatie die de ranking produceert zelf.

Ten tweede zijn er de *household surveys*. In dit soort surveys wordt de bevolking van de verschillende landen aangesproken om hun visie of ervaring met betrekking tot bepaalde aspecten van het openbaar bestuur te uiten.

Ten derde zijn er de *firm surveys* (ook wel *executive opinion surveys* genoemd). Deze veel gebruikte vorm van bevragingen, beperkt zich tot de peiling van de economische elite. Bedrijfsleiders, managers en kaderleden uit de top van het ondernemingsleven worden daarbij ondervraagd over de kwaliteit van de publieke sector. In de praktijk zal deze algemene vraag door de producent van de governance ranking of door de respondent zelf in de praktijk vertaald worden als: "In welke mate bemoeilijkt of vergemakkelijkt de overheid het economische verkeer?"

Dat dergelijke *firm surveys* zo vaak opduiken in internationale rangschikkingen van publieke sectoren, is een direct gevolg van het groot aantal governance rankings met een economische invalshoek. Als de

doelgroep van de ranking voornamelijk bestaat uit ondernemers, investeerders en managers wekt het geen verwondering, dat er vooral gezocht wordt naar de perceptie van de economische wereld ten aanzien van de publieke sector.

De aard van de respondent is ook belangrijk voor het identificeren van een potentiële bias in de meting. De vertekening in de antwoorden zal immers deels afhangen van de visie, ingesteldheid en achtergrondkenmerken van de respondenten. Indien de respondenten geen representatieve staal van de bevolking vormen, is het risico op bias sterk vergroot. Bij bedrijfssurveys is er dus op zijn minst een risico op economische of zogenaamde *private sector bias*. Daarenboven zijn in deze samples steevast mannen oververtegenwoordigd en is de gemiddelde respondent ook een stuk rijker.

Maar ook bij expertenpolls bestaat er een gevaar op vertekende antwoorden. Niet alleen zijn deze evaluaties vaak gebaseerd op een relatief klein aantal respondenten, wat het risico op bias vergroot, de experten komen bovendien ook vaak uit eenzelfde organisatie en delen meer dan waarschijnlijk nog een aantal andere achtergrondkenmerken. Een voordeel is wel dat, in tegenstelling tot de meeste *firm surveys* en *household surveys*, experten vaak een land moeten beoordelen waarmee ze geen rechtstreekse band hebben. Dit neemt het risico op *home country bias* weg. Respondenten die allen in het land dat gequoteerd moet worden, werden gerekruteerd, zullen potentieel een vertekend beeld hebben van hun moeder- of vaderland.

Household surveys waarbij een doorsnee van de bevolking wordt bevraagd met betrekking tot de prestaties van de overheid houden nog het minste risico op vertekening in. Omdat hierbij meestal nauwkeuriger over de representativiteit wordt gewaakt, zou de aggregatie van deze opinies nog het dichtst moeten aanleunen bij de publieke opinie. Dat neemt niet weg dat ook bevolkingssurveys mogelijks het imago van een land weerspiegelen in plaats van de werkelijke prestaties te vatten. Dit is echter een algemeen probleem waar subjectieve indicatoren mee worstelen.

Aangezien de keuze voor een bepaalde respondentengroep een invloed uitoefent op de data en dus van groot belang is, zijn we ook nagegaan of de verschillende producenten van governance rankings hun keuze voor deze of gene surveymodaliteit hebben verantwoord. Net zoals de conceptuele

keuzes en de methodologische keuzes, zouden ook de keuzes met betrekking tot de respondenten idealiter onderbouwd moeten zijn. De realiteit wijst echter uit dat de producenten maar weinig aandacht hebben voor deze methodologische regel (Van de Walle, 2005).

Vervolgens hebben we uitgezocht hoe groot de samples zijn in de surveys. Het aantal respondenten of experten dat de vragenlijst(en) beantwoordt, is van belang voor de representativiteit en betrouwbaarheid van de data.

Tot slot hebben we ook de *questionnaire* van naderbij bekeken. We zijn nagegaan in welke mate de surveyvragen die aan de respondenten worden voorgelegd, concreet, helder en welomschreven zijn. De relevantie hiervan is dat concreet gestelde vragen die een inschatting doen van een welomschreven fenomeen, correctere resultaten zullen opleveren. Onduidelijke vragen in vage bewoordingen worden door respondenten vaker op een persoonlijke manier ingevuld, wat de meting minder betrouwbaar maakt.

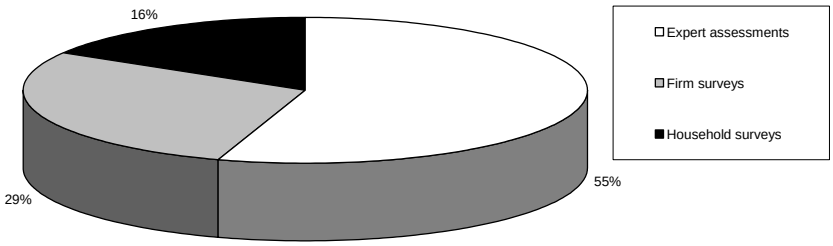
De analyseresultaten uit onze empirische studie tonen aan dat er in grote mate gebruik wordt gemaakt van subjectieve indicatoren om de prestaties van de overheid te meten en een rangschikking van publieke sectoren op te stellen. De meeste zachte data komen uit *expert-assessments* (zie figuur 6). Van de acht governance rankings uit onze selectie, zijn er drie gebaseerd op experten-analyses. Het gaat om de International Country Risk Guide (Political Risk Services Group), de Country Risk Briefing (Economist Intelligence Unit) en de Country Performance and Institutional Assessments (World Bank). Twee andere rankings baseren hun data op surveys waarbij voornamelijk bedrijfsleiders en managers naar hun mening en ervaring gevraagd wordt. Het gaat om het World Competitiveness Yearbook (International Institute for Management Development) en de Global Competitiveness Index (World Economic Forum). De laatste drie rankings, de Worldwide Governance Indicators (World Bank Institute), de Corruption Perceptions Index (Transparency International) en de Public Sector Performance (European Central Bank) doen beroep op externe databronnen voor de samenstelling van hun ranking en combineren zo experten-analyses met survey-resultaten.

Figuur 6: Uitsplitsing subjectieve indicatoren naar survey-type

	Objectieve indicatoren	**Subjectieve indicatoren**	*Expert opinion*	*Firm survey*	*Household survey*
Meting overheids-prestaties					
Government Efficiency (IMD)	**32**	**40**	0	32	0
Government Effectiveness (WBI)	**0**	**49**	27	14	8
Government Effectiveness Risk (EIU)	**0**	**7**	7	0	0
Government Inefficiency (WEF)	**0**	**4**	4	0	0
Quality of Public Administration (WB)	**0**	**4**	4	0	0
Administrative Performance (ECB)	**0**	**4**	1	3	0
Corruption Index (TI)	**0**	**14**	9	5	0
Bureaucratic Quality (PRS Group)	**0**	**1**	1	0	0
Totaal	**32**	**123**	53	54	8

Legende:

- WBI: World Bank Institute
- IMD: International Institute for Management Development
- WEF: World Economic Forum
- TI: Transparency International
- EIU: Economist Intelligence Unit
- WB: World Bank
- PRS Group: Political Risk Services Group
- ECB: European Central Bank



De experten-analyses die in onze selectie zitten, worden alle drie ingevuld door *in-house* experten. Het gaat om stafleden van de organisaties achter de rankings, die zich buiten het onderzochte land bevinden en een beoordeling maken vanuit het basisland van de producent. Qua aantallen, geeft enkel de Economist Intelligence Unit vrij dat de scores door 130 analisten en experten worden samengesteld. Veel critici spreken van een *private sector bias* bij de experten. Ze zouden sterk aanleunen bij de opinie van de economische topklasse. De *raters* van de Country Performance and Institutional Assessments van de World Bank worden in de internationale literatuur ook verweten een sterk neoliberale ingesteldheid te hebben.

De twee competitiviteitsrankings van het World Economic Forum en het International Institute for Management Development baseren zich op *executive opinion surveys.* Bedrijfsleiders, managers en leden van het topkader worden bevraagd over hun ervaringen, meningen en inschattingen met betrekking tot allerhande aspecten van competitiviteit. Het World Economic Forum verzamelde de opinies van 11.406 respondenten uit 131 landen, het International Institute for Management Development baseert zich op 3.700 bedrijfsleiders uit 55 landen. Deze surveys zijn niet

representatief voor de bevolking, aangezien ze vooral uit mannelijke respondenten bestaan, die bovengemiddelde rijkdom bezitten en een *private sector bias* in hun antwoorden leggen. Het World Competitiveness Yearbook (International Institute for Management Development) doet bovendien voornamelijk een beroep op vertegenwoordigers van *multinationals*. De respondentenlijst is immers in grote mate gebaseerd op de alumnilijst van de verwante, gerenommeerde *Business School*.

De rankings opgesteld door het World Bank Institute (Worldwide Governance Indicators), Transparency International (Corruption Perceptions Index) en de European Central Bank (Public Sector Performance) combineren externe databronnen van zowel survey-resultaten als experten-analyses. Ook zij ontsnappen daarbij niet aan de vertekening in de richting van de *business elite*. Het valt daarbij op dat ze alle drie de surveyvragen uit de *Executive Opinion Survey* van het World Economic Forum gebruiken.

> *5.3.3. Politiek-administratieve dimensie*

De derde en laatste dimensie die in ons empirisch onderzoek naar de kwaliteit van internationale rangschikkingen van publieke sectoren aan bod is gekomen, is de politiek-administratieve dimensie. In de internationale literatuur zijn er al enkele kritische studies gepubliceerd die de kwaliteit van internationale rangschikkingen van publieke sectoren onderzochten. De politiek-administratieve dimensie, die vragen beantwoordt over wie er achter de meting zit, welke objectieven de producent voor ogen heeft en voor wie de rankings worden opgesteld, bleef daarbij echter steevast onderbelicht. In deze studie willen we deze elementen mee in de analyse nemen, aangezien de producent en diens doelstellingen cruciale achtergrondinformatie vormt om de draagwijdte van de governance rankings te beoordelen.

Een eerste aspect dat van naderbij werd bekeken, is de aard van de producent. Wie er achter de meting schuilgaat en waar we deze producent moeten situeren, zijn twee belangrijke achtergrondkenmerken van een governance ranking.

Het tweede aspect is de visie die de producent er op nahoudt en de invalshoek van waaruit de internationale rangschikking van publieke sectoren werd opgesteld. Deze invalshoek, eerder economisch,

bestuurskundig of politiek getint, is meestal gemakkelijk in verband te brengen met de aard van de respondent.

Vervolgens hebben we in een derde item een overzicht gegeven van de doelstellingen die de producent van de governance ranking met zijn meting wil bewerkstelligen. Het gaat hier om een weergave van de zelfverklaarde objectieven van de meting, gecombineerd met het reële gebruik door users.

Tot slot hebben we ook de categorie van gebruikers in beeld proberen te brengen. De doelgroep, de users die de producent voor ogen houdt bij de opstelling van de meting en de berekening van de scores, moet voor een goed begrip van de resultaten ook in rekening gebracht worden.

De empirische analyse bevestigt het belang dat de politiek-administratieve dimensie speelt in het correct interpreteren van internationale rangschikkingen van publieke sectoren. De aard van de producent -of het gaat om een NGO, een multilaterale organisatie of een commercieel adviesbureau- speelt mee in de opbouw van de meting en kan in sommige gevallen een algemene bias in de data mee helpen verklaren.

> *5.3.3.1. Producent*

Het eerste aspect van de politiek-administratieve dimensie waar we ons licht hebben over laten schijnen, is de producent. Aangezien we een blik achter de cijfers willen werpen, komen we al snel uit bij de initiator van de meting, de producent. Een goed begrip van de aard van de producent zou moeten meegenomen worden in elke lezing van de ranking en de data.

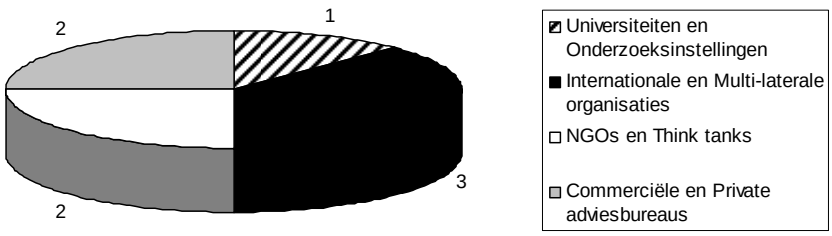
De aard van de producent zal immers tegelijkertijd meer duidelijkheid scheppen over de invalshoek van de ranking, als over de doelgroep, de potentiële biases enzovoort. Een universiteit of onafhankelijke onderzoeksgroep zal een governance ranking vermoedelijk vanuit een theoretisch kader opstellen. Het doel zal erin bestaan verklaringsmodellen van economische groei, ontwikkeling of *good governance* met comparatieve data op punt te stellen. Consultancy bedrijven of andere private instellingen zullen de governance data eerder als commercieel product beschouwen, wat ook gevolgen zal hebben voor de structuur en inhoud van de meting.

Er zijn verschillende categoriseringen van producenten voorgesteld in de internationale literatuur. Arnd en Oman (2006) zien in internationale investeerders, donoren van ontwikkelingshulp en *development*-specialisten de drie belangrijkste bronnen van internationale rangschikkingen van publieke sectoren. De auteurs van de Worldwide Governance Indicators, Kauffman, Kraay en Mastruzzi (2007a) maken een onderscheid tussen vier soorten databronnen, gelinkt aan de aard van de producent. Ten eerste zijn er de *cross-country surveys of firms and individuals*, ten tweede de *expert assessments from commercial risk rating agencies*, ten derde de *expert assessments produced by NGO's and Think Tanks* en ten vierde de *expert assessments from governments and multilaterals.*

Op dit laatste voortbouwend komen we hier tot een onderscheid van vier categorieën. Ten eerste zijn er de universiteiten en onderzoeksinstellingen. Ten tweede zijn er de multilaterale organisaties. Ten derde zijn er de NGO's en Think tanks. En ten vierde zijn er de private en commerciële organisaties. Deze vierdeling vat het spectrum van producenten van governance rankings, al vallen sommige producenten op de grens tussen twee categorieën.

De acht geselecteerde governance rankings zijn vrij evenwichtig over deze categorieën te verdelen (zie figuur 7). Het International Institute for Management Development (World Competitiveness Yearbook) is een organisatie die rechtstreeks gelinkt is aan de gelijknamige Zwitserse *Business School* en behoort dus tot de categorie van onderzoeks- en onderwijsinstellingen. Desondanks behandelen ze hun jaarlijkse governance ranking ook in grote mate als een commercieel product door er een aanzienlijke prijs op te plakken.

Figuur 7: Uitsplitsing (aard) producenten



De European Central Bank (Public Sector Performance), de World Bank (Country Performance and Institutional Assessments) en het World Bank Institute (Worldwide Governance Indicators) vallen onder de tweede categorie van internationale en multilaterale organisaties. Het gaat om rankings die door, voor en met landen worden opgesteld en er vooral moeten toe bijdragen gerangschikte landen tot vooruitgang aan te zetten.

Transparency International (Corruption Perceptions Index) en het World Economic Forum (Global Competitiveness Index) vallen dan weer onder de derde categorie, de groep van NGO's en Think tanks. De eerste voert al meer dan tien jaar strijd tegen de corruptie in de wereld. Het heeft zichzelf als doel gesteld corruptie op de internationale agenda te krijgen en is daar al bijzonder succesvol in geweest. Er zijn immers meerdere concrete voorbeelden aan te halen waarbij de corruptie-ranking landen ertoe aangezet heeft werk te maken van dit probleem (cf. Nigeria, Hong Kong, Zuid-Korea, Singapore). Het World Economic Forum werpt zich ook op als internationale NGO, al vertoont ook deze organisatie enkele kenmerken van een commerciële instelling. Zo worden de data niet gratis ter beschikking gesteld van de gebruikers en zijn ze in grote mate gericht op het adviseren van bedrijfslui.

De Political Risk Services Group en Economist Intelligence Unit ten slotte zijn *commercial business information providers*. Het gaat om private

adviesbureaus die risico-inschattingen doen ten behoeve van internationale investeerders. Ze verzamelen *country intelligence*, data over de politieke, economische en financiële situatie in een land, en bieden die onder verschillende producten aan geïnteresseerde wereldspelers aan. Zoals te verwachten is, besteden deze producenten minder aandacht aan de wetenschappelijke en theoretische onderbouwing van de rankings. Daarenboven hebben ze een overwegend economische inslag en hebben ze te lijden onder een *private sector bias*.

> *5.3.3.2. Visie*

De aard van de producent brengt ook een bepaalde invalshoek van de overheidsmeting en de governance ranking aan de oppervlakte. Het gaat hier over de visie die de producent heeft en die het perspectief bepaalt van waaruit de meting van publieke sectoren wordt opgesteld.

Van de Walle (2005) onderscheidt zeven verschillende invalshoeken van waaruit de meting en rangschikking van publieke sectoren op het vlak van de bestuurskwaliteit kan gestuurd worden. Ten eerste zijn er governance rankings met een politieke insteek. Daarnaast zijn er ook rankings met een bestuurskundige insteek. Ten derde bestaan er governance rankings met een economische invalshoek. Ten vierde identificeert Van de Walle rankings die vanuit de ontwikkelingsleer (door *development scholars*) werden opgesteld. Ten vijfde zijn er de econometrische studies. Ten zesde zijn er de *risk ratings*, met een risico-inschatting als basis. De zevende invalshoek is er een vanuit de *comparative Public Administration research.*

In dit kader situeert Van de Walle ook de misvattingen die er bestaan rondom het gebruik van het Global Competitiveness report (van het World Economic Forum) en het World Competitiveness Yearbook (van het International Institute for Management Development). Deze rankings zijn niet bedoeld om beleidsonderzoek of vergelijkende *Public Administration* te ondersteunen en mogen dus ook niet vanuit die optiek benaderd worden. Ze meten de kwaliteit van het overheidsbestuur als onderdeel van de competitiviteit van een land, wat voor bestuurskundig onderzoek een zeer selectief en onvolledig beeld oplevert (Van de Walle, 2006a).

De invalshoek van het Sociaal en Cultureel Planbureau (Publieke prestaties in perspectief (Sociaal en Cultureel Planbureau, 2007)) en de Organisation

for Economic Co-operation and Development (in het kader van *Government at a Glance*) heeft wel een bestuurskundig inslag (Callens, 2007). Beide organisaties opteren ervoor de overheid op te delen in deelsectoren waarvan nadien dan per deelproces de prestaties op verschillende indicatoren worden gemeten.

De analyse wijst uit dat er grote diversiteit bestaat in de invalshoek die gebruikt wordt door de verschillende governance rankings. Ze hebben elk hun eigen focus en eigen manier waarop ze de kwaliteit van de overheid benaderen. Wat vooral opvalt, is dat de meeste rankings groot belang hechten aan economische indicatoren.

De European Central Bank focust zich in haar Public Sector Performance op twee pijlers. Enerzijds zijn de economische prestaties van de onderzochte landen van belang. Anderzijds peilt de ranking aan de hand van enkele indicatoren ook naar de *quality of life*. De Worldwide Governance Indicators (World Bank Institute) hebben als uitgangspunt het verhogen van de governance. De meting moet het ontwikkelingsproces in de verschillende landen monitoren, en focust zich daarbij in belangrijke mate op de economische performantie. De Corruption Perceptions Index (Transparency International) focust zich volledig op één fenomeen, dat is corruptie.

Het World Economic Forum maakt een rangschikking van publieke sectoren op waarbij de competitiviteit als uitgangspunt wordt genomen. Het International Institute for Management Development deelt met haar World Competitiveness Yearbook deze invalshoek en is dus ook grotendeels vanuit een economische invalshoek op gesteld. Beide rankings houden er een non-interventionistische visie op de overheid op na. De Country Performance and Institutional Assessments van de World Bank moeten de ontwikkelingshulp mee helpen verdelen en leggen daartoe volledig de nadruk op het ontwikkelingsproces; het streven naar armoedereductie, het werken naar economische groei en het efficiënt aanwenden van ontwikkelingshulp in de Derde Wereld. Volgens enkele auteurs komt de neoliberale visie van de World Bank komt hierin sterk tot uiting.

De Political Risk Services Group en de Economist Intelligence Unit tot slot doen een risico-inschatting van een reeks landen op basis waarvan investeerders en bedrijfsleiders, die economische activiteiten met of in deze landen willen uitvoeren, een betere afweging van de risico's kunnen

maken. De eerste producent focust zich daarbij vooral op de politieke en overheidskenmerken, terwijl de laatste er eerder een algemeen economisch en financieel perspectief op nahoudt.

> *5.3.3.3. Doelstelling(en)*

De visie en invalshoek waarover in het vorige punt sprake was, vertaalt zich in concrete doelstellingen die de producent bij de meting en vergelijking van publieke sectoren beoogt. Onder de kwaliteitsvariabele 'doelstelling(en)' zijn we op zoek gegaan naar deze objectieven. Van deze zelfverklaarde doelstellingen, hebben we ook uitgezocht of ze stroken met het reëel gebruikspatroon.

Arndt (2007) onderscheidt in haar onderzoek over het waarom van internationale rangschikkingen van publieke sectoren een viertal doelstellingen. Ten eerste, en dit is ook vanuit ons (bestuurskundig) oogpunt de belangrijkste drijfveer, creëren de governance rankings een incentief voor landen om de prestaties van de eigen overheid te verbeteren. Hierachter gaan bepaalde mechanismen schuil.

Een ervan is het *naming and shaming*-principe. De idee achter dit principe is dat door het bekendmaken van goed en slecht presterende overheden (*naming*) in een weinig verhullende rangschikking, vooral deze laatste zullen aangezet worden om beter te doen om de 'schande' van de negatieve score (*shaming*) weg te werken.

Een verwant mechanisme is het *horse racing*-principe, dat zich dan weer voornamelijk aan de top van het klassement afspeelt. Hier speelt de competitieve druk die landen ertoe aanzet beter te presteren in hun strijd de *horse race* te winnen. De voorwaarde voor deze principes is dat de governance indicators, de gemeten aspecten van een goed overheidsbestuur, weldegelijk in een rangschikking worden gegoten. De visibiliteit van de relatieve posities naar de buitenwereld toe, is trouwens belangrijk voor deze beide mechanismen.

Ten tweede gebruiken internationale hulporganisaties en andere donoren governance scores en rankings om de transparantie en objectiviteit van *aid allocation decisions* mee te sturen. Een aantal overheden, financieringsinstellingen en donororganisaties beroepen zich immers op

governance data om de verdeling van ontwikkelingshulp aan de Derde Wereld mee te sturen of te verdedigen. Arndt (2007) merkt echter fijntjes op dat paradoxaal genoeg net deze internationale rangschikkingen van publieke sectoren vaak tekort schieten op het vlak van transparantie.

Een derde drijfveer voor het meten en vergelijken van publieke sectoren op internationale schaal ligt in het onderzoek. De data in verband met de overheidskwaliteit wordt immers in grote mate als input gebruikt voor kwantitatief onderzoek over governance, ontwikkeling en economische groei. Tot de belangrijkste historische werken in dit kader horen: *'Aid, policies and growth*' (Burnside & Dollar, 1997), *'Institutions and economic performance*' (Knack & Keefer, 1997) en *'Growth without governance*' (Kauffman & Kraay, 2002).

De vierde drijfveer die internationale organisaties ertoe aanzet om governance rankings te publiceren, zijn de zogenaamde institutionele motieven. Deze doelstelling is minder nobel dan de voorgaande drie, maar pertinent aanwezig bij governance rankings. Internationale rangschikkingen van publieke sectoren geven organisaties en instellingen een profiel, een gezicht. Bovendien is er een zeker kuddegedrag onder internationale organisaties vast te stellen: als een voorname instelling iets doet, zullen anderen volgen onafhankelijk van het nut en de efficiëntie van deze overweging. Er kan bovendien niet naast gekeken worden dat internationale rangschikkingen van publieke sectoren *trendy* zijn en veel belangstelling genieten bij pers en media. Tot slot wijst Arndt er nog op dat de diensten die binnen de internationale organisaties voor de rankings instaan een zeker aanzien genieten. Het conservatisme van instituten zorgt er dan ook voor dat deze trend wordt voortgezet, los van de praktijkrelevantie.

Aan dit lijstje wensen we nog enkele doelstellingen toe te voegen. Zo kunnen de internationale rangschikkingen van publieke sectoren die investeerders en bedrijven moeten adviseren bij hun economische en financiële activiteiten, gezien worden als *market information* in een globaliserende markt. De internationale bedrijven die zich op de mondiale markt begeven, zijn immers op zoek naar informatie over de overheidskwaliteit, de risicosituatie en het bedrijfsklimaat in bepaalde landen. Deze vinden ze in de governance rankings.

Vervolgens nemen governance rankings niet alleen een externe *accountability*-functie voor hun rekening, zoals bij beslissingen omtrent ontwikkelingshulp. Ze hebben ook een potentiële interne verantwoordingsfunctie. Overheden, ministers en beleidsmakers kunnen op basis van internationale rangschikkingen van publieke sectoren aan de kiezer tonen in welke mate hun beleid positieve effecten heeft gehad in termen van een betere score en positie op de internationale ladder.

Net zoals de analyse een grote diversiteit aan invalshoeken identificeerde in de acht geselecteerde internationale rangschikkingen van publieke sectoren, zijn ook de objectieven van de producenten en het effectief gebruik van de rankings erg uiteenlopend. Het is moeilijk één enkel objectief op elke ranking te plakken, aangezien ze in bepaalde mate voor de meeste functies worden gehanteerd. We beperken ons daardoor hier tot het duiden van enkele hoofddoelstellingen, het bespreken van afwijkend gebruik en het opwerpen van een aantal opmerkingen.

De Country Risk Briefing (Economist Intelligence unit), de International Country Risk Guide (Political Risk services Group) en het World Competitiveness Yearbook (International Institute for Management Development) brengen hun adviserende rol voor investeerders, bedrijfsleiders en managers naar voren als voornaamste objectief. Deze rankings worden in de praktijk ook voornamelijk afgenomen door de economische topklasse die er ondersteunend materiaal in ziet bij het nemen van financiële en economische beslissingen.

De Corruption Perceptions Index (Transparency International), de Global Competitiveness Index (World Economic Forum) en de Public Sector Performance (European Central Bank) zijn eerder in het leven geroepen om overheden over de wereld informatie te leveren over hun prestaties, alsook *best practices* aan te reiken om die prestaties te verbeteren.

De Country Performance and Institutional Assessments zijn er volledig op gericht *aid allocation decisions* mee te sturen. Ze worden dan ook rechtstreeks ingezet om de International Development Association-middelen van de World Bank mee te verdelen onder de ontwikkelingslanden. Hoewel de Worldwide Governance Indicators (World Bank Institute) mee het governance-gehalte de hoogte in willen drijven, wordt hun ranking ook in belangrijke mate gebruikt in onderzoek naar ontwikkeling en economische

groei. In een interview met een official van de World Bank Group werd aangegeven dat hun governance ranking in eerste instantie onderzoeksdoeleinden moet dienen. De auteurs van deze governance ranking -Kauffman, Kraay, Mastruzzi (en in een eerdere fase ook Zoido-Lobotán)- trachten factoren te identificeren die bijdragen tot de economische opleving en ontwikkeling van een land.

De geanalyseerde governance rankings worden echter niet altijd gebruikt waarvoor ze dienen. Sterker nog: in bepaalde gevallen worden de rankings ingezet voor taken waarvan de producenten gewaarschuwd hebben ze niet voor te gebruiken. Zo baseert de World Bank zich bij de allocatie van haar ontwikkelingshulp -financiële steun, renteloze leningen en materiële en logistieke steun voor lokale ontwikkelingsprojecten- op de Country Performance and Institutional Assessments. In een adem waarschuwt de World Bank dat de Worldwide Governance Indicators (van het gelinkte World Bank Institute) niet geschikt zijn om *aid allocation* mee te sturen. Een aantal landen slaan deze waarschuwing echter in de wind. Zo beroept de Amerikaanse overheid zich in grote mate op de Worldwide Governance Indicators in haar Millennium Challenge Account (zie kader 6). Aan de hand van zestien indicatoren, waarvan vijf rechtstreeks aan de Worldwide Governance Indicators worden ontleend, wordt besloten welke landen recht hebben op een deel van de ontwikkelingshulp van de Amerikaanse overheid en hoe groot elk aandeel is. Ontwikkelingslanden moeten voor de helft van de indicatoren boven de mediaan zitten, willen ze in aanmerking komen voor een deel van de 5 miljard dollar Amerikaanse ontwikkelingssteun.

Kader 6: Millennium Challenge Account

16 Indicatoren van de Millennium Challenge Account

Governing Justly:
- Civil Liberties (Freedom House)
- Political Rights (Freedom House)
- Voice and Accountability (World Bank Institute)
- Government Effectiveness (World Bank Institute)
- Rule of Law (World Bank Institute)
- Control of Corruption (World Bank Institute)

Investing in People:
- Public Primary Education Spending as Percent of GDP (World Bank/national sources)
- Primary Education Completion Rate (World Bank/national sources)
- Public Expenditures on Health as Percent of GDP (World Bank/national sources)
- Immunization Rates: DPT and Measles (World Bank/UN/national sources)

Promoting Economic Freedom:
- Country Credit Rating (Institutional Investor Magazine)
- Inflation (IMF)
- 3-Year Budget Deficit (IMF/national sources)
- Trade Policy (Heritage Foundation)
- Regulatory Quality (World Bank Institute)
- Days to Start a Business (World Bank)

> *5.3.3.4. User*

Tot slot hebben we met de user, ook over het eindpunt van het informatieproces ons licht laten schijnen. De beoogde doelgroep van de governance ranking hangt nauw samen met de producent, de invalshoek van de meting en de doelstellingen die worden nagestreefd met de rangschikking van publieke sectoren.

Een mogelijke benadering is de doelgroep van de governance rankings af te zetten tot de reële gebruikers en een inschatting te doen van de mate waarin de internationale rangschikkingen van publieke sectoren informatie

bieden die relevant is voor de beoogde gebruikers. In deze studie hebben we de governance rankings echter vanuit het oogpunt van de *Public Administration* bekeken. We zijn nagegaan in welke mate de internationale rangschikkingen van publieke sectoren, of het onderdeel dat de kwaliteit van het openbaar bestuur evalueert, nuttig is voor beleidsmakers. De centrale vraag daarbij luidde in welke mate de kwaliteit van de overheid gereflecteerd wordt in de governance scores en rankings.

In de internationale literatuur wordt er grofweg een onderscheid gemaakt tussen een viertal user-groepen (Arndt & Oman, 2006). Ten eerste zijn er de journalisten. Zij trachten nieuws aan de man te brengen en zien in de internationale rangschikkingen van publieke sectoren de ideale toetssteen om de nationale overheid te beoordelen. Cijfers doen het al gemakkelijker in de pers, rankings -simpel, duidelijk en *trendy*- vormen al helemaal een gedroomd instrument voor de media. Op die manier sijpelen de resultaten uit de internationaal vergelijkende studies door naar de bevolking. Zo wordt de burger geïnformeerd en wordt de *accountability*-functie van de overheid voor een deel uitgeoefend.

De praktijk wijst echter op enkele problemen van de journalistieke berichtgeving. Zo worden de cijfers uit de rankings ondanks de vele kwaliteitsgebreken vaak zonder veel kritiek of nuance overgenomen. Dat dit op zijn minst een misleidend beeld schept, wordt in dit onderzoek uitvoerig geïllustreerd. Daarnaast geven journalisten te vaak een onvolledig beeld door er enkele aspecten uit te lichten, die enkel in hun volledig kader correct tot uiting komen. Tot slot is het gevolg van het journalistieke motto 'goed nieuws is geen nieuws' dat de governance data en rankings meestal in een *bad news show* worden weergegeven.

Naast de journalisten behoren ook de internationale investeerders en bedrijfsleiders tot de gebruikers van internationale rangschikkingen van publieke sectoren. Zij gebruiken deze rangschikkingen en governance data als *back-up* voor de traditionele macro-economische risico-indicatoren. Deze laatste slaagden er eind jaren '70, begin jaren '80 immers niet in om de zware financiële crisissen te voorspellen. Meer en meer doen managers en investeerders beroep op de rankings bij het nemen van investeringsbeslissingen of bij relocatie-overwegingen. Zoals hoger beschreven (cf. 1.3), was de onvrede met de bestaande risico-inschattingssystemen bij investeerders een van de *triggers* van

internationale rangschikkingen van publieke sectoren. Politieke instabiliteit, een zwakke *rule of law*, een starre administratie of een weinig efficiënte overheid kunnen investeerders gemakkelijk ervan weerhouden in een land te investeren, aangezien ze het risico op financieel verlies bij economische operaties vergroten.

De derde categorie van gebruikers wordt gevormd door de donoren van ontwikkelingshulp. In de jaren '90 kwam de idee op om de verdeling van de ontwikkelingshulp te verbinden met de governance-prestaties in de Derde Wereldlanden. De theorie hierachter, in grote mate verleend aan *'Aid, Policies and Growth*' (Burnside & Dollar, 1997), was dat ontwikkelingshulp enkel landen vooruit kan helpen die een goed bestuur als basis hebben, zonder een te veel aan corruptie. De mate waarin dat het geval was voor de hulpbehoevende landen moesten de governance rankings uitwijzen. Bovendien zijn de internationale rangschikkingen van publieke sectoren ook instrumenteel in het belonen van goed presterende landen op vlak van de overheidskwaliteit en het straffen van landen die er niet in slagen vooruitgang te boeken op het vlak van governance.

De internationale rangschikkingen van publieke sectoren zijn echter kwalitatief niet in staat om een dergelijk zware last op de schouders te dragen. Er is een groot transparantieprobleem en de kwaliteit van de data is nog niet goed genoeg. Ook de auteurs van de governance rankings hebben reeds meermaals gewaarschuwd dat hun indicatoren niet gemaakt zijn om landen met elkaar te vergelijken. Initiatieven zoals de Millennium Challenge Account van de Amerikaanse overheid waarbij ontwikkelingssteun op basis van governance indicatoren wordt verdeeld, wijzen uit dat deze waarschuwing vaak in dovemansoren is gevallen.

De vierde groep gebruikers zijn de analisten, onderzoekers en academici. Zij gebruiken de data uit de governance rankings in hun kwantitatieve studies. Maar ook hier laten de gebreken van de data zich gelden, wat zorgt voor studies die vaak niet meer geloofwaardigheid bevatten dan de data waarop ze gebaseerd zijn.

In dit lijstje mogen echter twee andere voor de hand liggende gebruikers niet vergeten worden. Ten eerste zijn dat beleidsmakers. Zij zijn sterk geïnteresseerd in internationale rangschikkingen van publieke sectoren en dan vooral in de positie die de eigen overheid daarin bekleedt.

Beleidsmakers voeren een bepaald beleid en zijn verantwoordelijk voor de werking van de publieke sector. Voor deze opdracht heeft het beleid nood aan *benchmarks*. Overheden willen weten waar ze staan tegenover andere landen die andere accenten leggen en gelijkaardige uitdagingen misschien op een andere manier aanpakken. Dit zorgt voor een internationale positionering en biedt de mogelijkheid leereffecten te putten uit *best practices* in andere landen. Vanuit dit oogpunt kijken we in deze studie naar de verschillende internationale rangschikkingen van publieke sectoren, en meer bepaald naar de meting en vergelijking van het overheidsapparaat in verschillende landen.

De tweede categorie van gebruikers die we aan het lijstje willen toevoegen zijn de politici, de oppositieleden en de hervormers. Zij staan vaak langs de zijlijn vanuit hun al dan niet officiële functie om het beleid te controleren. Internationale rangschikkingen van publieke sectoren worden vanuit deze optiek aangegrepen om bepaalde beleidsopties te bekritiseren of hervormingen naar voren te schuiven.

Ook hier wijzen de analyseresultaten op een grote heterogeniteit onder de gebruikers. Geen van de eerder genoemde user-groepen is uit te sluiten voor de verschillende rankings. Het is wel duidelijk dat de Country Risk briefing (Economist Intelligence Unit), de Global Competitiveness Index (World Economic Forum), de International Country Risk Guide (Political Risk Services Group) en het World Competitiveness Yearbook (International Institute for management Development) zich vooral richten op internationale bedrijven die beroep doen op de rankings bij het nemen van financiële en economische beslissingen.

De Worldwide Governance Indicators (World Bank Institute), de Corruption Perceptions Index (Transparency International) en de Country Performance and Institutional Assessments (World Bank) worden vooral gebruikt door donoren en internationale financiële instellingen gericht op ontwikkeling.

De Public Sector Performance (European Central Bank), de Worldwide Governance Indicators (World Bank Institute) en de Corruption Perceptions Index (Transparency International) richten zich in belangrijke mate tot de overheden van de landen uit de rangschikking. Ze trachten hen een instrument aan te reiken om de prestaties van de publieke sector te verbeteren.

Uiteraard zijn de meeste van deze governance rankings ook voer voor journalisten en media. Daarenboven maken ook veel onderzoekers, wetenschappers en analisten gebruik van de data in verklarende studies over ontwikkeling, governance en economische groei. Tot slot laten ook politici, ministers en hervormers de kans niet liggen rankings aan te wenden in hun retoriek.

> 5.4. Conclusies

Kader 7: Evaluatie governance rankings

Governance ranking	Worldwide Governance Indicators	World Competitiveness Yearbook
Producent	World Bank Institute	International Institute for Management Development
Component	Government effectiveness	Government efficiency
Pluspunten	- Grote landensample (212 landen) - Grote (kosteloze) beschikbaarheid van indicatoren en methodologie - Geen absolute aggregatie - Combinatie van *expert opinion*, *firm surveys* en *household surveys* - Verantwoord wegingsschema (precisieweging) - Expliciteert meetfouten in betrouwbaarheidsintervallen	- Goede mix van subjectieve (40) en objectieve (32) indicatoren - Precedent van regionale inclusie: proefproject 2003/2004/2005 met enkele regio's - Vroege lancering data: mei van elk jaar - Goede beschikbaarheid van methodologie, indicatoren en data
Minpunten	- Louter subjectieve evaluatie - Geen comparabiliteit over de jaren heen - Indices die in verschillende landen mogelijks op verschillende manier tot stand gekomen zijn, beperken de comparabiliteit van de ranking - *Private sector* bias - Geen regio's in ranking - Chaotische conceptualisering door een *overload* aan indicatoren (# 49) - Basisscores enkel in herschaalde vorm beschikbaar - Weinig oog voor concrete leermogelijkheden	- Kostelijke beschikbaarheid (€ 3.656) - Beperkte landensample (# 55) - *Sample bias*: enkel geïndustrialiseerde landen in de sample - Chaotische conceptualisering door een *overload* aan indicatoren (# 72) - Vertekende respondentengroep: bedrijfsleiders uit multinationals - Non-representativiteit survey: N = slechts 3.700 voor 55 landen - Economische invalshoek 'government efficiency': bedrijfsvriendelijk klimaat als voorname prioriteit - Geen meetfouten in rekening gebracht - Geen regio's in ranking - Problematische comparabiliteit over de jaren heen, wegens methodologische wijzigingen

Governance ranking	Global Competitiveness Index (WEF)	Country Risk Briefing
Producent	World Economic Forum	Economist Intelligence Unit
Component	Government inefficiency	Government effectiveness risk
Pluspunten	- Grote landensample (131 landen) - Variabele weging ter aggregatie (afhankelijk van ontwikkelingsfase en gerelateerde noden) - Flexibele weging zorgt voor rekenschap nationale context - Beperkt aantal indicatoren (#4), bevordert de begrijpelijkheid	- Grote landensample (150 landen); zowel Westerse landen als ontwikkelingslanden - Geen enge rangschikking, ordinale ranking (A-E) - Dagelijkse update via website (mogelijk) - Regionale revisiepanels voor consistent scoringssysteem
Minpunten	- Veel methodologische veranderingen; nefast voor comparabiliteit - Geen regio's opgenomen - Economische invalshoek 'Government Inefficiency' - Non-representativiteit survey: N = slechts 3.700 voor 131 landen - Flexibele weging vermindert de comparabiliteit - *Private sector* bias: respondenten exclusief economische top management - Wijzigende landensample heeft negatieve invloed op de comparabiliteit - Geen meetfouten in rekening gebracht - Geen regio's in ranking - Geen theoretische onderbouw van meting	- Geen meetfouten in rekening gebracht - Geen regio's in ranking - Weinig transparantie; niet reproduceerbaar - Prijzige toegang tot basisdata; £5.500 - Weinig oog voor concrete leermogelijkheden - Economische invalshoek 'government effectiveness risk': focus op risico's voor bedrijven en economische actoren - *Expert-assessments*: *In-house*-experten missen mogelijk voeling met het te beoordelen land

Governance ranking	**Country Performance and Institutional Assessments**	**Public Sector Performance**
Producent	World Bank	European Central Bank
Component	Quality of Public Administration	Administrative Performance
Pluspunten	- Grote landensample (140): doelgroep bestaat uit ontwikkelingslanden - Naast economische, ook politieke en sociale invalshoek - Sinds enkele jaren verhoogde transparantie van data en methodologie - Intern systeem van *Checks and Balances* ingebouwd voor consistent scoringssysteem - Actionable indicators die oog hebben voor leermogelijkheden voor de overheden - Stabiele landensamenstelling	- Poging tot vaststelling effectiviteit en efficiëntie - Goede beschikbaarheid en transparantie van data en methodologie - Vrije toegang tot ranking en basisscores
Minpunten	- *Expert assessments*: *in-country experts* verhogen het risico op *home-country-bias* - Geen meetfouten in rekening gebracht - Geen regio's in ranking - Geen theoretische onderbouw voor indicatorenselectie	- Beperkte landensample (23); enkel geïndustrialiseerde landen in Europese context - Geen theoretische onderbouw voor indicatoren- en bronnenselectie - Perceptie verkeerdelijk aangenomen als *performance* - Grote overheden (veel inkomsten en uitgaven) worden systematisch benadeeld in de scores - Keuze voor wegingsschema niet onderbouwd - Geen meetfouten in rekening gebracht - Geen regio's in ranking - Geen continuïteit van de meting voorzien - Administratie op zeer enge en selectieve wijze opgevat

Governance ranking	Corruption Perceptions Index	International Country Risk Guide
Producent	Transparency International	Political Risk Services Group
Component	Corruption Perceptions Index	Bureaucratic Quality
Pluspunten	- Grote landensample (180) - Expliciteert meetfouten in betrouwbaarheidsintervallen - Intern systeem van *Checks and Balances* ingebouwd voor consistent scoringssysteem - Goede beschikbaarheid data, methodologie en onderliggende databronnen	- Grote landensample (140) - Geen enge rangschikking, ordinale ranking (*Very Low Risk - Moderate Risk - Very High Risk*) - Vrijgeven van informatie, bronnen en subjectieve analyse waar experten zich op baseren, verhoogt transparantie en leermogelijkheden - Vrij goede beschikbaarheid data, methodologie en indicatoren - Regionale revisiepanels voor consistent scoringssysteem - Maandelijkse update scores via website (mogelijk) - Stabiele landensamenstelling
Minpunten	- Volledig gebaseerd op externe databronnen - *Selection-bias* respondenten: *private sector bias*, geen bevolkingssurveys - Geen onderscheid gemaakt tussen administratieve corruptie en politieke corruptie in de meting - Indices die in verschillende landen op verschillende manier tot stand gekomen zijn, beperken de comparabiliteit van de ranking - Redelijk grote betrouwbaarheidsintervallen; vermindert het aantal significante vergelijkingen en verschillen - Geen regio's in ranking - Transparantie beperkt door de basisdata uit externe bronnen - Beperkte leermogelijkheid: indicatie van een probleem, niet aangepast om concrete oplossingen aan te reiken. - Veranderende methodologie en samplesamenstelling verhinderen comparabiliteit	- Component 'Bureaucratic Quality' stoelt op slechts één onderliggende subjectieve indicator (inschatting experten) - Economische invalshoek 'Bureaucratic Quality': sterkte van de administratie als stabiliserende functie van een land waarin bedrijven moeten opereren - Weinig transparantie; niet reproduceerbaar - Indicatorenselectie en keuze voor wegingsfactoren niet onderbouwd - Geen meetfouten in rekening gebracht - Geen regio's in ranking - Prijzige toegang tot basisdata; $7.940 - Experten onderhevig aan *private sector bias*

6. Klovenanalyse

Het corpus van deze studie wordt gevormd door de synthese van de kwaliteitsproblematiek rond de internationale rangschikkingen van publieke sectoren (cf. hoofdstuk 4) en de empirische analyse van een selectie van acht voorname governance rankings (cf. hoofdstuk 5). Naast deze kwalitatieve doorlichting van de internationale rangschikkingen van publieke sectoren, hebben we in onze studie ook een aanzet gedaan tot klovenanalyse. In dit hoofdstuk spiegelen we de waargenomen verwachtingen op het beleidsniveau op het vlak van internationaal vergelijkende data over de overheidsprestaties aan de realiteit, aan de doelstellingen en agenda's van de producenten die aan de basis van de governance rankings liggen.

In de eerste paragraaf beschrijven we wat de noden binnen het beleid zijn. Aan de hand van een aantal selectieve interviews bieden we een aanblik op de problemen, de noden en de wensen die op het vlak van de governance rankings bestaan. In bijlage geven we een oplijsting van de personen die we hieromtrent gepolst hebben, klankbordgroepleden voor dit project incluis.

Tegen deze horizon plaatsen we het effectieve actuele aanbod. Een *screening* van het governance rankingslandschap levert in de tweede paragraaf een beeld op van wat er bestaat op het vlak van internationale rangschikkingen van publieke sectoren. Contacten met vertegenwoordigers van de producenten van deze rankings hebben ook een beter zicht opgeleverd van de doelstellingen, beweegredenen en toekomstplannen die in het kamp van de producenten leven.

> 6.1. Naar wat zoeken de beleidsmakers?

We hebben ons in dit onderzoek voornamelijk gericht op governance rankings van bestuurskwaliteit. We hebben proberen uit te zoeken in welke mate ze geschikt zijn voor overheden om zich internationaal te positioneren en om de conclusies mee te nemen in het besluitvormingsproces. Om die vraag goed te kunnen beantwoorden moest er ook gekeken worden naar wat de beleidsmaker juist zoekt.

Overheden zien internationale rangschikkingen van publieke sectoren als beleidsinformatie. Het gaat om strategische informatie die als hulpmiddel kan dienen bij de uitvoering van de diverse overheidstaken. Zo zijn de data instrumenteel om zich als overheid internationaal te positioneren. Nationale overheden trachten in te schatten hoe goed of slecht er gepresteerd wordt in vergelijking met buurlanden of andere ijkpunten. Dit is het *benchmarking*-principe.

Met de *benchmarks* kunnen nationale overheden onderling hun prestaties vergelijken. Vanuit die wetenschap kunnen beleidsmakers vervolgens lessen trekken om de werking van het overheidsapparaat te verbeteren. De meting legt immers de pijnpunten in het eigen overheidsapparaat bloot, die aan de basis liggen van een gebrekkige effectiviteit en efficiëntie van de overheid. De vergelijking helpt zo mee om de onzekerheden beter in kaart te brengen en de met verschillende beleidsopties verbonden kansen en risico's nader te bepalen.

Best practices kunnen als voorbeeld gelden en de eigen prestaties op een hoger niveau tillen. "Het is misschien niet toevallig dat juist in de tijd dat iedereen de ideologische veren heeft afgeschud, de stofjas van de benchmarking zo gemakkelijk wordt aangeschoten. De vergelijking van de eigen prestaties met die van anderen laat als vanzelf zien wie de besten zijn en wie dus als toetssteen voor het eigen beleid moet gebruikt worden", aldus een veelbesproken rapport van de Nederlandse overheid over internationale vergelijkingen van publieke sectoren (Sociaal en Cultureel Planbureau, 2007).

Om nuttig te zijn, moeten indicatoren niet enkel een valide meting doen van de overheidskwaliteit, ze moeten ook een concreet aanknopingspunt vormen voor beleidsinput; ze moeten relevant zijn. Overheden zijn met

andere woorden het meest gediend bij *actionable* of *action-worthy indicators*. Dergelijke variabelen zijn verbonden aan concrete beleidsopties, waardoor ze geïdentificeerd kunnen worden met bepaalde beleidsoverwegingen. Bovendien is het belangrijk dat de variabelen binnen de invloedsfeer van de overheid vallen, zodat beleidsmakers door middel van concrete maatregelen een directe invloed kunnen uitoefenen op het gemeten fenomeen en van daaruit op de score en ranking.

> 6.2. Wat kunnen en willen producenten leveren?

De internationale organisaties, NGO's en adviesbureaus die aan de oorsprong van de internationale rangschikkingen van publieke sectoren liggen, hebben echter ook elk hun agenda. De aard en invulling van hun governance ranking is sterk afhankelijk van de aard, de objectieven en het doelpubliek van de producent zelf.

We kunnen bijvoorbeeld enkele argumenten opperen vanuit de producentenhoek ten aanzien van de aangehaalde kwaliteitsgebreken. Een eerste is een praktisch obstakel waar alle *rankers* mee dienen rekening te houden. Het betreft de beperkte beschikbaarheid van data over de performantie van de publieke sector. De kwaliteit van de overheid is een weinig tastbaar fenomeen. Dit heeft tot gevolg dat bepaalde aspecten van de overheidwerking moeilijk meetbaar zijn. Als er al officiële cijfers bestaan over governance-aspecten, is het bovendien vaak erg moeilijk om gelijkaardige statistieken te vinden over de landsgrenzen heen. Veel landen missen immers net de institutionele kwaliteiten en structuren om dergelijke data op een consistente en betrouwbare manier bij te houden.

Daarnaast moet voor ogen gehouden worden dat een groot deel van de internationale rangschikkingen van publieke sectoren niet in de eerste plaats voor de beleidsmakers worden opgesteld, maar wel voor geïnteresseerde 'klanten'. Zo worden de governance rankings meer nog dan een leerinstrument; een product. *Risk rating agencies*, private adviesbureaus of andere organisaties met winstoogmerk brengen daarop comparatieve data op de markt die het ondernemingsklimaat, de interne risicofactoren en overheidskwaliteiten in verschillende landen inschatten.

Vanuit dit perspectief is het minder interessant naar *actionable indicators* te zoeken, dan wel in te spelen op de noden van de afnemers. Dat er bijvoorbeeld geen governance-theorie aan de basis van de indicatorenselectie ligt, zal weinig bedrijfsleiders en managers interesseren. Zij zijn immers op zoek naar data die de investeringsrisico's inschatten in een land waar ze economische activiteiten willen ontplooien of financiële transacties mee willen afsluiten. Illustratief is de repliek die we van een medewerker aan de International Country Risk Guide (Political Risk Services Group) kregen. "*Rather than being theoretical exercises, our choices in such matters are business decisions based on business needs*." Ook het feit dat de

government effectiveness-meting een bestuurskundig zeer selectief beeld oplevert van de prestaties van de overheid, is vanuit dit opzicht van minder belang.

Ook het niet expliciteren van meetfouten en berekenen van betrouwbaarheidsintervallen door veel *commercial business information providers* moet in deze zin opgevat worden. Het is niet vergezocht dat commerciële overwegingen het in veel gevallen halen op wetenschappelijke gronden. Betrouwbaarheidsintervallen ondermijnen vanuit die visie veeleer de geloofwaardigheid van de data en de rankings dan deze te versterken.

Ook op het vlak van transparantie staan commerciële belangen wel eens in de weg. Waar wij het vragen van een hoge prijs voor de toegang tot de basisdata van de verschillende rankings een obstakel voor de databeschikbaarheid, transparantie en uiteindelijk ook de leereffecten beschouwen, hebben veel producenten van rankings daar een heel andere visie over. Indien zij hun data vrij de wereld insturen, verdwijnen hun winsten en kunnen zij misschien hun jaarlijkse meting opdoeken. Er kruipt immers flink wat geld en energie in de samenstelling van een governance ranking; het geldt daarvoor moet ergens vandaan komen.

7. Slotconclusie

Tot slot willen we besluiten door de belangrijkste conclusies uit deze studie naar de kwaliteit van rangschikkingen van publieke sectoren weer te geven. In eerste instantie overlopen we wat er zoal misloopt met de governance rankings van vandaag. De meta-synthese van de kwaliteitsproblematiek, gecombineerd met de empirische analyse van governance rankings geeft een goed overzicht van de tekortkomingen van de huidige comparatieve overheidsmetingen. Enerzijds geven we een overzicht van de kwaliteitsproblemen en de lessen die ze meebrengen voor de gebruiker, de producent en het Vlaamse beleid. Anderzijds werpen we ook een blik op de toekomst. We kijken naar de initiatieven van morgen op het vlak van comparatieve metingen van publieke sectoren. Daarnaast doen we ook suggesties naar waar de wetenschap nog een bijdrage aan de problematiek rond de rankings kan leveren.

In deze slotconclusie trachten we vanuit de wetenschap van de kwaliteitsproblemen van governance rankings een stap verder te zetten. We zullen in een eerste paragraaf de oplijsting van kwalitatieve pijnpunten vertalen in lessen voor drie verschillende actoren. Ten eerste vertalen we de gebreken die de comparatieve data en rankings kenmerken in concrete aanbevelingen voor de gebruiker. Dat de internationale rangschikkingen van publieke sectoren nog fouten en tekortkomingen vertonen, wil immers niet zeggen dat ze nutteloos zijn. We mogen het kind niet met het badwater weggooien.

Zo zijn de rankings, vanuit het perspectief van de beleidsmaker, vooral instrumenteel voor het detecteren van de interne ondermaat. Relatief slechte scores op bestuursaspecten wijzen de overheid op negatieve prestaties. Op die manier gelden de indicatoren als flikkerlichten die aandachtspunten kunnen aanwijzen.

Er kunnen echter door de gebruiker uit de rankings pas juiste beleidslessen getrokken worden, wanneer die zich bewust is van de kwaliteitsgebreken van deze data. Onder de vorm van een lijst aanbevelingen, trachten we de gebruiker tot een geïnformeerde en gewaarschuwde gebruiker te maken. Een besef van wat de data wel en niet bewijzen, van welke vergelijkingen betekenisvol zijn en welke niet, en van het perspectief van waaruit de data

moeten geïnterpreteerd worden, moet bijdragen aan een omzichtig, genuanceerd en weloverwogen gebruik van internationale rangschikkingen van publieke sectoren.

In tweede instantie doen we een lezing van de kwaliteitsproblemen van de huidige rangschikkingen van publieke sectoren in termen van kwaliteitswerkpunten voor de initiator van de rankings, de producent, de *ranker*. Zoals overheden in een constant proces zitten ter verbetering van de efficiëntie en effectiviteit van de publieke sector, bestaat er ook een continue dynamiek onder producenten van governance rankings om de kwaliteit van hun meet- en vergelijkingsinstrument te verbeteren. Met het weergeven van de conceptuele, technische en methodologische problemen van de governance rankings van vandaag, trachten we de producenten aan te zetten verder werk te maken van een productverbetering. Enkel een verbeterde kwaliteit van comparatieve metingen van publieke sectoren zal de betrouwbaarheid en hanteerbaarheid van de rankings doen stijgen. Hierbij zijn zowel de (commerciële) producenten als de gebruikers gebaat.

Ten derde belichten we het huidige rankingslandschap ook specifiek vanuit de Vlaamse beleidscontext. Enerzijds wijzen we op de zo goed als onbestaande regionale component in de 'inter-nationale' rangschikkingen van publieke sectoren en trachten we de motieven achter deze lacune te identificeren. Anderzijds doen we ook een aantal aanbevelingen voor de Vlaamse administratie die een betere hanteerbaarheid van de rankings voor Vlaanderen naar de toekomst toe in de hand kunnen werken.

In een tweede paragraaf doen we een aanzet naar de toekomst toe. Enerzijds zullen we enkele scenario's schetsen voor de internationale rangschikkingen van publieke sectoren van morgen. We belichten onder andere het initiatief *'Government at a glance'* dat de Organisation for Economic Co-operation and Development in de nabije toekomst zal opstarten. Anderzijds willen we ook het wetenschappelijk onderzoeksveld enkele pistes aanreiken voor toekomstige studie. De academische en onderzoekswereld moet kritisch blijven toezien op de diverse initiatieven van comparatieve metingen van publieke sectoren. De grote aandacht die de media aan dergelijke instrumenten besteden, het grote gewicht dat de rankings onder andere in het ontwikkelingsdebat krijgen toebedeeld en de niet te onderschatten invloed die de internationaal-comparatieve data hebben op het imago van een land, vragen immers om objectieve

lijnrechters. De wetenschap mag zich daarbij niet laten misleiden door de verkoopspraatjes waarmee de producenten van de governance rankings hun producten voorstellen. Ze moet de kwalitatieve problemen, lacunes en misbruiken aan de kaak blijven stellen en op die manier een constante druk uitoefenen op de producenten om de kwaliteit van de rankings hoog te houden.

> 7.1. Algemene bevindingen

In onze studie hebben we in eerste instantie getracht een meta-synthese te maken van de kwaliteitsproblematiek rond de rangschikkingen van publieke sectoren. De governance rankings van vandaag staan volop in de belangstelling, maar zijn verre van onbesproken. De pijnpunten van comparatieve metingen van 'de goede overheid' zaten vooralsnog enigszins versnipperd in de internationale literatuur. Met ons onderzoek hebben we deze problemen samengebracht en aan de hand van een eigen empirische analyse getracht verder zicht te krijgen op de stand van zaken. De eerste drie hoofdstukken van dit rapport vormden een aanloop naar de kern van het onderzoek. Ze schetsen de context van de governance rankings en illustreren vanuit welke probleemstelling dit onderzoek is ontsproten. We hebben er enkele concepten afgelijnd, stelden de onderzoeksvragen op en duidden de theoretische en methodologische basis van de studie.

Hoofdstuk 4 en 5 vormen het zwaartepunt van het onderzoek 'Internationale vergelijkingen vergeleken: Een meta-analyse van rangschikkingen van publieke sectoren'. In hoofdstuk 4 doen we een overzichtelijke oplijsting van de kwaliteitsproblemen van de rankings. We kwamen zo aan negen grote lijnen van kritiek. Hieronder geven we ze nog eens in een overzicht mee (zie kader 8).

Kader 8: Listing pijnpunten governance rankings

1. Foutief gebruik van subjectieve indicatoren
2. Non-representativiteit van respondenten in surveys
3. Conceptuele tekortkomingen in de meting
4. Meetfouten
5. Onderbelichting van politiek-administratieve dimensie
6. Onderhuidse bias
7. Over-aggregatie
8. Gebrek aan transparantie
9. Vergelijkbaarheidsproblemen

Hoofdstuk 5 beschrijft de empirische analyse waarbij we de concrete kwaliteitsaspecten van acht hoogaangeschreven rangschikkingen van publieke sectoren nauwgezet onderzocht hebben. Paragraaf 5.4 gaf reeds

een overzicht van de conclusies uit de meta-fiches. Voor de basisresultaten van de analyse, kan u bij de bijgeleverde meta-fiches terecht.

In deze paragraaf zetten we een stap verder en richten we ons tot de gebruikers en producenten van internationale rangschikkingen van publieke sectoren door hen op basis van onze onderzoeksbevindingen enkele aanbevelingen te formuleren. We kijken wat deze respectievelijke actoren kunnen leren uit de vastgestelde kwaliteitsproblemen. Door zowel *rankers* als *users* op hun verantwoordelijkheden te wijzen, trachten we de relevantie van de governance rankings te versterken. Aan de ene kant zijn er de kwaliteitsgebreken, die de producenten verder moeten zien weg te werken. Aan de andere kant moeten gebruikers echter ook de bestaande gebreken en nuances van de data meenemen in hun conclusies. Momenteel worden er nog te vaak overhaaste conclusies met de comparatieve data van publieke sectoren verbonden. De waarschuwingen die bij de governance rankings geleverd worden, worden door gebruikers en media in veel gevallen in de wind geslagen.

In eerste instantie lijsten we een aantal aandachtspunten voor de gebruikers op. Deze punten in acht nemen zou voor een meer genuanceerde en correctere interpretatie van de data en de vergelijkingen uit de governance rankings moeten zorgen. Onze stelling is dat de internationale rangschikkingen van publieke sectoren pas op een correctere manier zullen gebruikt worden als gebruikers zich rekenschap geven van de kwalitatieve tekortkomingen die in de rankings vervat zitten.

Ten tweede schotelen we de producenten een aantal werkpunten voor. We wijzen hen op de reeks kwaliteitsmankementen, die aan het licht kwamen uit onze synthese van de kwaliteit van governance rankings en geïllustreerd werden in onze empirische analyse. We spreken de internationale organisaties en adviesbureaus daarbij aan op hun verantwoordelijkheid om de kwaliteit van de data en de methodologie van de rankings te bewaren.

In een derde paragraaf besteden we even aandacht aan de specifieke noden van het Vlaamse beleidsniveau op het vlak van internationale rangschikkingen van publieke sectoren. Vlaanderen wil zich als regio ook versterken, profileren en positioneren in de wereld. De huidige governance rankings spelen echter nog niet op deze noden in.

> *7.1.1. Aanbevelingen gebruikers*

De kwaliteitsgebreken van internationale rangschikkingen van publieke sectoren, die in kritische studies als deze aan het oppervlak komen, zijn in eerste instantie de verantwoordelijkheid van de producenten. Maar ook de gebruikers van governance rankings, moeten op een meer verantwoorde manier met de data omspringen. Er worden nog te veel ongenuanceerde en overhaaste conclusies getrokken uit de rangschikkingen. Er worden vergelijkingen getrokken tussen landen die significant geen verschil tonen, er wordt gezocht naar evoluties zonder dat er betekenisvol vergeleken kan worden over de jaren heen, en men verwaarloost de genuanceerde informatie die de basisindicatoren bieden. Bovendien neemt men het vaak niet al te nauw met de waarschuwingen die de producent bij wijlen meegeeft.

Men vergeet te snel dat de governance data vaak aan de noodzakelijke nauwkeurigheid ontberen, dat vergelijkingen niet zelden zonder betekenis zijn en dat er in veel gevallen niet gemeten wordt wat er beweerd wordt te meten. 'Met cijfers kan men alles bewijzen', luidt een boutade. Het is belangrijk dat gebruikers een goed beeld van de kwaliteit van metingen van publieke sectoren krijgen, waardoor de data op een nuttige en relevante manier kunnen aangewend worden.

In deze studie hebben we totnogtoe de internationale rangschikkingen van publieke sectoren op de rooster gelegd en hun zwaktes en tekortkomingen besproken. Dit neemt niet weg dat vergelijken nuttig blijft. Ze zijn voornamelijk instrumenteel in het detecteren van relatieve zwaktes van het overheidsapparaat. Deze kunnen leiden tot het verhogen van de aandacht voor bepaalde domeinen of tot het verleggen van de beleidsprioriteiten. Hierbij kan het helpen om te kijken naar landen meet soortgelijke achtergrond en gelijkaardige uitdagingen, om er *best practices* uit over te nemen. Dit kadert in een strategie om de effectiviteit en de efficiëntie van de publieke sector constant te doen stijgen. Belangrijk daarbij is de governance rankings te lezen op wat ze werkelijk aantonen. In wat volgt doen we een reeks aanbevelingen om tot een correctere en nuttigere interpretatie van de governance rankings te komen.

o

Heb oog voor wat echt gemeten wordt

Het eerste en waarschijnlijk ook belangrijkste aandachtspunt betreft het onderwerp van de meting. Een concept zoals 'de efficiënte overheid', 'een goede werkende administratie' of 'de kwaliteit van het openbaar bestuur' mag niet blindelings overgenomen worden. De producenten claimen deze waardegeladen fenomenen met hun meting te vatten, maar een onderzoek van de onderliggende indicatoren openbaart hun ware inhoud. Illustratief in dit kader is het feit dat de verschillende governance rankings veelal een heel andere invulling aan gelijkaardige concepten geven. De hamvraag wordt dan ook: Wat wordt er juist gemeten door de internationale rangschikking van publieke sectoren? In eerste instantie moet er daarvoor gekeken worden naar de operationalisering. De basisvariabelen vormen de basisbestanddelen van de meting. Daarnaast spelen ook de meetprocedures een belangrijke rol. Daarbij is het belangrijk de bijdrage van het gebruikte wegingsschema te evalueren. De wegingsfactoren vormen de beste indicatie van het belang dat producenten aan een bepaald aspect in een te meten fenomeen toekennen. Idealiter ligt een governance-theorie aan de basis van de indicatorenselectie en het wegingsschema; op die manier worden de (causale) verbanden vanuit wetenschappelijk oogpunt geduid, wat andere, individuele, commerciële of politieke motieven tot een minimum moet herleiden.

- **Focus op de juiste ijking**

Een tweede aandachtspunt houdt in dat er best vergeleken worden met de juiste ijking. Hiervoor is het belangrijk een aantal referentielanden te nemen waaraan de eigen prestaties kunnen gespiegeld worden. In de meeste gevallen zullen de directe buurlanden hiervoor het best geschikt zijn (Van Roosbroek & Van Dooren, 2006). Zij opereren vaak in een gelijkaardige context en ondervinden dezelfde externe invloeden. Bovendien zijn ze ook visueel het sterkst aanwezig voor de bevolking, de politici en de media. Vanuit deze optiek is het ook belangrijk te kijken naar welke landen nog deel uitmaken van de landensample. 'In het land der blinden is eenoog koning'; de (vaak relatieve) scores en gerelateerde rankings zijn immers afhankelijk van de ijkingspunten.

-

Houd rekening met de analyse-eenheid

Een derde aspect waar men moet naar kijken is de analyse-eenheid. Zoals we in de vorige paragraaf (cf. 6.3) hebben onderstreept, vertegenwoordigen de data voor ons land steevast de prestaties van de Belgische publieke sector. Deze data mogen dus niet verward worden met de Vlaamse of Waalse situatie, die in vele gevallen andere cijfers zou voortbrengen. Onvolledige data leiden tot foute input voor het beleid (Van Roosbroek & Van Dooren, 2006). Om deze reden moet het opsplitsen van internationale vergelijkende indicatoren (zoals de Studiedienst van de Vlaamse Regering dat reeds in zekere mate doet) verder een prioriteit blijven. Dit is een noodzakelijke voorwaarde om op elk niveau de juiste beleidsconclusies te trekken.

- **Let op de meetfouten**

Ten vierde mag men niet voorbijgaan aan meetfouten. Meetfouten, zo geven veel producenten van governance rankings zelf toe, zijn onoverkomelijk. Het probleem is onder andere dat veel van de rankings geen metingen zijn, maar schattingen door middel van surveys. Inherent aan het werken met surveys is dat context en methode een sterke invloed hebben op de resultaten (Billiet, 1993; Billiet, 1993). Zo kan een kortstondige crisis, al dan niet uitvergroot in de media, een grote invloed uitoefenen op de perceptiescores voor een land. Bovendien zal zelfs een goed uitgevoerde survey enkel een schatting kunnen geven van een fenomeen (Van Roosbroek & Van Dooren, 2006).

Bij surveys kunnen de marges nog relatief gemakkelijk berekend worden. Bij 'objectieve' data die zijn aangetast door meetfouten of een gebruik van uiteenlopende definities, is deze inschatting echter zo goed als onmogelijk. Sommige auteurs stellen dan ook dat harde data mogelijk een grotere onbetrouwbaarheid en onnauwkeurigheid herbergen dan perceptiedata. Ondanks de wetenschap van de imprecisie van deze metingen, zwijgen de meeste producenten ze liever dood. Betrouwbaarheidsmarges die uit een expliciete erkenning van meetfouten voortvloeien, devalueren in hun ogen hun product. Dit weerhoudt vooral de commercieel ingestelde producenten ervan om betrouwbaarheidsmarges aan de ranking toe te voegen.

Er zijn niettemin enkele uitzonderingen op deze regel. Het World Bank Institute en Transparency International wijzen expliciet op de

onnauwkeurigheid van de meting en berekenen betrouwbaarheidsmarges (*confidence intervals*) voor de scores. Wanneer we rekening houden met deze marges, blijkt pas echt hoe weinig betekenisvolle vergelijkingen er nog overblijven. Tussen landen die zich in de rankings van elkaar onderscheiden door kleine scoreverschillen is in statistische termen geen onderscheid te maken. Dit wordt verder versterkt door het veel voorkomende fenomeen van '*the messy middle*' (Besançon, 2003). Vele rankings vertonen enkele *top performers*, een groepje achterblijvers en daarnaast een groot peloton met de rest van de landen. In deze middengroep zijn vaak weinig of geen significante verschillen op te tekenen.

- **Houd rekening met de oorsprong van de ranking**

Aangezien governance rankings in grote mate gebaseerd zijn op percepties van bevolkingsgroepen of experten, is het ten vijfde belangrijk voor ogen te houden wie er bevraagd wordt. Concepten als 'de efficiënte overheid' of 'een goed presterend overheidsbestuur' zullen door bedrijfsleiders op een andere manier ingevuld worden dan door burgers of ontwikkelingsexperten. Opnieuw is het belangrijk dat we achter de cijfers kijken door de methodiek bloot te leggen. Pas dan kunnen er juiste conclusies getrokken worden. Naast de aard van de respondent, speelt ook de omvang van de sample mee in het inschatten van de betrouwbaarheid van de surveyresultaten. Hoe groter het aantal respondenten, hoe groter de kans op een representatief staal.

Hierbij aansluitend hebben we proberen aan te geven dat elke governance ranking vertrekt vanuit een bepaalde invalshoek, met een bepaald doelpubliek voor ogen en met een duidelijke doelstelling in gedachte. Deze vijfde aanbeveling luidt dan ook deze politiek-administratieve component in beschouwing te nemen bij gebruik van de governance rankings. Het is belangrijk te weten door welke bril de ranking moet bekeken worden. Worden de data gemeten om investeringsbeslissingen mee te sturen, wordt de ranking samengesteld als basis voor de allocatie van ontwikkelingshulp, of moeten de cijfers een mondiale problematiek onder de aandacht te brengen? Worden de internationale rangschikkingen van publieke sectoren in eerste instantie gebruikt door donoren, of door bedrijfsleiders, of door onderzoekers en economici? Allemaal aspecten die een rol spelen bij een goed begrip van de comparatieve data.

> *7.1.2. Werkpunten producenten*

Waar we gebruikers aan de hand van een reeks aandachtspunten aanmanen tot omzichtigheid bij het interpreteren van de internationale rangschikkingen van publieke sectoren, sporen we producenten aan verder werk te maken van de optimalisering van de kwaliteit van de rankings. De oplijsting van kwaliteitsproblemen van internationale rangschikkingen van publieke sectoren in hoofdstuk 4 kan opgevat worden als een aandachtslijstje met werkpunten voor de producenten van deze governance rankings. We sommen hier nog enkele concrete aanbevelingen op, waarvan we denken dat ze de geloofwaardigheid en kwaliteit van de rankings enkel kunnen verbeteren.

- **Stel een onafhankelijke revisie van de data in**

Ten eerste wordt er vanuit de internationale kritische literatuur gesuggereerd dat het waardevol zou zijn om een statistische autoriteit in te stellen die toekijkt op de kwaliteit van de gebruikte data en de geleverde statistieken. Enerzijds is er het probleem dat internationale organisaties vaak afhankelijk zijn van nationale statistische diensten voor de datatoelevering (Van Roosbroek & Van Dooren, 2006). Hier rijzen vaak problemen. Er moet gewerkt worden met uniforme definities en er moet controle zijn op de correcte naleving hiervan. Beide voorwaarden zijn vaak niet vervuld (Van de Walle, 2006b). Een goed voorbeeld van een controlerende instantie vinden we terug in het Verenigd Koninkrijk, waar het *Office for National Statistics* een kwaliteitslabel aan goede data geeft.

- **Motiveer de gemaakte keuzes**

Ten tweede is het van cruciaal belang dat de gebruiker kleur bekent. Meten is kiezen; in alle metingen moeten keuzes gemaakt worden betreffende het voorwerp, de methodologie en de output van de meting. Deze keuzes zijn van subjectieve en normatieve aard. Dat deze beslissingen door gekleurde motieven worden ingegeven, is op zich niet het grootste probleem. Maar de producent moet er op toezien dat deze keuzes verantwoord worden en het de motieven expliciteert die achter de overweging schuilgaan.

De openheid waarvoor hier gepleit wordt, moet zich op diverse domeinen uiten. Zo hebben we in deze tekst uitvoerig het belang van de politiek-administratieve dimensie van de governance rankings geïllustreerd: Door wie? Voor wie? En met welke doelstelling voor ogen wordt er gemeten? Antwoorden op deze vragen helpen gebruikers de data vanuit de juiste invalshoek te interpreteren. De beleidsvisie achter de index moet dan ook geëxpliciteerd worden (Van Roosbroek & Van Dooren, 2006).

Bovendien moet er meer zorg besteed worden aan het conceptuele kader en de indicatorenselectie. Keuzes voor deze of gene indicator, databron of wegingsmodaliteit worden zelden of nooit onderbouwd. Het is echter van groot belang dat de auteurs van de internationale rangschikkingen van publieke sectoren aangeven op welke basis bepaalde inhoudelijke en methodologische pistes in de meting werden verkozen boven andere.

Het gebrek aan eensgezindheid over de inhoud en structuur van internationale metingen van publieke sectoren speelt een grote rol in de diversiteit van internationale rangschikkingen van publieke sectoren. Vragen als 'Wat is een goede overheid?' en 'Hoe moet een goede overheid gemeten worden?' spelen een bepalende rol in het debat over de kwaliteit van governance rankings. Dit debat sleept al lang aan en zal niet op korte termijn beslecht worden. We doen niettemin twee suggesties die een bijdrage kunnen leveren aan dit debat.

Enerzijds kan het nuttig zijn dat de producenten van internationale rangschikkingen van publieke sectoren dezelfde weg bewandelen die de producenten van *university rankings* hebben doorlopen. Met het oog op meer uniformiteit en betrouwbaarheid in de wereld van de *university rankings*, heeft de United Nations Educational, Scientific and Cultural Organization (UNESCO) de verschillende *rankers* over de wereld met elkaar in contact gebracht. In de hoop dat dit zou leiden tot verbetering en afstemming van de gebruikte criteria werd in mei 2006 zelfs een conferentie belegd. Dit mondde uit in de *Berlin Principles of Rankings*, een tiental technische en kwaliteitsregels die de ranker belooft te hanteren (Langouche, 7 A.D.). Dit is een interessante piste die ook voor *rankers* van publieke sectoren haar nut zou kunnen bewijzen.

Het gebrek aan unanimiteit over de bouwstenen van een meting van de overheidskwaliteit hangt samen met de diversiteit aan politieke ideeën die

er in onze samenleving bestaat. Het kerntakendebat over wat een goede overheid wel en niet zou moeten doen, willen we hier niet voeren. Wel doen we de suggestie om deze invulling door mogelijks in samenspraak met de burger zelf te doen. Waarom niet aan de hand van subjectieve bevragingen uitzoeken wat de burger werkelijk verlangt van de overheid? De indicatorenselectie en keuze voor wegingsfactoren kan dan op deze basis tot stand komen. Het afzetten van objectieve indicatoren tegen een subjectieve vraag, kan een meer betekenisvolle rangschikking van overheden opleveren (Yang & Holzer, 2006).

- **Breng onnauwkeurigheden in beeld**

Een derde aandachtspunt betreft het expliciteren van meetfouten. Zoals hoger gesteld zijn de governance rankings niet vrij van meetfouten, de meeste producenten ontberen echter de eerlijkheid om er voor uit te komen dat de metingen niet perfect nauwkeurig zijn. Het berekenen en toevoegen van betrouwbaarheidsintervallen die moeten aangeven welke marge moet in acht gehouden worden vooraleer van een significante vergelijking of verandering te mogen spreken, zou niet mogen ontbreken. Totnogtoe zijn de producenten die betrouwbaarheidsmarges berekenen (World Bank Institute, Transparency International) eerder de uitzondering dan de regel. Dit werkt vergelijkingen in de hand die zonder betekenis zijn. Indien producenten hun rankings meer geloofwaardigheid en betrouwbaarheid willen bieden, moet er meer aandacht besteed worden aan het in rekening brengen van meetfouten.

- **Zorg voor een transparante rangschikking**

Ten vierde is er ook op het vlak van de transparantie en beschikbaarheid van data en methodologie binnen veel governance rankings nog heel wat werk aan de winkel. Het moet gezegd worden dat de laatste jaren veel internationale rangschikkingen van publieke sectoren transparantie hoog op de agenda hebben geplaatst. Er is dan ook enige vooruitgang op dit vlak waar te nemen. Zo verhinderde het World Bank Institute nog niet lang geleden de toegang tot de basisscores van alle landen. Na aanhoudende kritiek en enkele interne hervormingen, worden nu zo goed als alle databronnen die aan de basis van de landenscores liggen, geopenbaard (doch enkel in genormaliseerde toestand).

Ook de almaar stijgende prijzen voor toegang tot vele databronnen van governance rankings is een obstakel voor de transparantie en belemmert de databeschikbaarheid. Zolang dergelijke data deel uitmaken van het commerciële proces van vraag en aanbod, is dit echter een onstuitbare evolutie.

- **Verlies de leerfunctie van de ranking niet uit het oog**

Tot slot luidt een vijfde aanbeveling voor producenten dat ze meer aandacht moeten hebben voor de leerfunctie van de rankings. Vandaag de dag worden nationale overheden wel op een rangschikking van slecht naar goed presterende landen geplaatst, maar wordt er weinig informatie aan diezelfde overheden aangereikt over het waarom van hun slechte beoordeling. De leermogelijkheden van de rankings, waarbij aangegeven wordt wat de concrete zwaktes zijn en vooral hoe die best kunnen aangepakt worden, zouden echter een hogere prioriteit moeten krijgen. Vanuit deze optiek komt er ook meer en meer kritiek op het feit dat er te veel gefocust wordt op de ranking zelf, terwijl de relevante en interessante informatie net in die basisscores te vinden is. De mediatisering van de governance rankings heeft hier in negatieve zin toe bijgedragen. De *horse races* trekken de aandacht weg van de ranking als leerinstrument.

Sommige critici pleiten er dan ook voor de internationale metingen van publieke sectoren niet door te trekken tot een rangschikking. De data die uit de meting naar voren komen kunnen ook per land weergegeven worden, waarbij landen simpelweg in alfabetische volgorde gerangschikt worden (Dill & Soo, 2005). Zo komt de nadruk weer te liggen op de basisscores en is de producent bovendien verlost van het aggregatieprobleem waarbij elke indicator een specifiek gewicht moet krijgen.

> 7.1.3. Aanbevelingen Vlaamse beleid

Alle zaken waar de doorsnee beleidsmaker naar op zoek gaat in internationale rangschikkingen van publieke sectoren, gelden ook voor het Vlaamse niveau. Al moet er in het geval van Vlaanderen nog een belangrijke kanttekening bij gemaakt worden. Vlaanderen bevindt zich als autonome regio met eigen perspectieven en aspiraties eerder in een relatief uitzonderlijke situatie. De specifieke noden van Vlaanderen vallen in de

internationaal-comparatieve oefeningen op het vlak van de publieke sector steevast in dovemansoren.

In de individuele gesprekken met vertegenwoordigers van de Vlaamse overheid, kwam dit laatste sterk naar voren. Het gebrek aan internationaal-comparatieve data voor Vlaanderen wordt als een belangrijk gemis ervaren. De empirische analyse van acht van de meest gebruikte internationale rangschikkingen van publieke sectoren illustreert deze lacune. De governance rankings van vandaag focussen zich volledig op nationale overheden. België vindt wel ingang in de meeste governance rankings, van Vlaanderen is echter nergens sprake.

Dit is vanuit het oogpunt van de producent ook niet onlogisch, aangezien de analyse-eenheid in de regel op landen wordt vastgelegd. In een federaal land met sterke regio's als België wordt dit echter als problematisch ervaren. Zo worden bijvoorbeeld prestaties op het vlak van onderwijs op Belgisch niveau beoordeeld. Deze *country scores* vinden vervolgens ingang in de internationale rangschikkingen van publieke sectoren. In ons land is onderwijs echter een zo goed als volledig geregionaliseerde bevoegdheid, wat betekent dat Vlaanderen in principe een eigen afrekening van haar beleid zou moeten krijgen. Ook voor een heel aantal zaken op het vlak van het bestuursapparaat is er in ons land een regionale aansturing, terwijl de governance rankings vooralsnog een nationale afrekening maken. De rankings die de kwaliteit van het overheidsbestuur meten en internationaal vergelijken, kunnen voor ons land pas de gepaste conclusies aanbrengen als Vlaanderen en Wallonië een afzonderlijke beoordeling en vermelding krijgen.

Voor internationale organisaties blijkt het echter weinig interessant om regio’s in de rankings op te nemen. Ten eerste is er op mondiaal niveau eerder beperkt vraag naar dergelijke regionale data. Naast Vlaanderen zijn er nog wel andere gekende regio’s met autonome bevoegdheden, zoals bijvoorbeeld Québec, Catatonie of Baskenland. Maar op wereldschaal is dit voor de producenten van governance rankings helaas een verwaarloosbaar aandeel. Zo bleek uit een reactie van een vertegenwoordiger van het World Economic Forum, die de Global Competitiveness Index opstelt. Deze *ranker* heeft in het verleden uitzonderlijk wel al eens regionale data opgesteld, namelijk voor regio’s in Oekraïne, Marokko en Brazilië. Dit gebeurde echter op externe vraag en enkel op voorwaarde dat de kosten hiervoor volledig

vanuit ontwikkelingsorganisaties werden gefinancierd. "*We would only include regions if this fits in our strategy*", voegde hij er nog aan toe. Vlaanderen opnemen in de ranking past in hun ogen niet in dat plaatje.

Voorlopig moeten we het op internationaal-comparatief vlak, wat de regionale component betreft, stellen met een uiterst beperkt aantal individuele oefeningen waarin de cijfers die België positioneren in de internationale governance rankings worden herberekend voor Vlaanderen. Zo rekenden de Leuvense onderzoekers Moesen, De Witte en Bowen de cijfers van het World Economic Forum (de Global Competitiveness Index, toen nog Growth Competitiveness Index) voor 2005 om naar een Vlaamse score (Bowen et al., 2007). Dit gebeurde op initiatief van toenmalig Minister Moerman en werd uitgevoerd binnen het Steunpunt Ondernemerschap, Ondernemingen en Innovatie.

De onderzoekers kwamen uit op een 19e plaats voor Vlaanderen in vergelijking met de 31e plaats die België in die ranking behaalde. De conclusies voor Vlaanderen waren echter niet waterdicht, ze berustten in grote mate op *proxy data*, of bij afwezigheid daarvan zelfs gewoon op de statistieken voor België. Aan het eind van de rit werd er dan ook besloten om de studie niet te continueren. De energie en kost die het vergde om de herberekening te maken voor Vlaanderen droeg in belangrijke mate bij tot deze beslissing. Hier kwam nog bij dat de ranking van het World Economic Forum de laatste jaren grootschalige methodologische veranderingen heeft doorgemaakt. Zo werden er in het jaar van onderzoek (2005) slechts 35 indicatoren betrokken in de meting. De laatste versie van de World Competitiveness Index bevat maar liefst 116 variabelen, wat een veelvoud aan herberekeningswerk en een stijging van de datacomplexiteit met zich mee zou brengen.

Een andere niet onbelangrijke vaststelling die in onze gesprekken met betrokkenen op Vlaams niveau naar boven kwam, is dat het gebruik van internationale rangschikkingen van publieke sectoren in ons land momenteel weinig gecoördineerd verloopt. Vanuit diverse hoeken kwam het geluid dat men niet wist welke ranking te gebruiken als maatstaf. Dit levert een ongecontroleerd *'ranking shopping'* op, waarbij niet duidelijk is wie wat meet en welke waarde er aan gehecht moet worden. Er worden aan de diverse diensten ook vanuit de gespecialiseerde overheidsinstellingen geen richtlijnen meegegeven over de omgang met internationaal-comparatieve

data van de overheidsprestaties. Dit levert een moeilijke evenwichtsoefening op voor diverse instanties binnen de publieke sector die bij de publicatie van negatieve rankings voor ons land, steevast in het defensief worden gedwongen door de media.

Niet alleen ontbreekt het aan een gecoördineerde opvolging van de governance data, diverse producenten duiden er op dat ons land achterop hinkt op het vlak van gebruik van internationale rangschikkingen van publieke sectoren. In vergelijking met ons land is er vooral in de Angelsaksische landen, maar ook in onze buurlanden een veel grotere afzet binnen de overheidssector van de periodieke vergelijkende data over de prestaties van de publieke sector.

> 7.2. Toekomstig onderzoek

Deze laatste paragraaf bewaren we voor een korte vooruitblik op de toekomst. We hebben in het onderzoek 'Internationale vergelijkingen vergeleken: Een meta-analyse van rangschikkingen van publieke sectoren' geïllustreerd dat de huidige governance rankings nog veel gebreken vertonen. Dit beperkt de bruikbaarheid van dergelijke beleidsinstrumenten voor de overheden van de verschillende gerangschikte landen. In de vorige paragraaf hebben we enkele aanbevelingen gepresenteerd ten aanzien van de gebruikers en de producenten van internationale rangschikkingen van publieke sectoren die moeten leiden tot een verbeterd gebruik van de rankings.

In deze slotparagraaf trachten we een aanzet tot verbetering van de rankings te geven. In eerste instantie herhalen we enkele suggesties uit de internationale litteratuur, die er voor moeten zorgen dat de rankings van morgen een betere kwaliteit vertonen. We schetsen ook kort het initiatief dat op til staat in de Organisation for Economic Co-operation and Development. Onder het project 'Government at a glance' zal deze organisatie in de komende jaren trachten een nieuwe dynamiek in het vergelijken van publieke sectoren op gang te brengen.

In tweede instantie doen we ook enkele suggesties naar de wetenschap toe. We geven aan waar er door de academische en onderzoekswereld nog studie kan verricht worden. Op die manier trachten we de wetenschap verder aan te moedigen mee de weg van de vooruitgang te plaveien op het vlak van het vergelijken van publieke sectoren.

> 7.2.1. De governance rankings van morgen

De internationale rangschikkingen van publieke sectoren van vandaag hebben het moeilijk hun rol van beleidsondersteunend instrument te spelen. Niet alleen zijn de vergelijkingen in veel gevallen hoogst precair wegens de kwalitatieve gebreken, ze zijn ook in veel gevallen niet concreet genoeg om er lessen uit te trekken voor een beleidsmaker. Enkele auteurs zien bijgevolg eerder een rol als *traffic light* weggelegd voor de governance rankings (Boyle, 2007; Kauffman et al., 2000). De rankings wijzen dan voornamelijk de landen aan die met (ernstige) governance-problemen kampen. Enkele auteurs suggereren dat dit proces bovendien beter zou

verlopen als men het ranking-idee zou verlaten, en naar brede categoriseringen zou overstappen. "*The categorical approach points to the need for reform without encouraging fruitless debate about precise scores or rankings assigned to particular countries*. (Kauffman & Kraay, 2007b)"

Om in een volgende stap ook de concrete pijnpunten te identificeren, zijn de metingen veelal te vaag op indicatorenniveau. Zoals de auteurs van de Worldwide Governance Indicators moeten erkennen: "*They are a blunt tool to use in formulating policy advice*." En is nu net niet dát aspect zo belangrijk voor overheden in hun strijd om verbeterde overheidsprestaties en hogere waarderingen in de gerelateerde rankings? De governance rankings van zouden moeten meer concrete aanknopingspunten voor beleidsvoering moeten aanbieden. Het is noodzakelijk meer *actionable indicators* in de rankings op te nemen, zodat overheden niet enkel de blaam van de rankings over zich krijgen, maar er ook worden door verder geholpen.

Een voorbeeld van een meetsysteem waarin concrete aspecten van de overheidswerking in beeld worden gebracht, is bijvoorbeeld de *Doing Business Database* van de World Bank, waarbij onder andere items als *'Paying taxes'*, *'Dealing with licenses'* en *'Starting a business'* aan de hand van concrete indicatoren worden gemeten. In plaats van vage governance-concepten, meten ze *real life issues* die relevant zijn voor burgers en andere actoren die in aanraking komen met de publieke sector. Rankings kunnen immers pas ten volle van nut zijn, aldus Manning (2006), indien ze informatie bevatten over diensten waarmee de burger persoonlijk contact heeft.

Daarnaast komt uit ons onderzoek duidelijk de noodzaak naar voren om op internationale schaal meer consistente data te vergaren. Hoe verder men staat in het theoretische debat over governance-metingen, hoe minder rangschikkingen van het ene jaar op het andere ineens een andere methodologie zullen hanteren. Hoe verder de nationale datagaring op mondiaal vlak staat, hoe meer continuïteit de landensamples zullen vertonen. Beide aspecten zijn momenteel nog te volatiel om van een goede vergelijkbaarheid in de governance rankings te kunnen spreken.

Onder welke vorm dan ook, de meeste critici zijn het er over eens dat de governance rankings van morgen aan hun kwaliteit moeten sleutelen. Vooral

het gebrek aan transparantie wordt bekritiseerd. Ook het validiteitsprobleem wordt aangekaart; het is in veel gevallen haast onmogelijk geworden te begrijpen wat de meting nu eigenlijk gemeten heeft. Voorts moeten er in de toekomst meer specifieke, tastbare en dagdagelijks-relevante aspecten van de publieke sector aan bod komen in de governance rankings.

Hood, Beeston en Dixon (2007) stellen dat we op dit moment met een chaotische wirwar aan internationale rangschikkingen van publieke sectoren zitten. Ze typeren de huidige situatie als een "*relatively unregulated world in which there are no formally recognized common standards of quality, and low barriers to entry into the rankings market by new players*". Maar de mechanismen van vraag en aanbod zorgen ervoor dat de rankings niet meer weg te denken zijn. *Rankings are here to stay*!

Hoe moet het dan verder? Deze auteurs zien drie potentiële toekomstscenario's. Ten eerste zou er zich een Darwiniaanse *'survival of the fittest*' kunnen voordoen. Dit zou een reductie in het aantal governance rankings en mogelijks een stijging van de kwaliteit met zich mee kunnen brengen. De Wet van Gresham stelt echter dat een dergelijke natuurlijke selectie ook een negatief effect op de kwaliteit van de rankings tot gevolg kan hebben. Zeker met de media, die meer en meer op zoek zijn naar '*quick and dirty rankings*', de budgettaire druk op internationale organisaties en de hoge kosten die nauwgezette vergelijkingen van publieke sectoren met zich meedragen (Hood et al., 2007), is dit geen garantie voor beterschap.

Ten tweede kan er zich een *'data club process*' voordoen, waarbij verschillende overheden overeenkomen zeer specifieke informatie betreffende de publieke sector te delen, op voorwaarde dat ze niet gepubliceerd worden. Het *Railway Technology Strategy Centre*, dat een dergelijk systeem voor informatie op het vlak van openbaar vervoer (voornamelijk metro) in verschillende landen hanteert, geldt hierbij op kleinere schaal als voorbeeld (Thomson, 2005a).

Een derde hypothese stelt dat er een systeem in het leven zou kunnen geroepen worden waarbij er een reeks standaarden wordt opgesteld waaraan de governance rankings moeten voldoen. Een voorbeeld hiervan is een internationaal aanvaard systeem van *'rank the rankers*', waarvan deze studie een illustratie vormt. Hierbij is vooral van belang dat een dergelijk

evaluatiesysteem onafhankelijk is en over de nodige legitimiteit kan genieten.

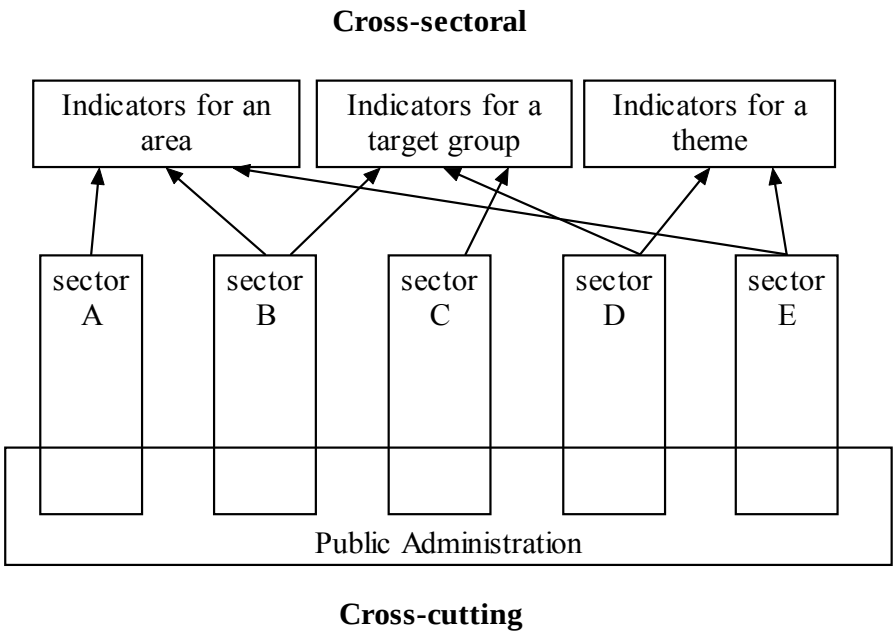
Naast de bestaande internationale rangschikkingen van publieke sectoren, staan er ook enkele nieuwe initiatieven op het vlak van internationale vergelijkingen van publieke sectoren op til. Er wordt daarbij vooral uitgekeken naar het OESO-project *'Government at a glance'* dat vermoedelijk in 2009 van start zal gaan.

De Organisation for Economic Co-operation and Development zal in dit project twee belangrijke keuzes maken. Ten eerste richt *Government at a glance* zich uniek op de prestaties van governance-processen. Men zal dus geen andere sectoren (onderwijs, gezondheidszorg, etc.) betrekken in de analyse. Deze governance-processen betreffen bijvoorbeeld *budgeting*, HRM, regulering, administratieve vereenvoudiging en informatisering.

Deze keuze heeft belangrijke gevolgen voor het concept *performance*. Governance-processen gaan meestal vooraf aan andere overheidsprocessen. Een goed HRM-beleid kan ervoor zorgen dat pakweg het onderwijsdepartement goed functioneert. De outcome van het HRM-beleid is dan ook vooral dat andere departementen competent personeel kunnen aanwerven binnen een redelijke termijn. Hoewel de burger/klant deze outcome niet direct voelt, is het toch belangrijk om de uiteindelijke doelstelling van goed onderwijs te realiseren.

Het gevolg is dat de outcomes van bestuur (*public administration* in figuur 8) doorheen beleidssectoren snijden. Dit is fundamenteel verschillend van andere indicatorensets die indicatoren van verschillende sectoren bijeenbrengen. Zo brengt *'Society at a glance'* van de Organisation for Economic Co-operation and Development indicatoren samen uit verschillende sectoren en dit per land (*area*). *'Pensions at a glance*' gaat dan weer dieper in op een doelgroep (*target group*) en *'Agriculture at a glance*' op een thema (*theme*). De meting vindt plaats per sector en de datasets brengen gegevens samen. De meting van de outcomes van bestuur daarentegen vindt plaats binnen alle sectoren, en is daarom een stuk complexer.

Figuur 8: OESO-aanpak 'Government at a glance'



(Van Dooren, 2008)

Een tweede keuze die zal gemaakt worden, is om een aantal systeembrede indicatoren te ontwikkelen naast de indicatoren op het niveau van concrete governance-processen zoals HRM en begroting. *Trust in government* is hier een voorbeeld van. Men ziet *trust* echter niet langer als een proxy van objectieve *performance*. Men interpreteert *trust* als een meting van de attitude van de burger ten aanzien van de overheid. *Fiscal sustainability* wordt voorgesteld als nog een andere systeembrede indicator.

In haar stappenplan heeft de OESO ervoor gekozen in eerste instantie naar en strakke definitie van '*Government*' te gaan. Hierbij is het vooral zaak overeenkomsten met het *'System of National Accounts'* te maken. Vervolgens zullen er strikt beschrijvende data worden opgesteld voor elk land. In derde instantie wil de OESO de vraag en aanbod van betrouwbare data verhogen. Het proces wordt afgesloten met het opstellen van een stel *evaluative indicators* op het niveau van de *Public Administration* (Manning et al., 2006).

> *7.2.2. Voortzetting wetenschappelijk onderzoek*

Ook voor de wetenschap zijn er nog talrijke uitdagingen in het debat rond de internationale vergelijkingen van publieke sectoren. Zolang de kwaliteit van de governance rankings niet voldoende is om ze als betrouwbare beleidsinformatie te gebruiken bij beleidsprocessen, moet de wetenschap in eerste instantie optreden als *monitoring institution*. We herhalen hier de taakomschrijving die de bekende auteur Thomson voor de onderzoekswereld weggelegd ziet: "*Academics could act as a 'parallel source of intelligence' on international rankings, serving to expose those of poorer quality and to act as a stabilizing influence on the 'media frenzy' that attended new publications*" (Thomson, 2005b).

Het is onze overtuiging dat de wetenschappelijke wereld deze rol ter harte moet nemen, en op die manier druk kan uitoefenen op de producenten van governance rankings. Het verleden heeft uitgewezen dat internationale organisaties niet doof zijn gebleven voor de kritiek in de internationale literatuur en reeds op enkele kwaliteitsdomeinen progressie hebben gemaakt.

Vervolgens is het ook belangrijk dat de wetenschap een rol op zich neemt in het uitwerken van uniforme definities. Zo lang er zelfs tussen experten grote onenigheid bestaat over dergelijke kwesties, is het haast onmogelijk om tot een internationaal aanvaarde governance ranking te komen. Zoals aangegeven in dit rapport, is dit geen simpele opdracht. Ze houdt onder andere in dat er overeenstemming gezocht wordt over de taken van een 'goed presterende overheid'. Die vraag kadert in het kerntakendebat, dat echter vooral ook in politieke kringen moet gevoerd worden.

Pas als dit plaatje is ingevuld, kan er ook gekeken worden naar de meetsystemen en procedures die in acht moeten genomen worden om de publieke sector te meten, te vergelijken en te rangschikken. Zoals geïllustreerd spelen deze schijnbaar technische kwesties immers een niet onbelangrijke rol in de data. De wetenschap moet helpen uitmaken hoe de ideale meting van een concept zoals 'de efficiënte overheid' er qua structuur en meetprocedures uit moet zien.

8. Bijlage

Lijst geïnterviewde personen in het kader van 'Internationale vergelijkingen vergeleken'

Naam	Organisatie	Functie
Hatipoglu, Saruhan	Business Environment Risk Intelligence	Managing Director & Acting CEO
Chabane, Nadia	Economist Intelligence Unit	Regional Sales Executive - Benelux
Alasdair, Ross	Economist Intelligence Unit	Director, Wire Services
Rosselet, Suzanne	IMD World Competitiveness Centre	Deputy Director
Walsh, Mary Lou	PRS Group Inc.	President
Gerken, Thomas	PRS Group Inc.	
Geiger, Thierry	WEF - Global Competitiveness Network	Economist
Grosso, Guillaume	African Development Bank Group	
Garrity, Colum	World Bank Institute	Consultant Public Sector Governance
Boucké, Bart	Flanders Investment and Trade	Verantwoordelijke Beheersovereenkomst - Staf v/d gedelegeerd bestuurder
Callens, Marc	Studiedienst van de Vlaamse Regering	Senior onderzoeker
De Cock, Jo	Departement Internationaal Vlaanderen	Beleidsmedewerker Afdeling Internationaal Ondernemen
De Mulder, Jan	Departement Bestuurszaken - Vlaamse overheid	Beleidsmedewerker Coördinerende Dienst
Martens, Gijs	Departement Bestuurszaken - Vlaamse overheid	Beleidsadviseur Afdeling beleid
Van Cleynenbreugel, Jos	Departement Internationaal Vlaanderen	Directeur Cel Financiën & Begroting

9. Bibliografie

1. Alfonso, A., Schuknecht, L., & Tanzi, V. (2005). Public sector efficiency: An international Comparison. *Pubk, 123,* 321-347.

2. Andersen, C. & Tverdova, Y. (2003). Corruption, Political Allegiances, and Attitudes Toward Government in Contemporary Democracies. *American Journal of Political Science, 47,* 91-109.

3. Arndt, C. (2007). *The Politics of Governance Ratings* Paper prepared for the workshop "Ranking and rating public services" of the International Public Management Network in Oxford, 7-9 augustus 2007: Maastricht Graduate School of Governance and Harvard University.

4. Arndt, C. & Oman, C. (2006). *Uses and abuses of Governance Indicators* Development Centre of the Organisation for Economic Co-operation and Development.

5. Bertelsmann, S. (2005). *Bertelsmann Transformation Index 2003*. Gütersloh: Bertelsmann Stiftung.

6. Besançon, M. (2003). *Good Governance Rankings: The Art of Measurement*. Cambridge: World Peace Foundation.

7. Bevan, G. & Hood, C. (2006). What's measured is what matters: targets and gaming in the English Public Health Care System. *Public administration, 84,* 517-538.

8. Billiet, J. (1993). *Ondanks beperkt zicht*. Brussel: VUBPress.

9. Blank, J. L. T. (2000). *Public Provision and Performance: Contributions from Efficiency and Productivity Measurement*. Elsevier.

10. Bollen, K. A. (1989). *Structural equations with latent variables*. New York: John Wiley & Sons.

11. Bouckaert, G. (1992). Productivity Analysis in the Public Sector: the Case of the Fire Service. *International Review of Administrative Sciences, 58,* 175.

12. Bouckaert, G. (1993). Efficiency Measurement from a Management Perspective: A Case of the Civil Registry Office in Flanders. *International Review of Administrative Sciences, 59,* 11.

13. Bouckaert, G. & Auwers, T. (1999). *Prestaties meten in de overheid*. Brugge: Die Keure.

14. Bowen, H., De Witte, K., Chavez, J. G., & Moesen, W. (2007). *The competitiveness of Flanders in the ranking of the World Economic Forum*.

15. Boyle, R. (2007). *Comparing Public Administrations: An assessment of the quality and efficiency of public administration in Ireland compared with selected European and OECD countries* Dublin: Institute of Public Administration.

16. Boyne, G. A. & Walker, R. M. (2005). *Introducing the "determinants of performance in Public Organizations" Symposium* Cardiff University.

17. Brutsaert, S. (2006a). Rol van de overheid ondermaats: 10 miljard euro inefficiëntie? *FORWARD,* 30-35.

18. Brutsaert, Sofie (2006b). Rol overheid ondermaats in België: 10 miljard euro inefficiëntie? *FORWARD*.

19. Burnside, C. & Dollar, D. (1997). *Aid, Policies, and Growth* (Rep. No. 569252). World Bank.

20. Callens, M. (2007). *Efficiëntie en effectiviteit in de publieke sector - Draft nota* Brussel: Studiedienst Vlaamse Regering.

21. Cherchye, L., Moesen, W., Rogge, N., & Puyenbroeck, T. V. (2007). An Introduction to 'Benefit of the Doubt' Composite Indicators. *Social Indicators Research, 82,* 111-145.

22. De Batselier, N. (2007). De efficiëntie van de overheid in België. In *MOVI-Colloquium Vlaanderen "anders" bekeken* (pp. 14).

23. De Groot, H. (2004). Waarom Benchmarking? *Waarom benchmarking?: Vergelijken van prestaties als managementinstrument* (.

24. De Neufville, J. L. (1975). *Social Indicators and public policy: Interactive processes of design and application*. New York: Elsevier Scientific Publishing Company.

25. De Witte, K. & Moesen, W. (2007). *The Growth Competitiveness Report - Information Sheet*.

26. Demidov, B. (2006). *Corruption Perceptions Indices - a True instrument or an Unreliable Mirror?*

27. Dill, D. & Soo, M. (2005). Academic quality, league tables and public policy: A cross-national analysis of university ranking systems. *Higher Education*.

28. Donahue, A. K., Selden, S. C., & Ingraham, P. W. (2000). Measuring Government Management Capacity: A Comparative Analysis of City Human Resources Management Systems. *Journal of Public Administration and Research Theory, 10,* 381-412.

29. Fearon, J. & Laitin, D. (2003). Ethnicity, Insurgency, and Civil War. *American Political Science Review, 97,* 75-90.

30. Frey, B. & Stutzer, A. (2000). Happiness, Economy and Institutions. *Ecnomic Journal, 110,* 918-938.

31. Galtung, F. (2005). Measuring the immeasurable: Boundaries and Functions of (Macro) Corruption Indices. (Ashgate (UK): Forthcoming.

32. Girvan, N. (2002). Problems with UNDP Governance Indicators: The Greater Caribbean This Week.

33. Gormley, W. & Weimer, D. L. (1999). *Organizational Report Cards*. Londen: Harvard University Press.

34. Grindle, M. S. (2007). Good enough governance revisited. *Development Policy Review, 25,* 553-574.

35. Handler, H., Koebel, B., Reiss, P., & Schratzenstaller, M. (2005). *The size and performance of public sector activities in Europe*.

36. Heliwell, J. (2003). How's life? Combininf individual and national variables to explain subjective well-being. *Economics of Modeling, 20,* 331-360.

37. Hood, C. (1991). A public management for all seasons. *Public administration, 69,* 3-19.

38. Hood, C. (2007). Public Service Management by Numbers: Why Does it Vary? Where Has it Come From? What Are the Gaps and the Puzzles? *Public Money & Management, 27,* 95-102.

39. Hood, C. & Beeston, C. (2005). How Does Britain Rank and How Do We Know? - International Rankings of Public Service Performance op de conferentie 'Where Does Britain Rank? International Public Service Rankings'. Londen.

40. Hood, C., Beeston, C., & Dixon, R. (2007). *Rating the rankings: assessing international rankings of public service performance* Paper prepared for the workshop "Ranking and rating public services" of the International Public Management Network in Oxford, 7-9 augustus 2007: University of Oxford.

41. Kauffman, D., Kraay, A., & Zoido-Lobatón, P. (2000). Governance matters: from measurement to action. *Finance and Development: A quarterly magazine of the IMF, 37*.

42. Kauffman, D. & Kraay, A. (2002). *Growth without governance* World Bank.

43. Kauffman, D. & Kraay, A. (2007a). *Measuring Governance Worldwide: Possibilities, Pitfalls, and Applications of the Woirldwide Governance Indicators (WGI) & others*.

44. Kauffman, D. & Kraay, A. (2007b). *On Measuring Governance*.

45. Kauffman, D., Kraay, A., & Mastruzzi, M. (2006). *Governance matters V: Governance Indicators for 1996-2005*.

46. Kauffman, D., Kraay, A., & Mastruzzi, M. (2007a). *Governance matters VI: Aggregate and individual Governance Indicators for 1996-2006* World Bank Policy Research Working Paper 4280.

47. Kauffman, D., Kraay, A., & Mastruzzi, M. (2007b). *The World Wide Governance Indicators Project: Answering the Critics*.

48. Kaufmann, D., Kraay, A., & Zoido-Lobatón, P. (1999). *Governance Matters* (Rep. No. Policy Research Working Paper 2196). The World Bank Development Research Group Macroeconomics and Growth and World Bank Institute Governance, Regulation and Finance.

49. Killias, M. (2005). Comparing Crime and Police Performance op de conferentie 'Where Does Britain Rank? International Public Service Rankings'. Londen.

50. Knack, S. & Kugler, M. (2002). *constructing an index of objective indicators of good governance* PREM Public Sector Group, World Bank.

51. Knack, S. (2006). *Measuring corruption in Eastern Europe and Central Asia: a critique of the cross-country indicators* (Rep. No. Policy Research Working Paper 3968). The World Bank Development Research Group Macroeconomics and Growth and World Bank Institute Governance, Regulation and Finance.

52. Knack, S., Kugler, M., & Manning, N. (2003a). *Second-generation governance indicators*.

53. Knack, S., Kugler, M., & Manning, N. (2003b). Second-generation governance indicators. *Internatinal Review of Administrative Sciences,* 345-364.

54. Knack, S. & Keefer, P. (1997). Does Social Capital Have An Economic Payoff? A Cross-Country Investigation. *Quarterly Journal of Economics, 112,* 1251-1288.

55. Knack, S. & Manning, N. (2000). *Towards more operationally relevant indicators of governance* (Rep. No. PREM note 49). Washington, DC: World Bank.

56. Knack, S. & Mannig, N. (2000). *Towards consensus on governance indicators: selecting public management and broader governance indicators*.

57. Kuha, J. & Gaskell, G. (2005). Methodological issues in constructing league tables op de conferentie 'Where Does Britain Rank? International Public Service Rankings'. Londen.

58. Kuhry, B. & Van der Torre, A. (2002). *De vierde sector*. Den Haag: Sociaal-Cultureel Planbureau.

59. Kuhry, B. (2004). *Public Sector Performance, an international comparison of Education, Health Care, Law and Order and Public Administration in 29 western countries* Den Haag: Social and Cultural Planning Office.

60. Kurtz, M. J. & Schrank, A. (2005). *Growth and governance: models, measures and mechanisms*.

61. La Porta, R., Lopez-de-Silanos, F., Shleifer, A., & & Vishny, R. (1999). The quality of government. *The journal of law, economics and organization, 15,* 222-279.

62. Langouche, G. Sterktes en zwaktes van de belangrijke rangschikkingen: De rankers gerangschikt. *Campus Krant,* (in press).

63. Majone, G. (1996). *Regulating Europe* London: Routledge.

64. Manning, N., Kraan, D., & Malinska, J. (2006). *How and why should Government activity be measured in "Government at a Glance"?*

65. Ministère des Affaires étrangères et Européennes Française *Aid efficacy: Governance Indicators*.

66. Moesen, W. & Cousserier, J. (2006). *De kwaliteit van de publieke sector.*

67. Munck, G. L. & Verkuilen, J. (2002). Conceptualizing and measuring democracy: evaluating alternative indices. *Comparative Political Studies, 35,* 5-34.

68. O'Mahony, M. & Stevens, P. (2002). Measuring International Comparative Performance in the Provisions of Public Services: A Review.

69. O'Mahony, M. & Stevens, P. (2004). *International Compaisons of Performance in the Provisions of Public Services: Outcome based measures for Education*.

70. Orkin, M. (2002). *Breaking the impasse in measuring good governance and democracy: Challenges to and from NEPAD* BMZ/GTZ/IDEA conference in Berlin.

71. Parks, B. P. (1984). Linking objective and subjective measures of performance. *Public Administration Review,* 118-127.

72. Peeters, K. (2007). *Septemberverklaring van de Vlaamse regering*.

73. Pollitt, C. & Bouckaert, G. (2004a). *Public management reform. A comparative analysis.* New York: Oxford University Press.

74. Pollitt, C. & Bouckaert, G. (2004b). *Public management reform. A comparative analysis. (Second edition)*. New York: Oxford University Press.

75. Pollitt, C. (2008). Simply the best? The international benchmarking of reform and good governance. *Public sector administrative reform and the challenges of effective change* (McGill-Queens University Press.

76. Public Governance and Territorial Development Directorate OECD (2005). *Management in government: feasibility report on the development of comparative data* Den Haag: GOV/PGC(2005)10.

77. Rose, R. & Shin, D. C. (2001). Democratization Backward: The Problem of Third Wave Democracies. *Britisch Journal of Political Science, 31,* 331-354.

78. Rothstein, B. & Teorell, J. (2005). *What is the quality of government? A theory of impartial political institutions* (Rep. No. 6). Göteborg.

79. Rutgers, M. R. (2004). Comparative Public Administration: Navigating Scylla and Charybdis - Global Comparison as a Translation Problem. *Administrative Theory and Praxis, 26,* 150-168.

80. Saltelli, A. (2005). *Composite Indicators for advocacy on a KEI Workshop in Maastricht* European Commission - Joint Research Centre.

81. Schmitz, T. (2007). *Governance criteria and political choices: The good, the bad and the governance*.

82. Schneider, F. (2002). *The size and development of the shadow economies of 22 transition and 21 OECD countries* (Rep. No. 514). Bonn: IZA.

83. Sociaal en Cultureel Planbureau (2007). *Publieke prestaties in perspectief* Den Haag: Sociaal en Cultureel Planbureau.

84. Street, A. (2005). Comparing Health Performance and the 2000 WHO health rankings experience op de conferentie 'Where Does Britain Rank? International Public Service Rankings'. Londen.

85. Thomas, M. A. (2007). *What do the worldwide governance indicators measure?*

86. Thomson, W. (2005a). Management in Government: Development of OECD Comparative Country Data op de conferentie 'Where Does Britain Rank? International Public Service Rankings'. Londen.

87. Thomson, W. (2005b). The Way Forward: Using and Developing International Comparisons of Public Service Performance op de conferentie 'Where Does Britain Rank? International Public Service Rankings'. Londen.

88. Tiihonen, S. (2004). *From governing tot governance*. Tampere: Tampere University Press.

89. Uitterhoeve, W. (2000). *Sociaal en cultureel rapport 2000, Nederland in Europa* Den Haag: Sociaal en Cultureel Planbureau.

90. United Nations (1998). *Cooperating for Development, Annual Report of the Secretary-General on the Work of the Organization* (Rep. No. Document A/53/1). New York.

91. Usher, A. & Savino, M. (2006). *A World of Difference: A Global Survey of University League Tables* Toronto: Educational Policy Institute.

92. Van de Walle, S. (2005). *Measuring bureaucratic quality in governance indicators*.

93. Van de Walle, S. (2006a). The State of the World's Bureaucracies. *Journal of Comparative Policy Analysis: Research and Practice, 8,* 437-448.

94. Van de Walle, S. (2007). *International Comparisons of Public Sector Performance: How to Move Ahaed?*

95. Van de Walle, S., Sterck, M., Van Dooren, W., & Bouckaert, G. (2004). *What you see is not necessarily what you get, Een verkenning van de mogelijkheden en moeilijkheden van internationale vergelijkingen van publieke sectoren op basis van indicatoren* Leuven: Instituut voor de Overheid.

96. Van de Walle, S. (2006b). Een grote, dure en niet efficiënte overheid? Indicatoren voor het vergelijken van publieke sectoren in de EU. *Vlaams Tijdschrift voor Overheidsmanagement (VTOM), 11,* -27.

97. Van Dooren, W. (2008). *Outcome indicators for Government at a Glance* Paris.

98. Van Dooren, W., Manning, N., Malinska, J., Sterck, M., & Bouckaert, G. (2007). *Issues in Output Measurement for "Governement at a Glance" OECD GOV Technical Paper 2*.

99. Van Dooren, W., Thijs, N., & Sterck, M. (2006). Achter de cijfers: Aandachtspunten bij het vergelijken van overheden. *Over Werk, Tijdschrift van het Steunpunt WAV, 1-2*.

100. Van dooren, W. (2006). *Performance measurement in the Flemish public sector: A supply and demand approach*. Leuven: Katholieke Universiteit.

101. Van Roosbroek, S. & Van Dooren, W. (2006). *Meten is weten!/? België/Vlaanderen in internationale indicatoren en indices* Leuven: Instituut voor de Overheid.

102. Vervenne, W. Belgische overheid is zeer inefficiënt. *De Tijd,* (in press).

103. Vlaamse regering (2006). *Vlaanderen in actie!* Brussel: Bart De Moor.

104. Working Group 'Assessment of quality in statistics' - Eurostat (2003). *Methodological documents: definition of quality in statistics* Luxembourg.

105. Yang, K. & Holzer, M. (2006). The performance-trust link: implications for performance measurement. *Public Administration Review, 66,* -114.

106. Zakaria, F. (2003). *The future of freedom. Illiberal democracy at home and abroad.* New York: W.W. Norton.