

# De invoeringstoets geborgd

## Randvoorwaarden voor de benutting van een nieuw evaluatie- instrument

---

*Raf Perrée*

*19-04-2023*

*MSc Publiek Management*

*Universiteit Utrecht*

# Colofon

19-04-2023

Student: R. (Raf) Perrée

Studentnummer: 6182259

1e begeleider: prof. dr Mirko Noordegraaf

2e begeleider: dr Karin Geuijen

Scriptie ter verkrijging van de graad Master of Science

Publiek Management

Departement Bestuurs- en Organisationswetenschap

Faculteit Recht, Economie, Bestuur- en Organisatie

Universiteit Utrecht

## Voorwoord

Voor u ligt de masterscriptie 'De invoeringstoets geborgd: Randvoorwaarden voor de benutting van een nieuw evaluatie-instrument'. In deze scriptie staat centraal onder welke randvoorwaarden het nieuwe evaluatie-instrument de invoeringstoets structureel kan worden benut. De context van de Rijksoverheid is enorm dynamisch en complex waardoor veel factoren van invloed zijn op hoe de resultaten van evaluaties opvolging krijgen. De vraag is dus hoe de invoeringstoets ingekleurd kan worden zodat deze rijksbreed ook daadwerkelijk zijn doel bereikt.

Tijdens mijn master Publiek Management heb ik middels de vele interessante gastcolleges en werkbezoeken kennis mogen maken met alle hoeken van de Rijksoverheid. Ik heb hierdoor in de eerste plaats veel geleerd over hoe zij georganiseerd is, maar vooral ook over hoe zij opereert. Eén gastcollege springt hier bovenuit: namelijk dat van secretaris-generaal van het ministerie van Binnenlandse Zaken en Koninkrijksrelaties: Maarten Schurink. Kort daarvoor zijn wij op werkbezoek geweest bij de Sociale Verzekeringsbank (SVB) waar Maarten Schurink de 'Garage de Bedoeling' geïntroduceerd heeft. Hiermee heeft hij bewerkstelligd dat de cultuur bij de SVB niet langer is om te kijken naar de wet of regeling zelf, maar naar de bedoeling ervan. Hierdoor heeft de dienstverlening van de organisatie een enorme kwaliteitsimpuls gekregen die voor het leven van de (toch vaak kwetsbare) burger enorme positieve gevolgen kan hebben. Het was dan ook ontzettend inspirerend om te horen hoe Maarten hier tegenaan keek. Dergelijke succesverhalen maken dat ik Publiek Management interessant vind. Door goed management kan een enorme impact worden gemaakt op de burger en maatschappij.

Sinds de start van mijn bachelorscriptie Bestuurs- en organisatiewetenschap houd ik mij bezig met de invoeringstoets. Destijds lag de focus op onder welke randvoorwaarden het instrument geïmplementeerd kon worden bij de Belastingdienst. Toen Maarten Schurink in zijn verhaal vertelde over het werken volgens de bedoeling van de wet heb ik hem de vraag gesteld hoe hij aan keek tegen de komst van de invoeringstoets: een evaluatie die focust op of de bedoeling van een wet gerealiseerd wordt. Hij antwoordde dat hij één van de aanjagers van de komst van het instrument was en graag mijn scriptie zou willen lezen. Aangezien ik de invoeringstoets nog steeds een zeer interessant idee vond heb ik de scriptie naar hem toegestuurd met de vraag of ik bij de implementatie van het instrument iets kon betekenen. Een aantal maanden later ben ik begonnen als stagiair bij het ministerie van Justitie en Veiligheid en tevens gestart met het schrijven van mijn volgende scriptie over de invoeringstoets!

Het schrijven van de scriptie is een proces geweest van vallen en opstaan. Niet alleen bleek het teveel gevraagd om de stage te combineren met het schrijven van de scriptie, maar was het feit dat de invoeringstoets nog zeer nieuw was (/is) ook een grote barrière voor het empirische onderzoek. Door het kleine aantal uitgevoerd toetsen bleek het ontzettend lastig om relevante respondenten te vinden waardoor het onderzoek maar lastig van de grond kwam en het voor mij persoonlijk ook een mentale worsteling werd. Daarom wil ik ook graag mijn stagebegeleider Pim Beirnaert bedanken, die zelfs maanden na het afronden van de stage nog bereid was om mij verder te helpen en mij uiteindelijk ook aan voldoende respondenten heeft geholpen. Daarnaast zou ik ook graag Mirko willen bedanken die mij al die tijd geduldig heeft bijgestaan en zeker tegen het einde ontzettend heeft geholpen met alle feedback en zich extra heeft ingezet met meetings zodat ik het op tijd kon afkrijgen voor mijn nieuwe baan. Tot slot wil ik graag mijn ouders bedanken die mij niet alleen onvoorwaardelijk hebben gesteund, maar ook hebben geprikkeld om door te gaan toen het tegen zat.

Voor mij is het nu eindelijk tijd voor een nieuwe uitdaging, waarbij ik de ontwikkeling die ik gedurende dit proces heb doorgemaakt zeker ga koesteren.

Raf Perrée

19-04-2023

# Inhoudsopgave

Colofon.....	2
Voorwoord.....	3
Inhoudsopgave.....	4
Managementsamenvatting .....	7
1. Inleiding.....	9
1.1 Aanleiding.....	9
1.2 Stand van zaken implementatie invoeringstoets .....	9
1.3 De invoeringstoets in de beleidscyclus.....	10
1.4 Benutting van beleidsevaluaties.....	11
1.5 Benutting van de invoeringstoets geborgd .....	12
1.6 Probleem- en doelstelling.....	12
1.7 Vraagstelling .....	13
1.8 Onderzoekaanpak en theoretisch perspectief .....	14
1.9 Relevantie.....	14
1.10 leeswijzer .....	15
2. De invoeringstoets in context.....	16
2.1 Inleiding .....	16
2.2 De toeslagenaffaire ('Ongekend onrecht').....	16
2.3 Kabinetsformatie 2017 .....	18
2.4 Werk aan uitvoering .....	19
2.5 De Tijdelijke commissie Uitvoeringsorganisaties (TCU) .....	21
2.6 De nieuwe bestuurscultuur.....	22
2.7 De invoeringstoets .....	23

2.8 Conclusie.....	24
3. Benutting van beleidsevaluaties .....	26
3.1 Inleiding .....	26
3.2 Wat is de invoeringstoets en hoe is deze te positioneren binnen beleidsprocessen? .....	26
3.3 Wat is benutting van evaluatie en welke benaderingen voor benutting zijn te onderscheiden? .....	33
3.4 Welke randvoorwaarden voor het borgen van benutting van de invoeringstoets zijn te onderscheiden?.....	38
4. Het empirisch onderzoek.....	44
4.1 Inleiding .....	44
4.2 Kwalitatief onderzoek.....	44
4.3 Onderzoeksdesign.....	45
4.4 Casus- en respondentselectie.....	47
4.5 Operationalisatie.....	47
4.6 Dataverzameling en -analyse.....	49
4.7 Kwaliteit van het onderzoek .....	50
4.8 Conclusie.....	51
5. Resultaten .....	52
5.1 inleiding .....	52
5.2 Hoe past de invoeringstoets binnen de beleidscyclus van de Rijksoverheid?.....	52
5.3 Hoe en waar wordt de invoeringstoets benut binnen de Rijksoverheid? .....	55
5.4 In hoeverre en hoe wordt in de praktijk voldaan aan de randvoorwaarden voor het borgen van de benutting van de invoeringstoets? .....	58
5.4.1 Organische randvoorwaarden.....	58
5.4.2 Politieke randvoorwaarden.....	59
5.4.3 Rationele randvoorwaarden .....	60
5.5 Conclusie.....	63
6. Conclusies en aanbevelingen.....	65

6.1 Inleiding .....	65
6.2 Deelvragen beantwoord .....	65
6.3 Hoofdvraag beantwoord.....	68
6.4 Aanbevelingen .....	69
6.5 Tot slot.....	71
Bronnenlijst .....	72
Bijlagen .....	75

## Managementsamenvatting

Naar aanleiding van de toeslagenaffaire heeft het kabinet de komst van een invoeringstoets toegezegd. Dit is een nieuw-evaluatie instrument waarbij relatief vroeg na inwerkingtreding van wet- en regelgeving wordt gekeken naar de werking ervan in de praktijk. Het doel van het instrument is enerzijds om het kabinet beter in staat te stellen om in te spelen op signalen van burgers, bedrijven en organisaties over de werking van nieuwe regelgeving in de praktijk. Anderzijds moet het voor de uitvoering makkelijker worden om knelpunten uit de uitvoeringspraktijk terug te leggen op de beleidstafel. Het idee is goed, maar dan moet de toets wél benut worden. Uit het rapport 'Klem tussen balie en beleid' van de Tijdelijke commissie Uitvoeringsorganisaties (2021, p. 37) blijkt dat dit voor evaluaties bij de Rijksoverheid niet altijd het geval is. Dit onderzoek gaat daarom in op de vraag onder welke randvoorwaarden de benutting van de invoeringstoets geborgd kan worden.

Om deze vraag te beantwoorden is allereerst ingegaan op de context waarbinnen het idee van de invoeringstoets is ontstaan. Hiertoe bespreken we de rapporten over het programma 'Werk aan Uitvoering' (Heerschop & van der Vlist, 2020a; Heerschop & van der Vlist, 2020b), het rapport van de Tijdelijke commissie Uitvoeringsorganisaties 'Klem tussen Balie en Beleid' (Tweede Kamer, 2021) en de informateursrapporten van Tjeenk Willink uit 2017 en 2021 (Willink, 2017; Willink, 2021). Hieruit blijkt dat een aantal ontwikkelingen uit met name de jaren 10 ervoor heeft gezorgd dat de uitvoering enorm onder druk is komen te staan en deze tegelijkertijd uit het oog is geraakt door de politiek en de departementen. Toen dit na de toeslagenaffaire aan het licht kwam is onder andere de invoeringstoets toegezegd. Hiermee zou structureel geëvalueerd kunnen worden hoe regelgeving inklinkt in de praktijk en welke gevolgen het heeft voor de doelgroep en uitvoering.

Dit onderzoek gaat verder in op de theoretische achtergrond van de werking van de invoeringstoets en het concept 'benutting'. Te beginnen met de werking van de invoeringstoets. De literatuur leert dat met verschillende blikken naar de werking van beleidsprocessen kan worden gekeken: een rationeel, organisch en politiek perspectief. Om erachter te komen hoe de invoeringstoets werkt moet dus vanuit elk van deze perspectieven worden gekeken. Ook voor het concept benutting bestaan verschillende vormen die verband houden met deze perspectieven, namelijk: instrumenteel (rationeel), conceptueel (organisch), agenderend (politiek) en politiek-strategisch (politiek). Uit de officiële berichtgeving over de invoeringstoets blijkt dat de focus voornamelijk ligt op instrumentele en voor een deel ook agenderende doorwerking. De vraag is echter of dit ook de vormen zijn die het meest voorkomen. Tot slot zijn de artikelen van Bekkers, Fenger, Homburg & Putters (2004) en Johnson, Greenseid, Toal, King, Lawrenz & Volkov (2009) gebruikt om te komen tot een aantal factoren waarvan het aannemelijk is dat ze van invloed zijn op de benutting van de invoeringstoets.

In het empirische deel zijn interviews afgenomen met 14 respondenten die betrokken waren bij ofwel een invoeringstoets, ofwel een evaluatie met gelijke kenmerken, ofwel de implementatie van de invoeringstoets in het algemeen. Deze respondenten waren werkzaam bij verschillende uitvoeringsorganisaties en ministeries van de Rijksoverheid en konden vanuit hun ervaring en kennis veel zeggen over de werking van de invoeringstoets. Met deze interviews is onderzocht hoe de toets in de praktijk werkt, hoe de resultaten ervan doorwerken en benut worden en tot slot hoe aan de randvoorwaarden voor het borgen van de benutting van de invoeringstoets wordt voldaan.

Uit deze interviews bleek dat de toets niet altijd werkt zoals deze op papier staat omschreven. In plaats van dat die werking heel rationeel is zoals op papier staat blijkt deze vooral ook afhankelijk te zijn van de omstandigheden waarin deze werkt en de belangen die worden

nagestreefd door betrokkenen. Ook blijkt de toets niet alleen maar instrumenteel en agenderend door te werken zoals de doelen van de toets omschreven staan, maar regelmatig ook conceptueel en politiek-strategisch. Het bijzondere daarbij is dat de respondenten de toets, zodra daarom gevraagd werd, vooral rationeel en instrumenteel beschreven. Tussen de onderwerpen door vertelden zij juist over hoe de toets ook een organische en politieke werking heeft en ook op andere manieren doorwerkt. Hieruit blijkt een soort rationaliseringsreflex te werken. In een poging de toets te laten werken zoals op papier is bedacht, wordt de invulling ook zeer rationeel benaderd waardoor juist de invloed van omstandigheden en politieke belangen overschaduwd wordt. Om de toets te laten werken zoals beoogd zou juist de nadruk moeten liggen bij die omstandigheden en politieke belangen om deze te erkennen en hier mee om te kunnen gaan.

Tegelijkertijd wordt wel degelijk aandacht besteed aan het feit dat de invoeringstoets afhankelijk is van de context, door te kiezen voor een decentrale aanpak. De uitvoeringsorganisaties en ministeries zijn individueel aan het kijken hoe zij het beste invulling kunnen geven aan het instrument zodat dit voor de eigen situatie werkbaar is. Juist door de keuze voor deze aanpak blijft de toets echter zeer abstract en generiek waardoor deze universeel lastig te gebruiken is. De invoeringstoets krijgt moeilijk bekendheid en men is niet altijd voldoende overtuigd van het nut. Zeker zodra een wet er na veel tijd en energie eindelijk staat of wanneer veel onderhandeld is over de toets heeft men weinig behoefte om het onderwerp terug op tafel te leggen.

Om de invoeringstoets dus echt tot een succes te maken is het cruciaal dat de invoeringstoets verder ingekleurd wordt. Pas als men weet wat het instrument is, hoe deze werkt en wat deze toevoegt kan zij hier ook mee aan de slag en wordt ook de bekendheid gestimuleerd. Daarnaast moeten de omstandigheden en politieke belangen die van invloed zijn op de invoeringstoets juist geadresseerd worden zodat hier op ingespeeld kan worden en de benutting geoptimaliseerd wordt.



# 1. Inleiding

## 1.1 Aanleiding

In december 2020 werd het langverwachte rapport 'Ongekend onrecht' gepubliceerd, waarin de Parlementaire ondervragingscommissie Kinderopvangtoeslag (POK) verslag deed van de misstanden rondom de kinderopvangtoeslag (Tweede Kamer, 2020a). In het rapport wordt geconstateerd dat bij de uitvoering van de kinderopvangtoeslag grondbeginselen van de rechtsstaat zijn geschonden, niet alleen door de Belastingdienst/Toeslagen, maar ook door de wetgever en de rechtspraak (Tweede Kamer, 2020a, p. 7). De commissie constateert enerzijds een politieke behoefte om de uitvoering van de toeslagen efficiënt in te richten en anderzijds de politieke en maatschappelijke wens om fraude te voorkomen. Dit zou hebben geleid tot wet- en regelgeving en een uitvoering daarvan, die het niet of nauwelijks toeliet om de individuele situatie van mensen recht te doen, bijvoorbeeld als zij zonder kwade opzet een administratieve vergissing begingen. Het gevolg van de keiharde aanpak was dat enkele honderden ouders ten onrechte werden aangemerkt als opzettelijke fraudeurs, waardoor de kinderopvangtoeslag is stopgezet en teruggevorderd (Tweede Kamer, 2020a, p. 11). Dit had voor veel ouders enorme financiële, relationele en andere problemen als gevolg.

Een kleine maand later reageert minister-president Rutte met een Kamerbrief namens het kabinet op het rapport van de POK (Rutte, 2021). In deze Kamerbrief worden excuses aangeboden aan alle gedupeerde ouders en kinderen, maar wordt vooral ook ingegaan op wat er moet veranderen om een volgende toeslagenaffaire te voorkomen. Eén van de genoemde veranderingen is de komst van de invoeringstoets, waarmee wet- en regelgeving een jaar na start van de uitvoering nogmaals getoetst moet gaan worden (Rutte, 2021, p. 6). De invoeringstoets moet daarmee de tegenhanger van de reeds bestaande uitvoeringstoets worden, waarmee beleid juist getoetst wordt vóór de invoering. In de Kamerbrief wordt hierbij een aantal hoofdvragen genoemd die in de invoeringstoets gesteld moet gaan worden, namelijk: *werkt dit inderdaad zoals we bedoeld hadden? Kloppen de aannames uit de uitvoeringstoets nog steeds? Wat is de kwaliteit van de uitvoering? En zijn er misschien knelpunten of nieuwe inzichten waar we aan de voorkant niet aan gedacht hadden? Hoe pakt het uit voor de mensen waarvoor de wet bedoeld is?* (Rutte, 2021, p. 6). Dat de implementatie van de invoeringstoets voldoende prioriteit heeft lijkt bevestigd te worden in het coalitieakkoord van december 2021, waar de komst van de toets op pagina 1 wordt vermeld (VVD, D66, CDA & ChristenUnie, 2021). Belangrijk is echter niet alleen de implementatie van dit instrument, maar vooral de benutting van de resultaten daarvan. Dit onderzoek richt zich op de randvoorwaarden die de benutting van een instrument als de invoeringstoets kunnen borgen.

## 1.2 Stand van zaken implementatie invoeringstoets

Op 21 april 2022 stuurde Minister voor Rechtsbescherming Franc Weerwind een brief naar de Kamer met daarin de stand van zaken rondom de implementatie van de invoeringstoets (Weerwind, 2022-a). Hieruit blijkt dat er binnen de Rijksoverheid volop gewerkt wordt aan het verder ontwikkelen en implementeren van de invoeringstoets, die ook meer gestalte begint te krijgen. Daarbij wordt een zogenaamde 'lerende aanpak' gehanteerd, waarbij verschillende onderdelen van de Rijksoverheid experimenteren met het instrument (Weerwind, 2022-a, p. 3). De verschillende organisaties buigen zich daarbij over kernvragen als bij wie de verantwoordelijkheid komt te liggen voor het selecteren van de 'invoeringstoetswaardige' dossiers, wie primair uitvoerend moet worden en of de in de eerdergenoemde Kamerbrief van Rutte (2021, p. 6)

genoemde vragen de juiste zijn. Daarnaast kijken zij ook naar het moment binnen het wetgevingsproces waarop de invoeringstoets uitgevoerd moet gaan worden en de wijze waarop signalen en knelpunten op een goede manier opgehaald kunnen worden bij burgers, organisaties, bedrijven, medeoverheden, medewerkers in uitvoering, rechtspraak, media en politiek (Weerwind, 2022-a, pp. 3-4).

Verder komt naar voren dat door de gekozen lerende aanpak, ook de implementatie van de toets geleidelijk plaats zal gaan vinden (Weerwind, 2022-a, p. 4). Sinds 2022 zijn daarom, verspreid over verschillende overheidsorganisaties, enkele tientallen experimentele invoeringstoetsen uitgevoerd op eerder tot stand gekomen regelgeving (geïmplementeerd in een periode waarin nog geen sprake was van een invoeringstoets). Door op deze manier experimenten rijksbreed uit te voeren kan de kennis over de invoeringstoets volgens de Kamerbrief geleidelijk groeien. Dit jaar moeten ook invoeringstoetsen volgen die gebaseerd zijn op een uitgekristalliseerde werkwijze (Weerwind, 2022-a, p. 4). Kortom: er worden verschillende experimenten uitgevoerd met betrekking tot de werkwijze van de invoeringstoets, maar een volledig afgeronde implementatie zal nog enkele jaren op zich laten wachten.

#### **De invoeringstoets zoals omschreven in de Kamerbrief van de minister van rechtsbescherming**

De invoeringstoets kan beschreven worden als een lichtvoetige bestudering van de werking van nieuwe regelgeving in de praktijk, met bijzondere aandacht voor de gevolgen voor de doelgroep en de uitvoering. Deze toets wordt uitgevoerd op het vroegst mogelijke moment waarop iets nuttigs gezegd kan worden over de werking van regelgeving in de praktijk, en onderscheidt zich daarmee van de uitgebreidere en diepgaandere wetsevaluaties. Voor een volledig beeld van de beleidseffecten heeft regelgeving tijd nodig om in te klinken en een zichtbaar beleidseffect teweeg te brengen. Het doel van de invoeringstoets is daarmee niet om te kunnen vaststellen of de nieuwe regelgeving een succes is. Een vermoeden van het tegenovergestelde zou echter wel aan het licht kunnen komen.

Bron: Weerwind (2022)

### **1.3 De invoeringstoets in de beleidscyclus**

De invoeringstoets moet het karakter krijgen van een *ex durante* evaluatie (Weerwind, 2022-a, p. 2). Deze vorm van evalueren beoordeelt volgens de Kamerbrief wat er voor de goede werking van de regelgeving nog *moet* gebeuren op het moment dat corrigeren nog mogelijk is, in plaats van wat er *had moeten* gebeuren zoals dat bij beleidsevaluatie achteraf (*ex post*) gebeurt. Verder komt in de brief naar voren dat de toets uitgevoerd moet gaan worden op het '*vroegst mogelijke moment waarop iets nuttigs gezegd kan worden over de werking van regelgeving in de praktijk*'. De bedoeling is dat de toets zich daarmee onderscheidt van de uitgebreidere en diepgaandere wetsevaluaties die nu al bestaan en dus ook iets toevoegt aan het bestaande evaluatie-instrumentarium.

De invoeringstoets moet dus een jaar na start van de uitvoering plaatsvinden. Hiermee onderscheidt het zich inderdaad van de meeste andere bestaande beleidsevaluaties die in de beleidsvoorbereiding (*ex ante*), dan wel na volledige inwerkingtreding (*ex post*) plaatsvinden. Voor de beleidsuitvoering wordt bijvoorbeeld al een uitvoeringstoets gedaan, waarbij wordt getoetst of een voorstel ook uitvoerbaar en haalbaar is (KCBR, z.d.-a). Na volledige inwerkingtreding wordt daarnaast een *ex post* evaluatie uitgevoerd waarin de gerealiseerde effectiviteit en efficiëntie van het uitgevoerde voorstel worden beoordeeld (KCBR-b). Een evaluatie die juist relatief vroeg in de beleidsuitvoering plaatsvindt

ontbreekt echter, terwijl deze juist – in tegenstelling tot andere evaluatievormen – tijdig knelpunten kan blootleggen zodat deze ook op tijd opgelost kunnen worden.



#### 1.4 Benutting van beleidsevaluaties

De hierboven geschetste situatie is hoe de invoeringstoets op papier moet gaan werken. De vraag is of deze situatie overeenkomt met de praktijk. Stone (2012, pp. 11-12) schrijft in haar boek 'Policy paradox' over de vele paradoxen die voorkomen in de wereld van de beleidsvorming. Misschien wel de belangrijkste paradox die zij noemt is dat beleidswetenschappen gaan over het verbeteren van het openbaar bestuur, terwijl er binnen de wetenschappelijke discipline een diepe afkeer is tegen de ambiguïteiten en paradoxen van de politiek. Politiek wordt gezien als een obstakel voor rationele analyse en goed beleid. Zij stelt echter dat politiek belangrijk is om elkaar vanuit andere wereldbeelden te laten kijken en zo beter samen te kunnen leven. Daarmee is politiek een gegeven en dient het ook meegenomen te worden in de beleidswetenschappen. Ook voor de invoeringstoets geldt dat een puur rationeel perspectief geen betrouwbare voorspeller is voor de werking van het instrument in de praktijk. Om te voorkomen dat vanuit tunnelvisie wordt gekeken zal dus vanuit verschillende perspectieven naar de invoeringstoets worden gekeken, waaronder ook een politiek perspectief.

Als we vanuit een ander perspectief dan het rationele kijken is het maar de vraag of de invoeringstoets gaat werken zoals bedoeld. Met name de stap van het uitvoeren van een evaluatie zoals de invoeringstoets naar het benutten van de resultaten ervan is namelijk niet zo vanzelfsprekend. Op 11 mei 2021 schoof minister-president Rutte aan bij het actualiteitenprogramma Nieuwsuur om te praten over zijn plannen voor een nieuwe bestuurscultuur (NRC, 2021, 11 mei). Rutte vond dat de overheid 'menschelijker' moest worden en één van de ideeën om dat te bereiken was dat er een nieuwe club moest komen die moest gaan uitzoeken waar het misgaat in de communicatie tussen burger en overheid. De reactie hierop van de nationale ombudsman van Zutphen was: 'Hij moet hier maar eens langskomen, die club bestaat al sinds 1982 ... Laat hem maar een club oprichten die onze rapporten leest'.

Ook in het rapport Klem tussen balie en beleid van de Tijdelijke commissie Uitvoeringsorganisaties (2021, p. 37) van de Tweede Kamer komt, naast dat beleid te weinig wordt geëvalueerd, naar voren dat uitgevoerde evaluaties onvoldoende worden benut. Uitvoeringsorganisaties krijgen daarbij vaak pas aandacht als er incidenten zijn. Een gestructureerde gedachtewisseling aan de hand van evaluaties van beleid en wetgeving blijft volgens de Raad van State vaak achterwege door de focus op incidenten en een regelmatig volgende regelreflex (Tijdelijke Commissie Uitvoeringsorganisaties, 2021, p. 37). Beleidsevaluaties worden dus lang niet altijd even goed benut binnen de Rijksoverheid, waardoor de vraag opkomt of dit bij de invoeringstoets wel het geval gaat zijn.

## 1.5 Benutting van de invoeringstoets geborgd

Zoals eerder is aangetoond is de bedoeling dat de resultaten van de invoeringstoets structureel leiden tot herziening van wetgeving (waar nodig). Maar dan moeten deze wel benut worden. Uit interne gesprekken met de (destijds) projectmanager van de implementatie van de invoeringstoets<sup>1</sup> blijkt dat hier ook bij de invoeringstoets voorzichtige twijfels over zijn. De angst is dat resultaten niet worden gebruikt bijvoorbeeld zodra ze om politieke redenen niet uitkomen, de betreffende beleidsmaker inmiddels al met een andere klus bezig is of de beleidsmakers de toets niet serieus nemen of er geen tijd voor hebben.

Twijfels over de benutting lijken ook geïmpliceerd te worden in de Kamerbrief van Weerwind (2022). Dit gebeurt onder andere in de volgende passage:

*'De invoeringstoets lijkt daarmee volop bestaansrecht te hebben, maar deze toets alleen zal geen verandering teweegbrengen. Het gaat erom te bewerkstelligen dat we na invoering van regelgeving de aandacht vasthouden voor wat we hebben willen bewerkstelligen op het betreffende onderwerp. Dat we nieuwsgierig zijn naar de gevolgen van ons beleidswerk in de samenleving, voor de doelgroep en voor de collega's in de uitvoering. En dat we ook binnen de uitvoering de vraag stellen hoe de verdere vertaling van het beleid in bijvoorbeeld uitvoeringsbeleid en telefooninstructies uitpakt. Dat we de wil hebben om knelpunten die in een invoeringstoets worden geconstateerd op te lossen voor de doelgroep' (p. 3).*

In deze alinea wordt benadrukt dat de uitvoerders van de invoeringstoets nieuwsgierig moeten zijn naar de gevolgen van eigen beleidswerk in de samenleving en voor de uitvoering. Daarnaast moeten zij de wil hebben om knelpunten te constateren en op te lossen voor de doelgroep. Er lijkt een voorzichtige angst door te schemeren dat uitvoerders misschien niet op de toets zitten te wachten en deze als 'weer een extra taak' zullen beschouwen. Voor de effectiviteit van de invoeringstoets is een actieve houding belangrijk. De uitvoerders van de toets moeten actief signalen gaan verzamelen en eventuele knelpunten ook echt willen oplossen.

Het is dus van belang dat de resultaten van de invoeringstoets ook echt benut gaan worden en dat wet- en regelgeving op structurele basis wordt herzien en waar nodig wordt bijgesteld. Kijkend vanuit het model van beleidsvorming van Stone (2012, p. 7) is het om verschillende redenen, zoals politieke motieven, de cultuur binnen de overheid en de houding van beleidsmakers, echter maar de vraag of dit ook zomaar het geval gaat zijn.

## 1.6 Probleem- en doelstelling

De invoeringstoets moet binnen de Rijksoverheid gaan bijdragen aan een betere uitvoering van wet- en regelgeving. Door relatief vroeg na inwerkingtreding te beoordelen hoe een regeling in de praktijk werkt kan deze nog worden aangepast. Het idee is daarbij dat eventuele knelpunten in een vroeg stadium verholpen, en mogelijke grotere problemen voorkomen kunnen worden. Zonder iets te zeggen over of dat doel met de invoeringstoets inderdaad behaald wordt, is het in ieder geval zaak dat de resultaten uit de invoeringstoets überhaupt benut gaan worden. Anders wordt het immers een zinloze inspanning.

---

<sup>1</sup> Pim Beirnaert, persoonlijke communicatie, 29-4-2022.

De benutting van de invoeringstoets kan op verschillende manieren ondermijnd worden. De resultaten kunnen 'niet uitkomen' voor een bewindspersoon, de beleidsmakers kunnen er geen tijd voor hebben of al met een andere klus bezig zijn, of misschien is het dossier wel te complex waardoor een fatsoenlijke toets te veel tijd en geld kost. De bedoeling is juist dat de invoeringstoets, ook met moeilijke of gevoelige dossiers, leidt tot herziening en eventueel bijstelling van wet- en regelgeving (waar nodig). Het is dus onduidelijk hoe ervoor gezorgd kan worden dat de resultaten uit de invoeringstoets ook daadwerkelijk benut worden en hoe deze benutting ook geborgd kan worden in het beleidsproces. Het doel van dit onderzoek is dan ook om meer inzicht te krijgen in hoe de benutting van beleidsevaluaties geborgd kan worden in beleidsprocessen zodat deze ook écht iets toevoegen aan de verbetering van wet- en regelgeving. Dit wordt gedaan door te focussen op de invoeringstoets en interviews af te nemen met betrokkenen van experimenten met invoeringstoetsen alsook soortgelijke evaluaties, om te kijken onder welke randvoorwaarden de benutting van het instrument geborgd kan worden.

## 1.7 Vraagstelling

Dit onderzoek gaat in op hoe de benutting van de invoeringstoets versterkt kan worden en welke randvoorwaarden hierbij van belang zijn. Dit komt neer op onderstaande hoofdvraag:

*Onder welke randvoorwaarden kan de benutting van de invoeringstoets worden geborgd?*

Om tot beantwoording van deze vraag te komen is het allereerst zaak om dieper in te gaan op het hoofdonderwerp van dit onderzoek: de invoeringstoets. Hierbij zal vanuit verschillende perspectieven worden gekeken naar de beoogde werking van het instrument en hoe deze past binnen de beleidscyclus. Vervolgens is het van belang om vanuit de literatuur het concept benutting (van evaluaties) onder de loep te nemen. Hierbij wordt vanuit verschillende modellen een definitie gegeven en worden de verschillende vormen die benutting kan aannemen behandeld. Zodra de belangrijkste concepten zijn gedefinieerd en uitgewerkt kan het theoretische deel worden afgesloten door op basis van de literatuur te komen tot randvoorwaarden voor het borgen van benutting van de invoeringstoets.

Om de theorie te spiegelen aan de praktijk onderzoeken we eerst hoe de invoeringstoets past binnen de beleidscyclus van de Rijksoverheid. Niet alleen in theorie, maar vooral ook in de praktijk. Ook gaan we specifiek in op de benutting en vooral op waar en door wie de toets al dan niet wordt benut. Tot slot is het voor het beantwoorden van de hoofdvraag van belang om in de praktijk te kijken naar hoe en in hoeverre wordt voldaan aan de randvoorwaarden voor het borgen van de benutting van de invoeringstoets. Op basis hiervan kunnen de volgende theoretische en empirische deelvragen worden geformuleerd:

Theoretische deelvragen:

1. *Wat is de invoeringstoets en hoe is deze te positioneren in de beleidscyclus?*
2. *Wat is benutting van evaluatie en welke benaderingen voor benutting zijn te onderscheiden?*
3. *Welke randvoorwaarden voor het borgen van benutting van de invoeringstoets zijn te onderscheiden?*

Empirische deelvragen:

4. *Hoe past de invoeringstoets binnen de beleidscyclus van de Rijksoverheid?*
5. *Hoe en waar wordt de invoeringstoets benut binnen de Rijksoverheid?*
6. *In hoeverre en hoe wordt in de praktijk voldaan aan de randvoorwaarden voor het borgen van de benutting van de invoeringstoets?*

## 1.8 Onderzoekaanpak en theoretisch perspectief

Dit onderzoek bespreekt verschillende theorieën en opvattingen rondom de invoeringstoets. Hierbij gebruiken we eerst drie modellen voor het reconstrueren van beleidsprocessen om de werking van de invoeringstoets te duiden. We gebruiken het fasen-, stromen- en rondemodell om vanuit verschillende perspectieven te kijken naar de beoogde werking van de toets. Daarnaast gebruiken we de modellen ook om de beoogde werking in perspectief te plaatsen en een beeld te schetsen van hoe de invoeringstoets daadwerkelijk kan gaan werken binnen het beleidsproces. Nadat de werking van de invoeringstoets vanuit de theorie is besproken gaan we verder in op het concept benutting. Hierbij bespreken we het artikel van Weiss (1979) dat verschillende definities geeft voor het concept benutting. Ook het artikel van Bekkers, Fenger, Homburg & Putters (2004) wordt behandeld, waarbij het onderscheid in verschillende vormen die benutting kan aannemen centraal staat. Ten slotte noemt het artikel van Bekkers et al. (2004) alsook het artikel van Johnson, Greenesid, Toal, King, Lawrenz & Volkov (2009) verschillende factoren die van invloed zijn op benutting. Deze factoren gebruiken we om te komen tot een lijst randvoorwaarden voor het borgen van de benutting van de invoeringstoets om zo de derde theoretische deelvraag te kunnen beantwoorden.

## 1.9 Relevantie

### *Maatschappelijk*

Dit onderzoek is allereerst maatschappelijk relevant omdat het bijdraagt aan het verbeteren van wet- en regelgeving. Zoals eerder al naar voren is gekomen was de toelagenaffaire de aanleiding voor de komst van de invoeringstoets. De Kamerbrief van de minister van Rechtsbescherming noemt ook dat er binnen de Rijksoverheid meer open moet worden gestaan voor wat er beter kan. Signalen over fouten in het optreden van de overheid en problemen van wetgeving in de praktijk moeten worden opgepakt en opgelost (Weerwind, 2022-a, p. 1).

Concreet kent de invoeringstoets twee doelen (Weerwind, 2022-a, p. 3). Enerzijds het beter in kunnen spelen op signalen van burgers, organisaties en bedrijven over de werking van nieuwe regelgeving in de praktijk. En anderzijds het dienen van de positie van uitvoeringsorganisaties om beter in staat te worden gesteld knelpunten uit de uitvoeringspraktijk terug op tafel te leggen. Daarnaast komt uit het rapport *'Meer dan de som der delen'* van de werkgroep *versterking functies Tweede Kamer* naar voren dat de invoeringstoets ook het parlement zou kunnen helpen in het versterken van de wetgevende, controlerende en vertegenwoordigende functie. Het beter benutten van de invoeringstoets betekent dus dat signalen vanuit de doelgroep en uitvoering beter doorkomen en wet- en regelgeving hierop kan worden aangepast. Onderzoek naar het borgen van de benutting van de invoeringstoets, kan dus bijdragen aan de kwaliteit van wet- en regelgeving.

### *Wetenschappelijk*

Het onderzoek is daarnaast niet alleen maatschappelijk relevant, maar om verschillende redenen ook wetenschappelijk. Het belangrijkste concept in dit onderzoek is benutting, waarover al redelijk veel literatuur bestaat. Ook is er al onderzoek gedaan naar de factoren die van invloed zijn op de benutting van beleidsevaluaties bij overheden (Johnson et al., 2009) en factoren die van invloed zijn op de benutting van

adviezen van adviesraden in Nederland (Bekkers et al., 2004). Wat echter nog ontbreekt en wat dit onderzoek moet toevoegen zijn factoren (randvoorwaarden) die van invloed zijn op de benutting van evaluaties in de context van de Nederlandse overheid. Inzicht hierin is wetenschappelijk relevant omdat het kan worden gebruikt om te onderzoeken hoe beleidsevaluaties in Nederland benut worden en waarom specifieke evaluaties wel of niet impact hebben gehad.

Daarnaast gaat het met de invoeringstoets om een ex durante evaluatie. Deze vorm van evalueren lijkt in de literatuur relatief weinig belicht, maar wordt de laatste jaren steeds vaker genoemd (van der Vlis, Bunte & van Galen, 2007; Pattyn, 2020; Van Hattem & van der Horst, 2022). Door meer kennis te verzamelen over de werking van dit type evaluatie kan ook de bruikbaarheid ervan voor overheden worden vergroot. Dit onderzoek draagt bij aan de kennis ervan.

Tot slot is één van de achterliggende gedachten van de invoeringstoets dat wet- en regelgeving vroegtijdig geëvalueerd, herzien en waar nodig aangepast wordt. Hiermee worden de schakels van de beleidsevaluatie en de beleidsterugkoppeling nadrukkelijker aan elkaar verbonden en kan de beleidscyclus mogelijk beter gaan lopen. Door te onderzoeken hoe de benutting van een dergelijk instrument geborgd kan worden, kan dus ook meer inzicht verkregen worden in het sluitend maken van de beleidscyclus.

## 1.10 leeswijzer

Om tot beantwoording te komen van de vraag onder welke randvoorwaarden de benutting van de invoeringstoets kan worden geborgd, wordt eerst meer kennis verzameld over de invoeringstoets zelf, en vooral in welk licht het idee hiertoe is ontstaan. In hoofdstuk 2 wordt daarom de politiek-maatschappelijk context geschetst waarin het idee voor een invoeringstoets is ontstaan. In hoofdstuk 3 worden vervolgens vanuit de literatuurstudie de drie theoretische deelvragen beantwoord. Het onderzoek en de bijbehorende methoden worden in hoofdstuk 4 uiteengezet waarna in hoofdstuk 5 de resultaten hiervan worden gepresenteerd. Tot slot worden in hoofdstuk 6 de belangrijkste conclusies en implicaties besproken, waarna een aantal aanbevelingen wordt gedaan voor vervolgonderzoek en de implementatie van de invoeringstoets.

## 2. De invoeringstoets in context

### 2.1 Inleiding

Dit hoofdstuk schetst de maatschappelijk-politieke context waarin het idee van de invoeringstoets is bedacht en ontwikkeld. Met de maatschappelijk-politieke context wordt bedoeld op de ontwikkelingen in politiek en maatschappij van voornamelijk de jaren 10 van deze eeuw, die de grond vormt waarin het zaadje voor de invoeringstoets gepland heeft kunnen worden. Dit zijn bijvoorbeeld ontwikkelingen in de verhoudingen tussen kabinet en parlement, tussen burger en overheid en tussen kabinet en uitvoering. Deze context staat in dit hoofdstuk centraal.

De maatschappelijk-politieke achtergrond kan het best worden geschetst vanuit een aantal rapporten dat de afgelopen jaren is verschenen en ook centraal staat in dit hoofdstuk. In de inleiding is al genoemd dat de invoeringstoets één van de genoemde veranderingen is die het kabinet heeft aangekondigd als reactie op de conclusies over de toeslagenaffaire. In het rapport 'Ongekend onrecht' van de Parlementaire ondervragingscommissie Kinderopvangtoeslag (Tweede Kamer, 2020a) wordt hier verder op in gegaan. De komst van de invoeringstoets kan echter ook in een breder licht worden gezien, namelijk in dat van het rijksbrede programma 'Werk aan Uitvoering' (WAU) waarin de verbetering van de publieke dienstverlening centraal staat. Nog breder houdt het verband met de roep om een nieuwe bestuurscultuur, inclusief transparant ambtelijk gedrag en zuivere politiek-ambtelijke verhoudingen. Dit wordt verder uitgewerkt aan de hand van de rapporten over het programma (Heerschop & van der Vlist, 2020a; Heerschop & van der Vlist, 2020b), het rapport van de Tijdelijke commissie Uitvoeringsorganisaties 'Klem tussen Balie en Beleid' (Tweede Kamer, 2021) en de informateursrapporten van Tjeenk Willink uit 2017 en 2021 waarin hij aandacht vraagt voor de uitvoering en hij een oproep doet voor een nieuwe bestuurscultuur (Willink, 2017; Willink, 2021).

### 2.2 De toeslagenaffaire ('Ongekend onrecht')

Om de komst van de invoeringstoets te begrijpen kijken we allereerst naar de directe aanleiding ervan: de kinderopvangtoeslagaffaire. De affaire is namelijk typerend voor de verschillende maatschappelijk-politieke ontwikkelingen waar de inleiding van dit hoofdstuk op doelt. Ook in de inleiding van het onderzoek is de toeslagenaffaire al kort besproken, maar om écht de context achter de invoeringstoets te kunnen begrijpen is meer diepgang nodig. Hiertoe wordt het rapport 'Ongekend onrecht' van de Parlementaire ondervragingscommissie Kinderopvangtoeslag (Tweede Kamer, 2020a) verder besproken.

Nadat in 2005 de Wet Kinderopvang (WKO) – de wet die een maandelijkse bijdrage regelt voor de kinderopvang van kinderen met twee werkende ouders – werd ingevoerd, ging veel fout. In feiten was de uitvoering voor de afdeling toeslagen (destijds nog) bij de Belastingdienst té ingewikkeld, maar omdat de toeslag samenhang met de zorgtoeslag en een nieuw zorgstelsel, werd het toch ingevoerd. De organisatie werkte tot 2011 met een ontoereikend ICT-systeem waardoor aanvragen pas achteraf konden worden gecontroleerd en soms zelfs tot 5 jaar na het uitkeren. Het kabinet nam daarnaast het advies van de Raad van State om een hardheidsclausule op te nemen niet over. Deze clausule moest het mogelijk maken om in onvoorziene en schrijnende gevallen af te kunnen wijken van de standaardregels. Volgens het kabinet moest



Toeslagen vanwege de grote hoeveelheid aanvragen echter als een machine gaan werken waardoor uitzonderingen voor 'zand in de raderen' zou gaan werken (Tweede Kamer, 2020a, p. 13).

Om fraude tegen te gaan werd de toeslag van ouders groepsgewijs stopgezet als er bij de kinderopvanginstelling vermoedens van fraude waren. Als onvolkomenheden werden gevonden moest de hele toeslag worden terugbetaald en als opzet of grove schuld aannemelijk werd geacht door de Belastingdienst kregen de ouders ook geen persoonlijke betalingsregeling. Hierdoor moest het gehele bedrag in 24 maandelijkse termijnen worden terugbetaald met dwanginvorderingsmaatregelen bij verzaking. Nadat in 2013 aan het licht kwam dat een Bulgaarse bende misbruik maakte van het feit dat de toeslagen pas achteraf werden gecontroleerd en gecorrigeerd, werd de fraudeaanpak nog veel verder geïntensifieerd. De bende incasseerde op grote schaal huur- en zorgtoeslag door gebruik te maken van valse adressen, maar de aanvragers verdwenen voordat de foutieve uitkeringen konden worden gecorrigeerd. De druk op de toeslagenmachine nam hierdoor verder toe (Tweede Kamer, 2020a, p. 14).

Nadat de fraude van de Bulgaren aan het licht kwam werd onder andere meer gecontroleerd voorafgaand aan het uitkeren van toeslagen. Het in 2011 ingevoerde ICT-systeem maakte het mogelijk om op grote schaal controles uit te voeren. Middels een risico-classificatiemodel dat zelflerend was op basis van goede en foute aanvragen, werden de aanvragen waar vermoedens waren dat iets niet klopt gefilterd. Later in 2020 bleek dat dit systeem in ieder geval van 2016 tot 2018 een discriminerende werking had omdat het ook de nationaliteit meenam via de indicator 'Nederlandschap' (Tweede Kamer, 2020a, p. 14).

Verder werd een speciaal managementteam fraudebestrijding opgericht dat met een jaarlijks budget van €25 miljoen zoveel fraude moest opsporen dat ten minste hetzelfde bedrag terugverdiend zou worden door minder toeslagen onterecht uit te keren. Hier werd ook op gestuurd door de directie. Hiertoe werd in 2013 het Combiteam Aanpak Facilitators (CAF) opgericht met als doel systeemfraude aan te pakken door zich te richten op personen die bewust grootschalige systeemfraude organiseerden of in gang zetten (facilitators). De afhandeling van dergelijke CAF-zaken kreeg een veel grimmiger aanblik in de praktijk. Toeslagen van ouders werden groepsgewijs en plotseling stopgezet zonder dat onderzoek was gedaan bij individuele ouders. Informatie van ouders werd wantrouwend en excessief beoordeeld en de kleinste administratieve tekortkomingen waren al voldoende om een heel jaar aan toeslag terug te moeten betalen (Tweede Kamer, 2020a, pp. 14-15).

In 2013 ging ook het ministerie van Sociale Zaken en Werkgelegenheid zich mengen met de fraude aanpak (daarvoor hield zij zich louter inhoudelijk bezig met het beleid en niet met fraudebestrijding). Met het idee dat fraude en slechte kwaliteit veelal samengaan, werd het project kwaliteitsverbetering en fraudebestrijding gastouderbureaus (KEF) gestart. Slecht presterende gastouderbureaus moesten uit het stelsel worden gedrukt door (1) toezicht te houden op de kwaliteit en (2) de kinderopvangtoeslag te stoppen van ouders wiens administratie gebreken vertoonde, zodat de inkomsten van het gastouderbureau op zouden drogen. Het aanpakken van de bureaus werd daarmee gedaan via de portemonnee van ouders wiens administratie niet klopte (Tweede Kamer, 2020a, p. 15).

Ondanks verschillende casussen van kinderopvanginstituten waarbij de toeslagen van ouders onterecht groepsgewijs is stopgezet, meerdere brieven van getroffen ouders en een rapport van de Ombudsman over de knelpunten bij de Belastingdienst, werd pas in 2019 bij de politieke en ambtelijke top duidelijk wat de omvang was van de affaire. In dat jaar werd door vermoedens van de misstanden een onderzoek ingesteld naar

de fraudebestrijding door de ministerraad. Het rapport concludeerde uiteindelijk dat sprake was van institutionele vooringenomenheid. Later trad het kabinet hiervoor af (Tweede Kamer, 2020a, pp. 15-18).

## 2.3 Kabinetsformatie 2017

Een aantal zaken in de kern van de toeslagenaffaire is kenmerkend voor de ontwikkelingen rondom de uitvoering van de afgelopen dertig jaar. Tjeenk Willink omschrijft deze ontwikkelingen in een bijlage van zijn eindrapport als informateur voor de kabinetsformatie van 2017 (Willink, 2017). In de bijlage getiteld 'Over de uitvoerbaarheid en uitvoering van nieuw beleid (regeerakkoord)' gaat hij in op de frictie tussen 'politieke tijd' en 'tijd voor de uitvoering'. Hij doelt hiermee op het tijdspad waarin gemaakte politieke afspraken (uit regeerakkoorden) moeten worden gerealiseerd en de tijd die de uitvoering nodig heeft om alle beleidswijzigingen fatsoenlijk in te kunnen voeren. Hij stelt dat de afgelopen dertig jaar naar een kleinere en goedkopere centrale overheid is gestreefd en dat dit heeft geresulteerd in een aanzienlijke groei van het aantal regels en formulieren waar de burger mee te maken heeft. Individuele burgers worden hier vaak het slachtoffer van omdat zij niet in de standaardmodellen passen en moeite hebben om gehoor te krijgen bij problemen die één loket overstijgen (Willink, 2017).

In het kader van het aankomende nieuwe regeerakkoord stelt Tjeenk Willink dat departementen en uitvoeringsorganisaties al belast zijn met de uitvoering van bestaand beleid (Tjeenk Willink, 2017). Nieuwe beleidsvoornemens sluiten hier niet per se bij aan, maar bij frictie domineert de politiek en niet de uitvoering. Dit komt onder andere naar voren zodra de kinderopvangtoeslag vanwege het nieuwe zorgstelsel wordt ingevoerd, ondanks dat de uitvoering aangaf hier niet klaar voor te zijn.

Ook gaat Tjeenk Willink in op de vele bezuinigingsoperaties die sinds 2010 zijn doorgevoerd en grote impact hebben gehad op de uitvoeringscapaciteiten (Willink, 2017). Het heeft geleid tot veel reorganisaties en extra taken zonder extra budget. Dit betekent weer een vermindering van het inzicht in het eigen functioneren en daarmee ook op het toezicht door bewindslieden. Uiteindelijk is de burger hier de dupe van, want 'waar departementen en publieke uitvoeringsorganen noodgedwongen hun aandacht en energie intern richten op de uitvoering van reorganisaties en substantiële bezuinigingen, raken burgers onvermijdelijk buiten beeld'. Dit wringt volgens hem met de steeds breder gevoelde noodzaak om de verbindingen tussen burgers en overheid te bevorderen of te herstellen (Tjeenk Willink, 2017). Ook de vele bezuinigingen zijn onderdeel geweest van de misstanden bij de toeslagenaffaire. Zo werden bij Toeslagen onder kabinet Rutte 1 flinke bezuinigingen aangekondigd die gerealiseerd moesten worden door versobering en efficiënter werken. Daarnaast kon alleen extra budget worden vrijgemaakt voor de fraudebestrijding door het in 2013 opgerichte managementteam voor fraudebestrijding, als dit zichzelf ook weer terugverdiende. Met alle gevolgen van dien.

Ten slotte stelt Willink dat het zorgelijk is hoe anno 2017 de departementen geen scherper zicht hebben op de effecten van de uitvoering en de effecten van beleid en dat de politiek onmachtig blijkt te voorkomen dat juist burgers in het woud van regels de weg kwijtraken en zich door de overheid niet gehoord of begrepen voelen (Tjeenk Willink, 2017). Wederom maakt hij hiermee een kenmerkend punt voor de toeslagenaffaire, waarbij de hoeveelheid aan regels het vaak moeilijk maakte voor ouders om overal aan te voldoen. Willink sluit af door de overheid en publieke dienstverlening op te roepen om niet incidenteel maar stelselmatig te letten op de uitvoeringsaspecten van het beleid, de eigen beleidssector ook stelselmatig vanuit de ogen van de burgers te bezien en ervoor te zorgen dat de obstakels die burgers in hun contacten met overheid en semi-publieke diensten ondervinden ook daadwerkelijk worden weggenomen.

## 2.4 Werk aan uitvoering

In 2019 lijken de problemen in de uitvoering voor het eerst echt herkend en opgepakt te worden. Naar aanleiding van de bijlage van Tjeenk Willink geeft het kabinet namelijk aanleiding voor het programma 'Werk aan Uitvoering' (WaU) ter verbetering van de publieke dienstverlening (Heerschop & van der Vlist, 2020a), terwijl de Kamer de eerste stappen zet voor een parlementair onderzoek naar de problemen in de uitvoering (Tweede Kamer, 2020b). Beide ontwikkelingen worden besproken, te beginnen met het programma 'Werk aan Uitvoering'.

Het doel van het programma is als volgt geformuleerd: *'het reduceren van continuïteitsrisico's, het vergroten van de verandercapaciteit/wendbaarheid van uitvoeringsorganisaties en een integrale en fundamentele aanpak voor een verbeterde dienstverlening'* (Heerschop & van der Vlist, 2020a, pp 7). Binnen het programma wordt in februari 2020 het rapport voor de eerste fase: 'Probleemanalyse' gepubliceerd en in juli van hetzelfde jaar het rapport voor fase 2: 'Handelingsperspectieven en samenvatting analyse' (Heerschop & van der Vlist, 2020b). Het eerste rapport komt met 7 conclusies over de huidige stand van de uitvoering.

1. Stapeling van beleid, wet- en regelgeving	De stapeling van wet en regelgeving als gevolg van politieke druk, het negeren van opmerkingen uit de uitvoering in het parlementaire proces en de gebrekkige ruimte om te vereenvoudigen, zorgen voor een hoge mate van complexiteit die voor de burger niet uit te leggen is.
2. De ambtelijke besturingsdriehoek	Voor het bestuur van uitvoeringsorganisaties geldt de driehoek van de rollen eigenaar (de SG of pSG), opdrachtgever (de beleids-DG) en opdrachtnemer (bestuur of DG van de uitvoeringsorganisatie). De besturing richt zich vrij eenzijdig op budget, rechtmatigheid, doelmatigheid en tijdige implementatie van nieuw beleid. Hierbij is minder aandacht voor de kwaliteit en continuïteit van de dienstverlening, het delen van keuzes bij risicogerichte handhaving, innovatie en de lange termijn.
3. De impact van incidenten	Incidenten ontstaan door een stapeling van factoren, zoals ook bij de kinderopvangtoeslag. De politiek eist hierop een vlotte betaling waar de werkprocessen vervolgens op ingericht worden. Naar aanleiding van fraude vindt een overreactie plaats. Dit zigzag patroon tussen dienstverlening en handhaving is schadelijk voor burgers en ook voor de uitvoering onwenselijk.
4. De dynamiek van uitvoeringsorganisaties	Uitvoeringsorganisaties staan aan alle kanten onder druk. Regels moeten begrijpelijk zijn, nieuwe wetten moeten snel ingevoerd worden, er moet ruimte voor maatwerk zijn etc. Het is een permanente uitdaging om aan alle eisen en verlangens te voldoen.
5. ICT en geautomatiseerde processen	De Nederlandse overheid ligt voorop met de digitalisering, maar de geautomatiseerde systemen worden steeds complexer en er zijn veel kwetsbaarheden. Met name systemen die uit het zicht zijn geraakt door de IT-afdeling (schaduwgebruik) veroorzaken veel uitvoeringsleed en frustratie bij burgers. Daar komt nog bij dat sluipenderwijs achterstallig onderhoud leidt tot risico's voor de continuïteit van de dienstverlening en tot het niet (tijdig) kunnen inpassen van nieuw beleid.
6. Digitalisering en gebruik van data	Door de digitalisering en toegenomen data-uitwisseling worden gegevens eenmalig ingevoerd en via keten- of netwerkautomatisering veelvuldig gebruikt. Omdat de primaire systemen nooit voor dit concept zijn ontworpen wordt het voor burgers, bedrijven en instellingen ingewikkeld om fouten te herstellen.

7. Ruimte voor mensen	Jaren van bezuinigingen en afslanken hebben geleid tot standaardiseren en automatiseren van werkprocessen. Aansturing van mensen is ook jaren gericht op efficiency en prestaties (new public management) wat ook de cultuur aangetast heeft. Om de verschuiving naar werken gericht op de maatschappelijke opgave te realiseren zal ook het personeelsbeleid hiertoe moeten veranderen.
-----------------------	--

*Fase 1: Handelingsperspectieven en samenvatting analyse*

Daarop aansluitend komt het tweede rapport juist met handelingsperspectieven om de huidige situatie in de periode 2020 tot 2030 te verbeteren (Heerschop & van der Vlist, 2020b).

1. Verbeteren dienstverlening	Burgers en bedrijven kunnen rekenen op een deskundige overheid, die op eenvoudige wijze toegankelijk is voor iedereen. Naast de verdere opmars van de digitale dienstverlening zal de burger ook een beroep kunnen doen op fysiek persoonlijk contact met de overheid. Begrijpelijke taal, eenvoudige(re) procedures en waar nodig dienstverlening op maat.
2. Versnellen digitale agenda	Overheidsdienstverlening aan burgers en bedrijven is verschoven van een sterk proces-georiënteerde 'one size fits all' inrichting en aanpak van de publieke dienstverlening, naar een aanpak waarin de behoefte van burgers en bedrijven centraal komt te staan. Bij digitalisering betekent dit gemak en minder verrassingen in het contact met uitvoeringsorganisaties en een proactieve overheid.
3. Opruimen en vernieuwen wet- en regelgeving	De continuïteit en wendbaarheid van uitvoeringsorganisaties is vergroot, door de reductie van de hoeveelheid, complexiteit en gedetailleerdheid van beleid en wet- en regelgeving. Burgers en bedrijven kunnen dan rekenen op begrijpelijke wet- en regelgeving en eenvoudigere procedures.
4. Intensiveren samenwerking en sturing binnen de driehoek	Er is een intensieve en duurzame samenwerking en sturing op maatschappelijke opgaven tot stand gekomen tussen politiek, ministeries en uitvoeringsorganisaties.
5. Vergroten statuus en aantrekkelijkheid van de uitvoering	De uitvoering beschikt over voldoende middelen, kwantitatieve en kwalitatieve capaciteit, politieke dekking en een helder beleidskader om vanuit de bedoeling van de wet en de eigen deskundigheid de burgers en bedrijven efficiënter en effectiever te kunnen helpen. Een uitvoering die wendbaar is en waarbij de continuïteit van de dienstverlening en het belang vanuit het perspectief van burgers en bedrijven is gewaarborgd.
6. Rol van de politiek	Over tien jaar is het vertrouwen tussen burger en overheid verbeterd. Beleid is goed en duurzaam uitvoerbaar, goed uitlegbaar en begrijpelijk voor burgers. Er is ruimte voor maatwerk, waarover door de uitvoering verantwoording wordt afgelegd. Er is een directe en open wisselwerking tussen politiek en uitvoering op basis van betrouwbare informatie. Maatschappelijke opgaven staan centraal bij vormgeving en uitvoering van beleid.

Het rapport concretiseert hoe de komende 10 jaar gewerkt moet gaan worden aan een betere uitvoering (Heerschop & van der Vlist, 2020b). Stap voor stap moet worden gewerkt om het doel te bereiken dat de dienstverlening voor alle burgers robuust is verankerd in wendbare

samenwerkende uitvoeringsorganisaties. Het rapport geeft daarnaast ook het gevoel dat het doel serieus genomen wordt binnen Rijksoverheid en dat men ook commitment heeft om het te halen. Zo wordt aangegeven dat de gezamenlijke secretarissen-generaal (SGO), samen met de topmanagementgroep (TMG), en bestuurders van uitvoering de ambitie hebben om het rapport ook om te zetten in een meerjarig werkprogramma. Daarnaast is één van de acties die genoemd wordt de komst van de 'Staat van de Uitvoering', waarbij periodiek inzicht moet worden gegeven in cruciale aspecten die noodzakelijk zijn voor de continuïteit en wendbaarheid van een robuuste uitvoeringsorganisatie. Hiermee moet het belang van de uitvoering permanent op de topambtelijke en politieke agenda blijven (Heerschop & van der Vlist, 2020b).

## 2.5 De Tijdelijke commissie Uitvoeringsorganisaties (TCU)

Ook de Tweede Kamer herkent in 2019 de problemen bij de uitvoering en neemt een motie aan waarin het presidium verzocht wordt een voorstel voor te bereiden voor *'een parlementair onderzoek over de oorzaken van problemen bij uitvoeringsorganisaties, het verlies aan menselijke maat, de rol en informatiepositie van de Tweede Kamer zelf en hoe de uitvoerbaarheid van beleid beter gewaarborgd kan worden in het parlementair proces'* (Tweede Kamer, 2019). Dit onderzoek komt er en twee jaar later verschijnt het rapport 'Klem tussen balie en beleid'. 'Het onderzoek heeft als focus de terugkerende oorzaken van problemen bij uitvoeringsorganisaties, op diverse beleidsterreinen' en besteedt hierbij specifiek aandacht aan de Belastingdienst, het CBR en het UWV (Tweede Kamer, 2021, p. 9).

Het rapport concludeert dat Nederlandse uitvoeringsorganisaties in het algemeen goed presteren, maar dat ook burgers (soms ernstig) klem komen te zitten door uitvoeringsproblemen (Tweede Kamer, 2021, p. 10). Het concludeert dat de samenwerking tussen Tweede Kamer, departementen en uitvoeringsorganisaties bij de totstandkoming en uitvoering van nieuw beleid onvoldoende is. Ook zou tussen deze partijen te weinig onderling vertrouwen en interesse in en kennis van elkaars werkzaamheden zijn. Daar komt nog bij dat voorgenomen beleid niet altijd getoetst wordt op uitvoerbaarheid en problemen aan de balie veelal niet terecht komen bij de mensen die ze kunnen oplossen (Tweede Kamer, 2021, p. 10).

Met name die laatste conclusie is natuurlijk ook relevant voor het ontstaan van de invoeringstoets. Signalen over problemen moeten soms een leemlaag aan managers door voordat iemand een besluit kan nemen. Zelfs als signalen wél bij de ambtelijke top van de uitvoeringsorganisatie terecht komen, bereiken deze niet altijd het departement of de verantwoordelijke bewindspersoon. Dit kan komen omdat de ambtelijke top te door een gebrek aan kennis niet op waarde weet te schatten of omdat onvoldoende moed wordt getoond om het departement tegen te spreken. Daarnaast worden signalen soms niet gedeeld met de Kamer omdat de bewindspersoon erdoor in de problemen kan komen (Tweede Kamer, 2021, p. 11).

Een andere relevante conclusie is dat de Kamer een blinde vlek heeft voor de uitvoering. Dit komt enerzijds door een gebrek aan interesse vanuit de Kamer zelf, maar anderzijds ook omdat de departementen haar niet altijd volledig en tijdig informeert. Daarnaast concludeert het rapport dat kennis van uitvoeringsprofessionals onvoldoende wordt benut en dat beleid niet of nauwelijks wordt geëvalueerd op onbedoelde effecten voor burgers (Tweede Kamer, 2021, p. 13).

Al deze conclusies wijzen op problemen waartoe de invoeringstoets moet gaan helpen en die de komst ervan kunnen verklaren, wat ook in hetzelfde rapport wordt bevestigd. Het rapport doet namelijk zeven aanbevelingen voor een dienstverlening met een menselijker gezicht, waarvan het zo snel mogelijk signaleren en adresseren van problemen in de uitvoering er één is. Hiertoe wordt voorgesteld om een

'reflectietoets' in te voeren die één jaar na inwerkingtreding van een wet zicht geeft op de effecten ervan – en niet pas na vijf jaar. Voor deze evaluatie moeten alle betrokkenen een rol krijgen, dus burgers, belangenorganisaties, cliëntenraden en professionals op de werkvloer. Kortom: er moet een invoeringstoets komen.

## 2.6 De nieuwe bestuurscultuur

Nadat het kabinet Rutte III in 2020 aftreedt vanwege de Toeslagenaffaire, vinden op 15 maart 2021 nieuwe verkiezingen plaats en start een nieuwe formatie (Tweede Kamer, 2023). Ook hierin zijn de gevolgen van de toeslagenaffaire weer volop zichtbaar en is de uitvoering nadrukkelijk onderwerp van gesprek. Dit wordt bijvoorbeeld al zichtbaar wanneer (wederom) Tjeenk Willink als informateur wordt aangesteld en hij in gesprek gaat met alle verkozen fracties. In zijn rapport hierover (Willink, 2021) schrijft hij over hoe door velen fracties een pleidooi werd gehouden voor een meer dualistische verhouding tussen Kamer en kabinet en een open bestuursstijl.

Met name de verstoorde verhouding tussen Kamer en regering is vaak genoemd in de verschillende rapporten die zijn besproken in dit hoofdstuk en hangt ook sterk samen met de open bestuursstijl die genoemd wordt. Het gebrek aan balans komt volgens Tjeenk Willink onder andere voort uit de dichtgetimmerde regeerakkoorden die het controlerend vermogen van de Kamer ondermijnen. Daarnaast wijst hij erop dat de controleur (de Kamer) afhankelijk is van de informatie die hij krijgt van de gecontroleerde (het kabinet). Dat gaat ten koste van de systematische controle op de uitvoering en de effecten van beleid en wetgeving op de burger.

Naast de verstoorde verhouding tussen Kamer en regering spreekt Tjeenk Willink ook van een verstoorde verhouding tussen overheid en burger. Hij schrijft hierover dat de burger de overheid alleen zal vertrouwen als deze zelf ook betrouwbaar is en uitgaat van vertrouwen in burgers. Het functioneren van de overheid als een bedrijf dat 'gemanaged' moet worden waarbij de burger de rol van klant en kostenpost heeft, is volgens hem schadelijk geweest voor de basis van dit wederzijdse vertrouwen. Hij noemt hierbij de toeslagenaffaire als voorbeeld.

Een nieuwe bestuursstijl moet (in 2021) de relaties tussen Kamer en regering en overheid en burger dus herstellen. Hoe deze nieuwe bestuursstijl er precies uit moet komen te zien wordt niet concreet gemaakt, maar het een en ander kan wel worden opgemaakt uit het rapport van Tjeenk Willink. Kernpunten die vaak terugkomen zijn dat er meer wederzijds vertrouwen moet komen tussen Kamer en regering, overheid en burger en de verschillende politieke partijen; het dualisme moet worden versterkt en de controlerende functie van de Kamer moet worden versterkt door minder 'dicht te timmeren' in het regeerakkoord.

De nieuwe bestuurscultuur wordt in de maanden na het rapport van Tjeenk Willink een kernpunt van de formatie. Een belangrijke oorzaak hiervoor is de affaire rondom Kamerlid Pieter Omtzigt die op 1 april zorgde voor een motie van wantrouwen voor premier Rutte (Keultjes en van Ast, 2022). De term lijkt daarbij ook in zekere zin symbool te staan voor de ontwikkelingen die in dit hoofdstuk zijn omschreven en die ook de aanleiding vormen voor de invoeringstoets. Uit de rapporten 'Ongekend Onrecht', 'Klem Tussen Balie en Beleid' en 'Werk aan Uitvoering' fase 1 en 2 blijkt dat de uitvoering jarenlang verwaarloosd is, wat tot enorme crises als de kinderopvangaffaire heeft geleid en waar de regering te weinig transparant over is geweest. Deze ontwikkelingen vragen om radicale verandering en de nieuwe bestuurscultuur omvat deze verandering.

## 2.7 De invoeringstoets

Alle ontwikkelingen die in dit hoofdstuk zijn omschreven dragen bij aan het ontstaan van de invoeringstoets. De toeslagenaffaire is hierbij de directe aanleiding geweest, maar de echte grondslag ligt verder terug. De echte grondslag zit hem vooral in de jarenlange verwaarlozing van de uitvoering en de hoogopgelopen druk op uitvoeringsorganisaties, zoals omschreven in de rapporten 'Ongekend onrecht' van de TCU (Tweede Kamer, 2019), 'Werk aan Uitvoering: fase 1' en de bijlage van Tjeenk Willink uit 2017. Nadat 30 jaar gestreefd is naar een kleinere en goedkopere centrale overheid zijn na de financiële crisis in 2010 nog vele bezuinigingsoperaties doorgevoerd. Tegelijkertijd bleven de uitvoeringsorganisaties kampen met de stapeling van beleid, wet- en regelgeving, de last van bestaand beleid en de gevolgen van verouderde ICT-systemen. De focus vanuit de politiek lag tegelijkertijd alleen op incidenten die telkens direct opgelost moesten worden, ook al was de beste oplossing nog niet gevonden. Tot slot bleken de departementen ook nog eens onvoldoende in staat om zicht te houden op de effecten van de uitvoering en nieuw beleid. Voorgenomen beleid werd niet altijd getoetst op uitvoerbaarheid en problemen aan de balie kwamen veelal niet terecht bij de mensen die ze konden oplossen.

De toeslagenaffaire was een gevolg van al deze ontwikkelingen en het is dan ook niet gek dat het kabinet in januari 2021 een invoeringstoets aankondigt in de reactie op het rapport van die affaire (Rutte, 2021). De brief noemt naar aanleiding van de affaire verschillende maatregelen om te leren van de gemaakte fouten, waarbij de invoeringstoets moet helpen om de praktische werking van wet- en regelgeving te toetsen. Het instrument wordt hierin verder nauwelijks toegelicht. Er wordt alleen toegelicht dat het kabinet voornemens is om een jaar na start van de uitvoering een toets te doen op nieuwe wet- en regelgeving waarbij de vragen centraal staan die ook al in de inleiding zijn genoemd.

Bij het ministerie van Justitie en Veiligheid is men vervolgens aan de slag gegaan met de implementatie van de toets, waarna in 2022 de Kamerbrief van minister Weerwind is verstuurd (Weerwind, 2022-a). Deze brief geeft voor het eerst meer invulling aan het instrument en het idee erachter. In de aanleiding wordt gesteld dat signalen over fouten in het overheidsoptreden en problemen van regelgeving in de praktijk moeten worden opgepakt en opgelost. De overheid hoort daar volgens de brief actief naar op zoek te gaan en moet dus ook onderzoeken hoe regelingen in de praktijk uitpakken voor mensen. Zij moet nieuwsgiering zijn naar de daadwerkelijke gevolgen van beleidswijzigingen en minder vertrouwen op de verwachte uitkomsten (p. 1).

De invoeringstoets wordt in de brief omschreven als een lichtvoetige bestudering van de werking van nieuwe regelgeving in de praktijk, met bijzondere aandacht voor de gevolgen voor de doelgroep en de uitvoering (Weerwind, 2022-a, p.2). Hij wordt uitgevoerd op het vroegst mogelijke moment waarop iets nuttigs gezegd kan worden over de werking van nieuwe regelgeving in de praktijk. Het wordt verder omschreven als een *ex durante* evaluatie die zich richt op wat er *moet* gebeuren in plaats van wat er *had moeten* gebeuren. De brief erkent daarbij wel dat het een werkverzwaring betekent, maar dat die gezien de doelen van het instrument gerechtvaardigd worden geacht. Ook wordt het risico van een te instrumentele benadering benoemd: *'Als de invoeringstoets een doel op zich wordt en altijd voor alle regelgeving gedaan zou moeten worden, zou het ook in veel gevallen niets opleveren en dus een betekenisloze exercitie zijn die veel tijd en geld kost zonder merkbaar resultaat. We moeten dus verstandig en doelmatig met dit instrument omgaan en dus selectief en doelgericht zijn, in combinatie met een open blik voor onverwachte signalen omdat we ook blinde vlekken moeten voorkomen'* (Weerwind, 2022-a, p.2). De doelen van de toets zijn er volgens de Kamerbrief nadrukkelijk twee. Ten eerste moet het kabinet beter in kunnen spelen op signalen van burgers, organisaties en bedrijven over de werking van nieuwe regelgeving in de praktijk. Ten tweede moet het de uitvoering beter in staat stellen om knelpunten uit de

uitvoeringspraktijk terug te leggen op de beleidstafel. Daarop aansluitend wordt ook het eerdergenoemde doel erkend dat de Kamers wetgevende, controlerende en vertegenwoordigende functie ermee versterkt kan worden.

Verder geeft de Kamerbrief aan dat is gekozen voor een lerende aanpak waarbij verschillende onderdelen van de rijksoverheid zijn gaan experimenteren met de invoeringstoets (Weerwind, 2022-a, p.4). Voorbeelden en werkwijzen worden interdepartementaal uitgewisseld om te bepalen aan welke uitgangspunten een toets moet voldoen om een invoeringstoets genoemd te kunnen worden. Daarbij wordt wel gewaakt voor een one-size-fits-nobody-model en wordt de behoefte aan maatwerk en flexibiliteit erkend.

Als reactie op de brief van minister Weerwind komt de Kamer met een aantal aanvullende verzoeken waaronder een overzicht van alle uitgevoerde en geplande invoeringstoetsen (Weerwind, 2022-b). In september 2022 wordt deze brief beantwoord met onder andere een overzicht met 6 uitgevoerde invoeringstoetsen en nog eens 32 geplande invoeringstoetsen (Weerwind, 2022-b). Daarnaast wordt op een aantal aspecten van de toets extra verduidelijking gegeven. Zo worden bijvoorbeeld 4 redenen gegeven voor het selecteren van een onderwerp voor een invoeringstoets:

- om (met verschillende soorten regelingen) te oefenen met dit evaluatieinstrument (lerende aanpak);
- omdat er uit de praktijk signalen zijn gekomen over knelpunten voor de doelgroep of de uitvoerende organisaties;
- omdat anderen aanleiding zagen om een invoeringstoets te doen, bijvoorbeeld de Afdeling advisering van de Raad van State of leden van uw Kamer, waarna dat is toegezegd;
- omdat de betreffende uitvoeringsorganisatie een terugblik op de invoering in de standaardprocessen heeft verwerkt.

Naast deze redenen voor het selecteren van een onderwerp geeft de brief ook nog drie indicatoren om vervolgens een kwalitatieve, onderbouwde keuze te maken om al dan niet daadwerkelijk een invoeringstoets te doen:

- Uitvoerende fase: tijdens de implementatie en in de eerste periode waarin de regeling effectief is, ontstaan er signalen over de daadwerkelijke effecten in de praktijk die aanleiding kunnen zijn voor een invoeringstoets.
- Besluitvormende fase: in de besluitvorming kan de toezegging dat bepaalde effecten goed gemonitord zullen worden (met een invoeringstoets) van belang zijn voor bijvoorbeeld het verminderen van zorgen en vergroten van draagvlak.
- Beleidsvoorbereidende fase: specifieke kenmerken van een voorgenomen regeling kunnen ertoe leiden dat de inschatting wordt gemaakt dat er aanleiding is voor een invoeringstoets.

Het is dus duidelijk dat niet stil wordt gezeten bij het implementeren van de invoeringstoets, hoewel wel nog flinke slagen gemaakt moeten worden.

## 2.8 Conclusie

Het ontstaan van het idee van een invoeringstoets valt dus te verklaren uit een aantal ontwikkelingen uit voornamelijk de jaren 10 van deze eeuw. De vele bezuinigingen sinds de jaren 90 hebben de uitvoering enorm onder druk gezet, terwijl deze tegelijkertijd ook steeds verder uit beeld raakte van de departementen en de Tweede Kamer. Tegelijkertijd bleven de uitvoeringsorganisaties kampen met de stapeling van wet- en regelgeving, de last van bestaand beleid en de gevolgen van verouderde ICT-systemen. Voorgenomen beleid werd daarbij niet altijd



getoetst op uitvoerbaarheid en problemen aan de balie kwamen veelal niet terecht bij de mensen die ze konden oplossen. De kinderopvangtoeslagaffaire was hier één van de gevolgen van. Om een dergelijke affaire in de toekomst te voorkomen moest meer aandacht komen voor de uitvoering, waarna het kabinet onder andere de invoeringstoets aankondigde. Bij het ministerie is men vervolgens aan de slag gegaan met de implementatie en invulling van de toets. In 2022 wordt voor het eerst echt invulling gegeven aan het instrument, zijn de eerste invoeringstoetsen afgerond en staan nog eens enkele tientallen toetsen op de planning.

## 3. Benutting van beleidsevaluaties

### 3.1 Inleiding

Dit hoofdstuk bespreekt de belangrijkste inzichten uit de literatuur die van belang zijn voor het borgen van de benutting van de invoeringstoets. De inleiding van dit onderzoek heeft al kort besproken wat de invoeringstoets is en hoe deze op papier moet gaan werken binnen het beleidsproces van de Rijksoverheid. Hierbij zijn ook de drie theoretische deelvragen geïntroduceerd die in dit hoofdstuk centraal staan. Paragraaf 3.2 bespreekt dan ook in meer diepgang wat de invoeringstoets is en vooral hoe deze vervolgens kan gaan werken binnen de beleidsprocessen van de Rijksoverheid. Hierbij worden drie verschillende perspectieven gebruikt om naar het beleidsproces te kijken. Paragraaf 3.3 gaat vervolgens in op wat het concept benutting (van evaluaties) precies betekent en onderscheidt verschillende benaderingen van benutting. Tot slot presenteert paragraaf 3.4 verschillende factoren die van invloed zijn op benutting van evaluaties en adviezen van Nederlandse adviesraden. Van deze factoren wordt een selectie gemaakt die naar verwachting van invloed is op de benutting van de invoeringstoets, waarna deze worden vertaald naar randvoorwaarden voor het borgen van de benutting van de invoeringstoets.

### 3.2 Wat is de invoeringstoets en hoe is deze te positioneren binnen beleidsprocessen?

Voor deze eerste theoretische deelvraag wordt de invoeringstoets verder geduid en ingepast in het beleidsproces. Verschillende perspectieven en modellen voor het reconstrueren van beleidsprocessen worden gebruikt om de werking van de invoeringstoets weer te kunnen geven (Teisman, 1993, pp. 18-24). Door vanuit verschillende perspectieven naar beleidsprocessen te kijken en de invoeringstoets hierin te plaatsen kan een genuanceerd beeld worden geschetst van de werking van het instrument. Om dit te kunnen doen zal na het introduceren van de verschillende perspectieven een definitie van beleidsevaluatie worden gegeven. Aan de hand van de evaluatieliteratuur zal vervolgens vanuit de verschillende perspectieven worden gekeken naar de werking van de invoeringstoets binnen beleidsprocessen.

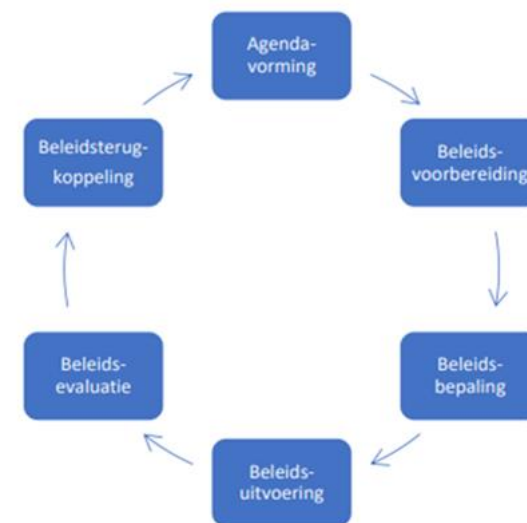
#### 3.2.1 Reconstructies van beleidsprocessen

Om de werking van de invoeringstoets op een goede wijze te kunnen analyseren moet goed worden gekeken naar de context waarin deze moet gaan werken: het beleidsproces. Om dit proces verder in kaart te brengen zal daarom een reconstructie gemaakt moeten worden. Teisman (1993, pp. 18-22) beargumenteert echter dat het maken van een dergelijke reconstructie selectief van aard is en dat de onderzoeker gestuurd wordt door de eigen veronderstellingen over de verschijningsvormen van beleidsprocessen. Hij zegt dus eigenlijk dat iedereen hier een ander beeld bij heeft en dit dus ook van invloed is op hoe zo'n reconstructie wordt gemaakt. Ook bij de analyse van de invoeringstoets moet met verschillende visies rekening gehouden worden. Een punt dat ook aansluit bij het in de introductie genoemde punt van Stone (2012, pp. 11-12) dat het voor accurate beleidsanalyses belangrijk is om vanuit verschillende perspectieven te kijken. Teisman (1993, pp. 18-22) concludeert verder dat beleidsonderzoekers verschillende uitgangspunten hanteren en komt daarmee uit op ruwweg drie modellen om beleidsprocessen mee te reconstrueren: het fasen-, het stromen- en het rondemodell. Door het beleidsproces rondom de invoeringstoets middels alle drie de modellen te reconstrueren kan een genuanceerd beeld worden gevormd. Er kan een analyse worden gemaakt van de

werking van de invoeringstoets zonder in een tunnelvisie te belanden, waarbij bijvoorbeeld alleen vanuit een rationele blik wordt gekeken. Elk model zal eerst uiteen worden gezet, waarna ze worden toegepast op de invoeringstoets en het beleidsproces daaromheen.

### *Het (rationele) fasenmodel*

Het eerste model dat Teisman (1993, pp. 19-20) onderscheidt is het fasenmodel en geeft een vrij rationeel perspectief op beleidsprocessen. Dit model – dat hij baseert op Hoogerwerf (1982) – geeft het beleidsproces namelijk weer als een opeenvolging van verschillende losstaande fasen: de beleidsvoorbereiding, beleidsbepaling, beleidsuitvoering, beleidsevaluatie en de beleidsterugkoppeling. Hoogerwerf (1982, p. 57) hanteert het model als ideaaltype, maar is zich ervan bewust dat de empirie hiervan afwijkt. Wel denkt hij dat het wenselijk is om beleidsprocessen te reconstrueren alsof ze verlopen in fasen. Als men dit kader namelijk normatief op de beleidspraktijk zou toepassen, zouden volgens Hoogerwerf (p. 66) vele tekortkoming aan het licht komen. Een manier waarop het fasenmodel veelal tot uiting komt is aan de hand van de beleidscyclus zoals weergegeven in figuur 2 (Polder, 1996, p. 22; van der Berg et al., 2021, p. 114; Janssen & Hellendoorn, 1999, p. 11; Pattyn, 2020, p. 6; van der Vlis et al., 2007, pp. 20-21; Korsten, 2013a, p. 9). Hierbij wordt het beleidsproces weergegeven als een gefaseerd verlopend, iteratief proces (Polder, 1996, p. 22). Het model kan dus worden gebruikt om beleidsprocessen te vergelijken met het ideaaltype om zo de knelpunten bloot te kunnen leggen



Teisman (1993, pp. 19-20) geeft ook een aantal kritiekpunten op het fasenmodel. Allereerst, bestaan beleidsprocessen – zoals Hoogerwerf (1982, p.57) ook toegeeft – in de praktijk niet uit opeenvolgende, volledig losstaande fasen. Verder zou het indelen van beleidsprocessen in vier of meer fasen niet zinvol zijn, omdat de analyse van de verschillende fasen niet tot specifieke theorievorming zou leiden. Tot slot gaat het model uit van één beleidsbepaler die het uiteindelijke beleid bepaalt. Die rol wordt à priori toegeschreven. Critici ontkennen echter dat dit van tevoren vastgesteld kan worden.

### *Het (organische) stromenmodel*

Waar het fasenmodel beleidsprocessen op rationele wijze omschrijft als opeenvolgende fasen en deze koppelt aan de beleidsbepaler, bestaan beleidsprocessen volgens het stromenmodel uit stromingen die los van elkaar en van een specifieke participant staan (Teisman, 1993, pp. 20-21). Dit organische perspectief is verder doorontwikkeld door Kingdon (1984, p. 20) om het niet alleen toepasbaar te maken op processen binnen organisatorische eenheden, maar ook op inter-organisatorische eenheden.

In deze versie bestaat een stroom van problemen, een beleidsstroom en een politieke stroom die wederom onafhankelijk van elkaar bestaan. Elke stroom ontwikkelt zich los van elkaar en volgens eigen dynamiek en regels, waarmee als het ware drie aparte werelden ontstaan. Hierbinnen worden specifieke producten ontwikkeld en gevormd die niet volgtijdelijk verbonden zijn zoals in het fasenmodel. Binnen de beleidsstroom zijn actoren met oplossingen op zoek naar problemen en politieke commitment. Politici zijn juist op zoek naar problemen en oplossingen waarmee zij kunnen scoren. Volgens het stromenmodel ontstaan pas beleidswijzigingen zodra deze drie stromingen elkaar vinden.

Dit gebeurt voornamelijk zodra er voldoende momentum is. Kingdon noemt dit het 'beleidsraam'. Het model poogt het verloop van beleidsprocessen begrijpelijker te maken door te focussen op hoe koppelingen en dus beleidswijzigingen tot stand komen.

#### *Het (politieke) rondemodel*

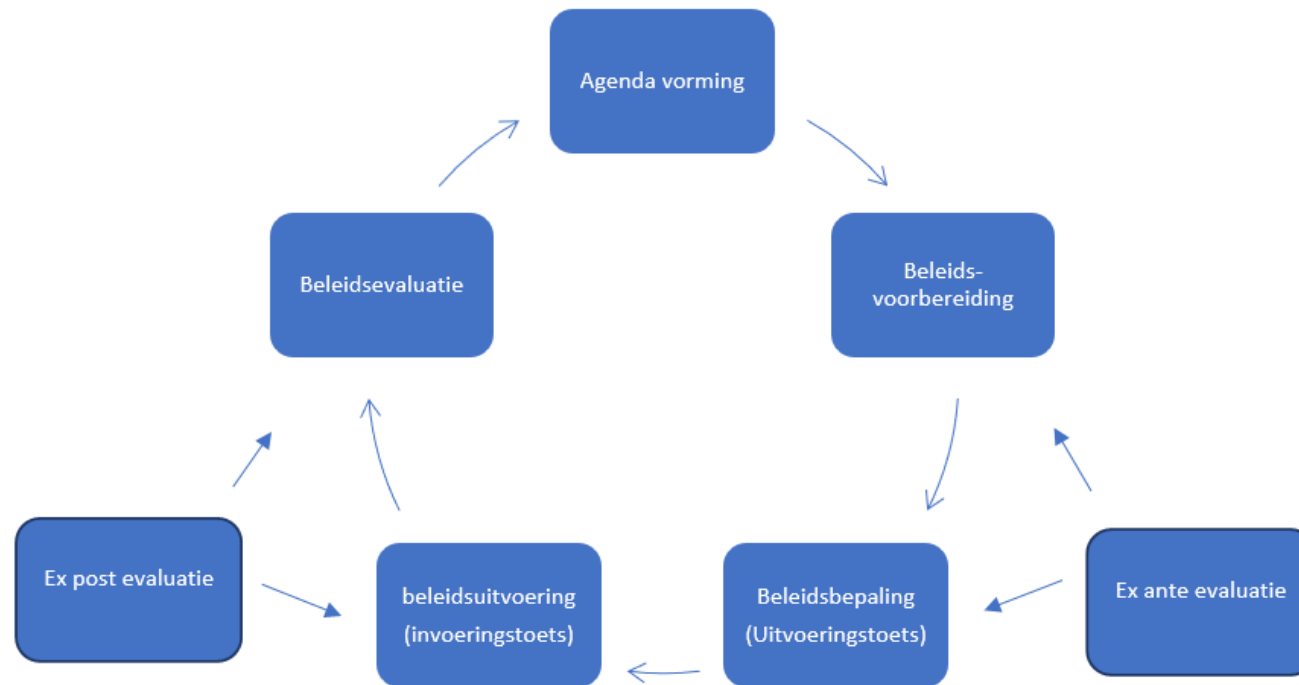
Het laatste model gaat weer uit van een centrale rol van actoren die bepalen welke problemen en oplossingen relevant zijn (Teisman, 1993, pp. 20-21). Echter, waar het fasenmodel uit gaat van één actor en één moment van besluitvorming, gaat het rondemodel uit van vele beleidsbepalers en vele beslissingen. Beleid komt in dit model tot stand in verschillende rondes van rollenspellen waarin actoren zelf beslissen of en wanneer zij initiatieven nemen of voorstellen van anderen aanpassen en overnemen. Het gevormde beleid is het resultaat van een reeks beslissingen, genomen door verschillende actoren in dergelijke rondes. Het beleidsproces is daarmee iteratief, omdat een beslissing aan het eind van de ronde vaak niet het einde is van het proces, maar slechts het startschot voor de volgende ronde. Hierin worden weer nieuwe beslissingen genomen die niet per se in lijn hoeven te liggen met de vorige. Het kan lijken alsof je eruit bent en het beleid staat, maar dat kan in een volgende ronde net zo goed weer tenietgedaan worden. Actoren hebben daarbij een eigen doel en perceptie van in welke fase zij zich bevinden. Eén actor kan zich bezighouden met de evaluatie van een wet, terwijl een andere die evaluatie gebruikt als argument om een besluit tot ander beleid door te duwen. De interactie tussen de verschillende actoren bepaalt vervolgens het resultaat van het beleidsproces. Centraal staat dan ook het doel en de strategie die de verschillende actoren kiezen en hoe zij deze aanpassen op de interacties die zij hebben met andere actoren. Daarmee heeft het model een sterk politiek karakter. Het rondemodel probeert beleidsprocessen te duiden door te focussen op het gedrag en de interactie van actoren. Het gaat hierbij niet zozeer om de juistheid van de strategie, de interactie of de perceptie, maar om de discongruenties in de verschillende strategieën (Teisman, 1993, pp. 20-21).

### 3.2.2 Beleidsevaluatie

De invoeringstoets is – zoals in de inleiding al is benoemd - een vorm van beleidsevaluatie. Beleidsevaluatie door een overheid slaat volgens Korsten (2013a) op het '*op een systematische manier beoordelen van overheidsbeleid*' (p. 15). Een andere definitie is die van Bressers en Hoogerwerf (1995): '*Beleidsevaluatie van overheidsbeleid is het beoordelen van beleidsprocessen – zoals bijvoorbeeld de probleemanalyse of de totstandkoming van beleid-, de beleidsinhoud, de organisatie van het beleid of effecten aan de hand van bepaalde maatstaven of criteria*' (p. 1). Hieruit kan worden opgemaakt dat bij een beleidsevaluatie verschillende keuzes gemaakt kunnen worden en op verschillende wijze focus aangebracht kan worden. Er bestaat daarom ook niet één vorm van beleidsevaluatie.

#### *Ex ante en ex post beleidsevaluatie*

Veelal wordt een simpel onderscheid gemaakt tussen de ex ante evaluatie die voorafgaand aan de besluitvorming plaatsvindt en de ex post evaluatie die na volledige inwerkingtreding of uitvoering van beleid plaatsvindt (Janssen & Hellendoorn, 1990, p. 11; Korsten, 2013b, p. 56; Winter, 2014, p. 156). Bij ex ante evaluatie wordt voorafgaand aan besluitvorming over beleid onderzocht welke (on)gewenste gevolgen er zijn van het beleid zelf en van (meerdere) beleidsalternatieven. Bij ex post evaluaties wordt het beleid juist na volledige inwerkingtreding beoordeeld. In de beleidscyclus van het fasenmodel zou dit weergegeven kunnen worden zoals in onderstaande figuur.



### *Ex durante beleidsevaluatie*

Naast de ex ante en ex post evaluatie wordt in de literatuur, in mindere mate, nóg een evaluatievorm genoemd, namelijk de ex durante evaluatie. De invoeringstoets wordt volgens de Kamerbrief van minister Weerwind als dusdanig gekenmerkt (2022, p. 2). Het wordt hierin uitgelegd als een vorm van evalueren die beoordeelt wat er voor een goede werking van regelgeving nog *moet* gebeuren als corrigeren nog mogelijk is, in plaats van wat er *had moeten* gebeuren. Ook van der Vlis et al. (2007, p. 20) en Pattyn (2020, p. 6) benoemen deze vorm. Van der Vlis et al. (2007, p. 20) omschrijven het als een vorm van evaluatie die plaatsvindt gedurende de looptijd van het beleid. Het doel is om de eerste prestaties te meten en te leren van de opgedane ervaringen zodat noodzakelijke aanpassingen kunnen worden gedaan gedurende de resterende looptijd. Daarnaast kunnen de geleerde lessen worden meegenomen bij de voorbereiding van nieuw beleid. Pattyn (2020) stelt daarnaast dat ex durante beleidsevaluaties ‘worden uitgevoerd om zicht te krijgen op de implementatie van beleid of op mogelijke verbeteringen, wanneer er bijvoorbeeld aanwijzingen zijn dat de uitvoering hapert en/of de doelstellingen niet bereikt worden. Ook kunnen ex durante evaluaties nuttig zijn om een vroeg inzicht te krijgen in de effecten van beleid, inclusief eventuele positieve of negatieve neveneffecten.’

(p. 6). Waar de ex ante evaluatie dus vóór de besluitvorming en de ex post evaluatie ná volledige inwerkingtreding of uitvoering van beleid plaatsvindt, peilt de ex durante evaluatie al tijdens de beleidsuitvoering of het beleid hout snijdt.

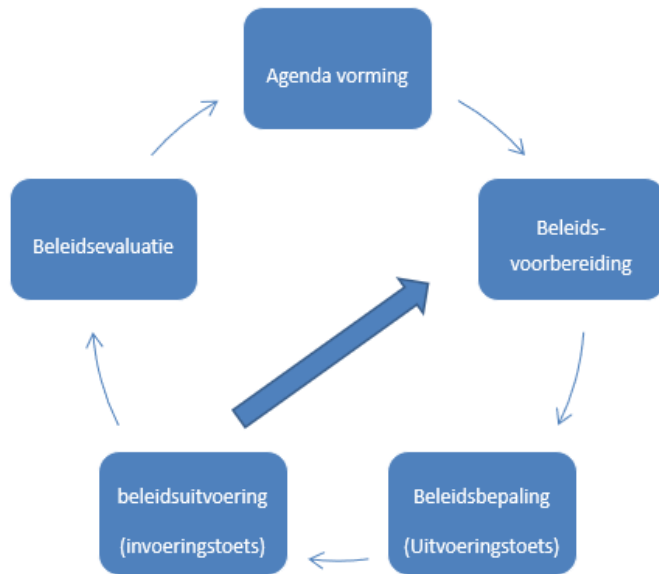
### 3.2.3 De invoeringstoets vanuit het rationele, organische en politieke perspectief

Nu verschillende manieren zijn gepresenteerd voor het reconstrueren van het beleidsproces en het concept beleidsevaluatie uiteen is gezet, wordt duidelijk wat de invoeringstoets precies kan zijn. Eerder is al besproken hoe vanuit verschillende perspectieven naar beleidsprocessen kan worden gekeken en op basis waarvan het verloop ervan anders kan worden verklaard. Deze zelfde perspectieven – het rationele, organische en politieke perspectief – kunnen ook gebruikt worden om verschillend naar evaluaties en specifiek de invoeringstoets te kijken. De vraag wat de invoeringstoets precies is en hoe deze te positioneren is binnen beleidsprocessen kan daarom ook vanuit elk van deze perspectieven worden beantwoord en een ander resultaat opleveren. Hetzelfde geldt overigens voor de hoofdvraag van dit onderzoek.

#### *De invoeringstoets vanuit rationeel perspectief*

Zodra vanuit een rationeel perspectief wordt gekeken naar de invoeringstoets kan het idee erachter beter worden omschreven met behulp van het fasenmodel en de beleidscyclus. Vanuit het rationele perspectief is het verloop van beleidsprocessen namelijk duidelijk en à priori vast te stellen. De invoeringstoets vindt daardoor standaard plaats op hetzelfde moment binnen het beleidsproces, oftewel: in dezelfde fase van de beleidscyclus.

In de Kamerbrief van Weerwind wordt de invoeringstoets beschreven als een instrument waarmee op lichtvoetige wijze de werking van nieuwe regelgeving in de praktijk bestudeerd kan worden (Weerwind, 2022-a, p. 2). Het wordt omschreven als een ex durante evaluatie waarbij vroegtijdig in de uitvoering wordt gekeken naar hoe de regeling uitpakt voor de doelgroep en voor de uitvoering. Het idee is dat mogelijke knelpunten al snel gesignaleerd en opgelost kunnen worden en dat deze niet pas aan het licht komen bij een volledige evaluatie.



Als gekeken wordt naar de andere evaluatie-instrumenten die doorgaans worden uitgevoerd bij de Rijksoverheid dan valt op dat deze allemaal het karakter van ex ante of ex post lijken te hebben. Korsten (2013b) heeft het bijvoorbeeld over de *klassieke beleidsevaluatie* (ex post) met daarbinnen de *doelbereikingsanalyse* en de *oorzakelijkheidsanalyse* (ex post), de *doelstellingsvrije beleidsevaluatie* (ex post), de *inputevaluatie* (ex ante), de *procesevaluatie* (ex ante/post), de *efficiency-evaluatie* (ex ante/post), de *retrospectieve* (ex post) en de *prospectieve evaluatie* (ex post). Met de ex ante evaluaties wordt vooraf gekeken naar de mogelijke werking van wet- en regelgeving en wordt het hierop aangepast. Een goed beeld van de praktijk is er dan echter nog niet. Bij de ex post evaluaties worden knelpunten juist relatief laat gesignaleerd, waarbij wetten vaak ook weer opnieuw de gehele beleidscyclus moeten doorlopen om tot verandering te komen. Het voordeel van de invoeringstoets is dat deze uitgevoerd wordt zodra er (waarschijnlijk) nog aandacht en nieuwsgierigheid is naar de werking van de wet- en regelgeving. Hierdoor kunnen snel aanpassing worden gedaan zonder dat een diepgaande analyse uitgevoerd hoeft te worden en de regeling weer opnieuw de gehele cyclus moet doorlopen. Hiermee fungeert het in theorie als een soort afsnijroute van de beleidsuitvoering direct terug naar de beleidsvoorbereiding.

### *De invoeringstoets vanuit organisch perspectief*

Helaas zou het naïef zijn om te verwachten dat de invoeringstoets ook precies zo gaat werken als beoogd. Niet voor niets krijgt het fasenmodel kritiek over het feit dat het een te simpele weergave van de werkelijkheid is, en dat het beleidsproces helemaal niet zo gefaseerd verloopt. Zodra minder rationeel gekeken wordt en juist meer een organische bril wordt opgezet, dan komt men op een veel chaotischer beeld uit waarbij beleidsdossiers veel afhankelijker zijn van andere beleidsdossiers, politieke commitment, momentum en andere omstandigheden. Vanuit dit perspectief heeft de invoeringstoets ook geen 'vast' moment binnen het beleidsproces, maar vindt deze plaats zodra de omstandigheden het toelaten. Volgens het stromenmodel is dit wanneer de drie stromingen van problemen, oplossingen en politieke commitment bij elkaar komen en een beleidsraam ontstaat. Zodra de omstandigheden weer veranderen of het beleidsraam weer dicht is kan het ook heel moeilijk worden om nog iets voor elkaar te krijgen. Dit roept bijvoorbeeld de vraag op in hoeverre de invoeringstoets als instrument in staat is om nog tot wijzigingen te komen als het beleid al een jaar in uitvoering is en het beleidsraam al dicht is. In dit kader zou je ook kunnen stellen dat de invoeringstoets een instrument is dat de drie stromingen bij elkaar moet brengen, bijvoorbeeld door op het juiste moment te verschijnen, of door actoren welbewust bij elkaar te brengen. Door structureel ongeveer een jaar na inwerkingtreding terug te blikken op beleid en dit middels de invoeringstoets terug te leggen bij de beleidsmakers, krijgen zij beter zicht op ontstane problemen. Deze kunnen zij vervolgens (in theorie) ook

weer onder de aandacht krijgen van de politiek. De vraag rondom commitment blijft dan staan. Zolang de beleidsmakers en betrokken politieke actoren niet ontvankelijke zijn voor de invoeringstoets zullen de signalen alsnog niet aankomen en wordt het een zinloze exercitie.

#### *De invoeringstoets vanuit het politieke perspectief*

Tot slot kan naar de invoeringstoets worden gekeken vanuit een politiek perspectief, waarbij de vraag rondom commitment nog scherper neergezet wordt. Volgens het rondommodel zijn actoren vrij om zelf te bepalen wanneer zij initiatieven nemen of voorstellen van anderen aanpassen en overnemen. De actoren kiezen zelf een strategie en passen deze aan op basis van het handelen van anderen. Daarnaast wordt benadrukt dat verschillende actoren niet per se in dezelfde fase van het beleidsproces hoeven te zitten.

Als vanuit dit perspectief naar de invoeringstoets wordt gekeken dan kunnen de twijfels over de beoogde werking nog sterker worden. Verschillende actoren kunnen namelijk invloed uitoefenen op een beleidsdossier en daarmee ook op een eventuele beleidswijziging. Zo kan het bijvoorbeeld zijn dat een bepaald dossier zeer politiek-geladen is en dat na vele debatten, onderhandelingen en achterkamertjespolitiek beleid tot stand is gekomen. Dan denkt men eruit te zijn, wordt er geëvalueerd en dan begint het hele proces weer van vooraf aan, omdat het middels de invoeringstoets weer onder de aandacht komt. Of wellicht proberen die actoren invloed uit te oefenen op de evaluatie zelf of schuiven plotseling hele nieuwe actoren aan in de discussie met weer hele nieuwe ideeën. Misschien wordt de evaluatie in zijn geheel genegeerd of is er überhaupt geen aandacht voor omdat het midden in een politieke crisis wordt gepubliceerd. Zelfs de evaluator(en) zelf zou(den) andere conclusies in de evaluatie kunnen opnemen of deze op een ongunstig moment kunnen publiceren om bijvoorbeeld de eigen bewindspersoon te beschermen. Kortom: de verschillende actoren kunnen allerlei redenen hebben om dwars te liggen bij een wijziging van beleid die in veel gevallen ook niet voorzien zullen zijn.

#### 3.2.4 Conclusie en verwachtingen

Voor de eerste empirische deelvraag is vanuit verschillende perspectieven gekeken naar de invoeringstoets en de werking ervan binnen beleidsprocessen. Daarmee kan nu de eerste theoretische deelvraag worden beantwoord. De invoeringstoets is een nieuw ex durante evaluatie-instrument waarmee ongeveer een jaar na inwerkingtreding van wet- en regelgeving teruggeblikt wordt op de werking ervan in de praktijk. Vanuit een rationele blik moet het instrument knelpunten vanuit de beleidsuitvoering direct terugleggen bij de beleidsvoorbereiding zodat het beleid of de wet- en regelgeving ook direct aangepast kan worden om zo ook de knelpunten te verhelpen. Vanuit een meer organische of politieke blik lijkt deze gedachtegang echter te kort door de bocht. De invoeringstoets bevindt zich in een dynamische context waarin veel belangen een rol spelen. De werking van het instrument is daardoor enerzijds sterk afhankelijk van de context en omstandigheden waarin deze zich bevindt. Hierbij spelen andere beleidsdossiers, politieke commitment, momentum een rol en blijft het ook de vraag of er een jaar na inwerkingtreding nog voldoende momentum is of momentum kan worden gevonden om beleid nog te wijzigen. Anderzijds is de werking van het instrument afhankelijk van het handelen van de verschillende actoren en de doelen die zij nastreven. Het idee achter het instrument is echter wel passend. Juist door beleid en wet- en regelgeving structureel te evalueren en ook in te bouwen dat de conclusies hiervan structureel teruggelegd worden bij de beleidsmaker, kan structurele herziening worden bewerkstelligd. Indien nodig kan de wet of regeling dan worden aangepast om de knelpunten te verhelpen.



Op basis van wat over de invoeringstoets is geschreven in de Kamerbrief van Weerwind (2022) en andere officiële publicaties is de verwachting dat de focus bij de implementatie te veel ligt op de rationele functie van het instrument. Het gaat hierin voornamelijk over hoe de toets wet- en regelgeving weer onder de aandacht van de beleidsmaker kan krijgen zodra hier problemen ontstaan en hoe dit vervolgens tot herziening kan leiden. De toets wordt echter geïmplementeerd in een hectische context waar zeer veel belangen van eveneens zeer veel actoren van invloed zijn. Niet alleen kunnen de hierboven genoemde omstandigheden de werking van de toets in de weg liggen, maar kunnen ook verschillende actoren vanuit hun persoonlijke doelen dwars gaan liggen. De verwachting is dus ook dat bij de implementatie van de invoeringstoets de nadruk te veel ligt op het rationele en dat ook organische en politieke factoren een rol spelen.

### 3.3 Wat is benutting van evaluatie en welke benaderingen voor benutting zijn te onderscheiden?

Bij deze tweede theoretische deelvraag zal het concept benutting van evaluatie worden besproken, waarbij gepoogd zal worden om het concept te definiëren en te operationaliseren. Hiertoe zal eerst worden ingegaan op een aantal gerelateerde begrippen om vervolgens aan de hand van een aantal modellen te proberen tot een bruikbare definitie te komen. Ten slotte zal met het oog op de operationalisering van het concept een aantal vormen waarop benutting van evaluatie kan voorkomen worden besproken.

#### 3.3.1 Onduidelijkheid over de definitie van benutting

Over het concept benutting en de daaraan gerelateerde concepten gebruik en doorwerking, is sinds de jaren 60 (met name in Amerika) veel geschreven (Leeuw & Gils, 2010, p. 165; Winter, Scheltema & Herweijer, 1990, p. 50; Bekkers, Fenger, Homburg & Putters, 2004, p. 11). Prof. Arno Korsten was één van de eersten die zich in Nederland bezighield met het onderwerp van benutting (Leeuw & Gils, 2010, p. 165) en publiceerde in 1983 zijn boek *'Wat is goed genoeg? Benutting van onderzoek naar overheidsbeleid.'* Hierin stelt hij dat een eenduidige operationele definitie voor het concept benutting niet bestaat. Daar komt bij dat verschillende terminologie door elkaar gebruikt wordt. Benutting, doorwerking, gebruik, *use* en *utilization* zijn allemaal termen die gebruikt worden door auteurs als Korsten (1983; 2013a), Leeuw & Gils (2010), Derksen & Fleurke (2010), Smet (2013), Winter et al. (1990), Bekkers et al. (2004) en Cousins & Leithwood (1986, p. 332), en waar ruwweg hetzelfde mee bedoeld wordt. De termen lijken dus onderling uitwisselbaar te zijn, terwijl een eenduidige definitie ontbreekt.

#### 3.3.2 De zeven betekenissen van benutting

Ook Weiss (1979) constateerde een gebrek aan consensus over een definitie. Weiss vond het belangrijk om meer inzicht te krijgen in hoe de sociale wetenschap tot dan toe bijgedragen had aan publiek beleid en hoe deze bijdrage in de toekomst vergroot kon worden. Om de basis hiervoor te leggen trachtte zij om de verschillende rondgaande definities te ontleden. Zij kwam hiertoe tot zeven betekenissen (die zij modellen noemt) die met het concept geassocieerd werden. Dit onderscheid is later door veel verschillende auteurs overgenomen (o.a. Korsten, 1983, p. 34-43; Winter et al., 1990, p. 50-52; Bekkers et al., 2004, p. 14-18). In aanvulling op beleidsprocessen en beleidsevaluaties kan ook benutting als rationeel, organisch of politiek worden gekenmerkt. De 7 modellen zullen daarom aan de hand van deze perspectieven worden uitgelegd.

##### *Rationele modellen*

De eerste twee modellen die genoemd worden zijn rationeel van aard. Het *'knowledge-driven'* of het *kennistoepassingsmodel* stamt voornamelijk af uit de natuurwetenschappen (Weiss, 1979, p. 427; Winter et al. 1990, p. 50). In de context van advisering betekent doorwerking

volgens dit model dat het beschikbaar komen van kennis uit een advies leidt tot gebruik ervan in de besluitvorming (Bekkers et al. 2004, p. 15), mits het advies aan bepaalde kwaliteitseisen voldoet. Het model gaat er dus vanuit dat kennis ook gebruikt wordt in de besluitvorming en dus van een lineair proces. Het *probleemoplossingsmodel* lijkt hier sterk op en is volgens Weiss (1979, p. 427) het meest voorkomend in de sociale wetenschappen. Het gaat namelijk uit van een bepaald beleidsprobleem met een tekort aan kennis om tot een oplossing te kunnen komen. Onderzoek moet de benodigde (empirische) kennis opleveren om het probleem op te kunnen lossen (Winter et al., 1990, p. 50).

#### *Organische modellen*

De volgende modellen gaan uit van beïnvloeding van het gebruik door de omgeving en omstandigheden en hebben dus een meer organisch karakter. Te beginnen met het *interactiemodel* dat meer oog heeft voor de verscheidenheid aan kennis die van invloed kan zijn op de besluitvorming (Weiss, 1979, p. 428; Winter et al., 1990, p. 51; Bekkers et al., 2004, p. 17; Korsten, 1983, p. 19). Beleidsvoerders hebben te maken met meerdere informatie-eenheden en informatiebronnen, waar resultaten van onderzoek er slechts één van is. Daarnaast kunnen ervaring, politiek inzicht, druk, inspraak en overleg van invloed zijn. Onderzoekresultaten worden dus niet an sich benut, maar zijn in combinatie met andere soorten kennis van invloed op de besluitvorming.

In het *'enlightenment'* model houdt benutting in dat sociaalwetenschappelijke concepten en theorieën in verloop van tijd hun weg vinden in de samenleving (Weiss, 1979, p. 429). Het gaat dan om concepten en theorieën die langzaam maar zeker bekender en geaccepteerder worden, waardoor ze uiteindelijk (onbewust) door besluitvormers worden gebruikt. Het idee is dat sociaalwetenschappelijke generalisaties en oriëntaties doordringen tot de maatschappij en vorm gaan geven aan de manier waarop mensen denken over sociale kwesties.

Het laatste organische model is *'onderzoek als onderdeel van de intellectuele onderneming van de samenleving'* en beschouwt sociaal wetenschappelijk onderzoek als afhankelijk van de heersende denkpatronen en trends in de samenleving (Weiss, 1979, p. 430; Winter et al., 1990, p. 51). Beleid en onderzoek beïnvloeden elkaar, maar ze worden beiden ook beïnvloed door de collectieve interesse in bepaalde maatschappelijke problemen. Zodra er in de maatschappij veel te doen is om een bepaald beleidsonderwerp zal hier vanzelf meer onderzoek naar worden gedaan wat uiteindelijk zal leiden tot herziening van de kwestie door besluitvormers. Anderzijds kan het ook zijn dat – door de maatschappelijke interesse - besluitvormers een beleidsonderwerp herzien en vervolgens juist meer onderzoek naar het onderwerp gedaan wordt.

#### *Politieke modellen*

Tot slot worden nog twee modellen onderscheiden die als politiek kunnen worden kenmerkt. De eerste van de twee wordt ook letterlijk het politieke model genoemd en gaat ervan uit dat in een aantal gevallen de besluitvormers vooraf al positie hebben ingenomen bij een bepaald beleidsprobleem door alle factoren die hierop van invloed zijn (Weiss, 1979, p. 428; Korsten, 1983, p. 19; Winter et al., 1990, p. 51; Bekkers et al., 2004, p. 17). Besluitvormers staan dan niet meer open voor nieuw bewijs uit sociaal onderzoek. Onderzoeken en evaluaties worden dan voornamelijk verricht om bestaand beleid te kunnen rechtvaardigen of juist om het standpunt van politieke tegenstanders te weerleggen, zelfs als de resultaten daarvoor uit verband getrokken moeten worden.

Ook kan het voorkomen dat het doel achter sociaalwetenschappelijk weinig te maken heeft met de inhoud ervan. Dergelijke gevallen worden omschreven in het *Tactische model*. Hierbij gaat het puur om het feit dat het onderzoek gedaan wordt en niet om de resultaten (Weiss, 1979, p.

429). Dit kan bijvoorbeeld gaan om een overheidsorgaan dat wil kunnen laten zien met een bepaald beleidsprobleem bezig te zijn of kritiek probeert te pareren door te beweren dat het handelen gebaseerd is op implicaties of aanbevelingen uit onderzoek.

Rationeel	Kennistoepassingsmodel	Kennis uit een advies of onderzoek leidt tot direct gebruik ervan in de besluitvorming
	Probleemoplossingenmodel	Gebrek aan kennis van een bepaald beleidsprobleem leidt tot onderzoek/een adviestraject waarmee het probleem opgelost moet kunnen worden
Organisch	Interactiemodel	Interactie met onderzoekers is slechts één van de vele informatiebronnen die van invloed is op de beeldvorming van besluitvormers. Veelvuldige en diepgaande interactie met onderzoekers draagt bij aan gebruik van de resultaten
	'Enlightenment' model	Sociaalwetenschappelijke concepten en theorieën vinden in verloop van tijd hun weg in de samenleving, waardoor ze (onbewust) door besluitvormers worden gebruikt.
	Onderzoek als onderdeel van de intellectuele onderneming van de samenleving	Beleid en onderzoek zijn (deels) afhankelijk van elkaar, maar ook van de heersende denkpatronen en trends in de samenleving. Benutting van onderzoek in beleid is dus tot zekere hoogte afhankelijk van dominante perspectieven in de samenleving
Politiek	Politieke model	Onderzoek wordt door besluitvormers slechts argumentatief gebruikt om bijvoorbeeld beleid te rechtvaardigen of tegenargumenten te weerleggen
	Tactische model	Onderzoek wordt puur uitgevoerd om tactische redenen en niet vanwege de inhoud, bijvoorbeeld om te laten zien met een onderwerp bezig te zijn of te beweren dat het eigen handelen gebaseerd is op onderzoek

Dit zijn de verschillende betekenissen die gebruik volgens Weiss (1979, p. 430) aan kan nemen. Elke betekenis zal in sommige situaties van toepassing zijn maar geen één van deze betekenissen kan volgens haar een volledig antwoord geven op hoe een staat het best haar onderzoeksmiddelen kan benutten. Ze benadrukt ten slotte wel het nut van het denken over de verschillende perspectieven van benutting. Het laat namelijk het nut zien van sociaalwetenschappelijk onderzoek, ook al heeft het in veel gevallen niet direct invloed zoals bij het probleemoplossende model.

### 3.3.3 Verschillende vormen van doorwerking

Uit de verschillende definities voor het concept gebruik die Weiss in 1979 introduceerde wordt al duidelijk dat onderzoek en evaluatie op verschillende wijzen kan doorwerken. Volgens Bekkers (2004, p. 21) wordt in discussies over doorwerking van advisering vaak impliciet een instrumenteel (rationeel) model van doorwerking genomen. Er wordt dus van uitgegaan dat adviezen een directe invloed hebben op de besluitvorming. Volgens Bekkers et al. komt deze vorm echter zelden voor en zijn andere vormen zoals 'enlightenment' daarom ook zeer belangrijk om te beschouwen.

Het onderkennen van deze onderscheidingen betekent volgens Bekkers et al. (2004, p. 21) ook dat geen eenduidige wijze van doorwerking vast te stellen valt, maar dat voor verschillende typen adviezen verschillende maatstaven moeten worden aangelegd. Bij het vaststellen van doorwerking moet daarom rekening gehouden worden met zowel het type advies als de doelgroep waar de doorwerking zich op richt. In dit onderzoek draait het niet om het vaststellen van doorwerking, maar wel om het vinden van randvoorwaarden voor het borgen van doorwerking. Ook hierbij is het van belang om een onderscheid te maken tussen de vormen van doorwerking, omdat een randvoorwaarde voor benutting waarbij een onderwerp wordt geagendeerd niet per definitie een randvoorwaarde is voor benutting waarbij beleid veranderd wordt. Bekkers et al (2004, pp. 23-30) onderscheidt hierbij vier vormen waarop adviezen kunnen doorwerken.

#### *Vier vormen van doorwerking*

Bekkers et al. (2004, pp. 24-26) maken dus een onderscheid in verschillende vormen waarop adviezen effect kunnen hebben. Allereerst noemen zij de instrumentele vorm van doorwerking waarbij advies direct effect heeft op het gedrag van individuen of organisaties. Zoals hierboven al wordt genoemd wordt vaak uitgegaan van deze vorm van doorwerking, terwijl juist ook vooral andere vormen voorkomen. Conceptuele doorwerking wordt bijvoorbeeld als een veel belangrijkere functie afgestempeld. Hierbij werken adviezen en aanbevelingen door tot de dominante denkbeelden en opvattingen van individuen of organisaties. Het betekent op langere termijn een verandering in discours. Daarnaast kan doorwerking in agenderende vorm voor komen. In dat geval zorgen adviezen of aanbevelingen niet direct voor verandering en ook niet pas na een langere periode, maar wordt het beleidsthema wel opnieuw onderdeel van gesprek. Dit kan zowel in het maatschappelijke als het politieke debat het geval zijn. Tot slot wordt politiek-strategische doorwerking onderscheiden. Bij deze vorm worden adviezen of aanbevelingen meer strategisch gebruikt voor het bereiken van politieke doelen door de machtspositie van bepaalde actoren te versterken.

Aangezien de verschillende vormen van doorwerking sterk verband houden met de drie centrale perspectieven zal een vertaalslag gemaakt worden om ze bruikbaar te maken voor het onderzoek. De instrumentele vorm van doorwerking heeft namelijk een sterk rationeel perspectief omdat het ervan uit gaat dat kennis per definitie leidt tot benutting. De conceptuele vorm is juist veel meer context- en omstandighedenafhankelijk en dus organisch. Tot slot zijn de agenderende en politiek-strategische vorm zeer politiek van aard.

<b>Perspectieven</b>	<b>Vormen van doorwerking</b>	<b>Omschrijving</b>
Rationeel	Instrumentele doorwerking	Het advies leidt tot een directe verandering van gedrag van individuen of organisaties in lijn met de aanbevelingen van het advies
Organisch	Conceptuele doorwerking	Het advies heeft een verandering in de kennis, opvattingen of causale redeneringen van individuen of organisaties tot gevolg
Politieke (en) doorwerking	Agenderende doorwerking	Het advies wordt gebruikt om een nieuw onderwerp of beleidsthema onderwerp van het maatschappelijke of politieke debat te maken of

	Politiek-strategische doorwerking	Het advies wordt gebruikt om de machtspositie van één of enkele spelers te versterken om op deze wijze de politieke doelen beter te kunnen bereiken
--	-----------------------------------	---

### 3.3.4 Conclusie en verwachtingen

Om tot een antwoord te komen op de tweede theoretische deelvraag: voor het concept benutting bestaat geen eenduidige definitie en de term wordt tevens afgewisseld door de termen gebruik en doorwerking. In plaats van een eenduidige definitie worden doorgaans zeven modellen gebruikt als de verschillende betekenissen van de term (o.a. Korsten, 1983, pp. 34-43; Winter et al., 1990, pp. 50-52; Bekkers et al., 2004, pp. 14-18), die aansluiten bij en daarom ook gereduceerd zijn tot rationele, organische en politieke definities van doorwerking. Om benutting beter te kunnen analyseren wordt door Bekkers et al. (2004, pp. 24-26) verder een onderscheid gemaakt tussen vier verschillende vormen waarop doorwerking kan plaatsvinden: instrumentele, conceptuele, agenderende en politiek-strategische doorwerking.

De verwachting voor het empirische deel van het onderzoek is dat wederom te rationeel naar het instrument wordt gekeken en dat dus een té grote nadruk ligt op instrumentele doorwerking. Ook hiervoor geldt dat in de officiële berichtgeving over de toets vooral herziening en wijziging worden genoemd als doel van de invoeringstoets. Daarbij wordt ook het terugleggen van signalen uit de uitvoering bij de beleidsmakers genoemd als belangrijke functie, waardoor ook agenderende doorwerking erkend wordt als doel van het instrument. In de brief staat echter wel dat gewaakt moet worden voor een te instrumentele benadering, maar dat wordt vervolgens uitgelegd als dat het geen doel op zich moet worden en dat verstandig en doelmatig met het instrument om moet worden gegaan. Die te instrumentele benadering gaat dus alleen maar over de gebruikte methode, maar niet zozeer over de benutting van het instrument. Conceptuele en politiek-strategische doorwerking worden dan ook volledig achterwegen gelaten. In paragraaf 3.2 is uiteengezet hoe de invoeringstoets wordt geïmplementeerd in een dynamische context waar veel belangen een rol spelen. Met name dit laatste punt lijkt daarmee ook zeer relevant voor de manier waarop de invoeringstoets zal doorwerken. Stakeholders streven allemaal andere doelen na en zullen ook die invoeringstoets benutten op een manier die het eigen belang dient. Tot slot zal de invoeringstoets ook regelmatig de conclusie hebben dat het beleid goed loopt en geen knelpunten zijn gevonden. In dat geval heeft het juist een informerende functie en vindt vooral conceptuele doorwerking plaats. De verwachting is dan ook dat zeker politiek-strategische, maar ook conceptuele doorwerking vaker voor zal komen dan officiële berichtgeving nu doet geloven.

### 3.4 Welke randvoorwaarden voor het borgen van benutting van de invoeringstoets zijn te onderscheiden?

Deze laatste theoretische deelvraag heeft als doel om vanuit de wetenschappelijke literatuur tot randvoorwaarden voor het borgen van benutting van de invoeringstoets te komen. Hiertoe zal in deze paragraaf eerst een aantal problemen bij het beantwoorden van bovenstaande vraag worden benoemd, waarbij wordt aangegeven hoe hiermee wordt omgegaan. Vervolgens zal de literatuurbespreking van Johnson et al. (2009) over gebruik van evaluaties worden besproken, waarin ook een lijst met factoren die van invloed zijn op gebruik is opgenomen. Tot slot zal besproken worden of/welke van deze factoren ook toepasselijk zouden kunnen zijn voor deze derde theoretische deelvraag, waarna de factoren worden vertaald naar randvoorwaarden.

Het beantwoorden van de vraag welke randvoorwaarden voor het borgen van benutting van de invoeringstoets in de literatuur te onderscheiden zijn, is lastig om drie redenen. Ten eerste is de invoeringstoets een nieuw instrument dat nog grotendeels vormgegeven moet worden, waardoor dus logischerwijs nog niet of nauwelijks onderzoek is gedaan over het onderwerp. Laat staan onderzoek dat specifiek is gericht op de randvoorwaarden voor benutting van de invoeringstoets.

Het tweede probleem zit in het woord 'borgen'. Er is veel onderzoek gedaan naar factoren die van invloed zijn op benutting (Johnson et al., 2009), maar dit gaat voornamelijk over specifieke evaluaties. Dergelijke evaluaties worden eenmalig uitgevoerd waardoor de benutting niet systematisch 'geborgd' hoeft te worden. In dit onderzoek gaat het om een beleidsevaluatie-instrument waarmee systematisch wet- en regelgeving geëvalueerd moet gaan worden. In dit geval gaat het er dus om dat de benutting als het ware ingebouwd wordt in het beleidsproces en dus geborgd wordt, zodat invoeringstoetsen ook op lange termijn nog voldoende benut blijven worden.

Het laatste punt is in mindere mate problematisch, maar toch goed om te benoemen. Dit gaat namelijk over het feit dat er in dit onderzoek gezocht wordt naar 'voorwaarden' voor 'benutting', terwijl er in de literatuur voornamelijk gesproken wordt over 'factoren van invloed' en voor de term 'benutting' vaak ook 'gebruik' of 'doorwerking' wordt gebruikt (Johnson et al., 2009). Om deze problemen te omzeilen is gezocht naar bronnen met (1) factoren die van invloed zijn op (2) de benutting/het gebruik/de doorwerking van (3) (bij voorkeur beleids-)evaluaties.

#### 3.4.1 Literatuur met factoren van invloed op gebruik

Over gebruik van beleidsevaluaties is – in tegenstelling tot de invoeringstoets – in de literatuur veel te vinden. Er zijn zelfs twee generaties aan extensieve literatuurstudies naar gedaan: Cousins & Leithwood (1986) over de jaren 1971 tot en met 1985 en Johnson et al. (2009) over de jaren 1986 tot 2005. Naast deze twee literatuurbesprekingen die zich focussen op Engelstalige geschreven artikelen, heeft ook Bekkers et al. (2004) zich gericht op gebruik, maar dan van adviezen van Nederlandse adviesraden. Voor dit onderzoek zal gebruik worden gemaakt van zowel de factoren van Johnson et al. (2009), als van Bekkers et al. (2004). Op deze manier kunnen namelijk zowel de uitgebreide opzet van Johnson et al. (2009) als de toespitsing op de Nederlandse context door Bekkers et al. (2004) worden meegenomen. Cousins & Leithwood (1986) zal wel buiten beschouwing worden gelaten aangezien Johnson et al. (2009) hier met recentere onderzoeken op doorgaat.

Om tot randvoorwaarden voor het borgen van benutting van de invoeringstoets te komen, worden de factoren van Johnson et al. (2009) en Bekkers et al. (2004) samengevoegd. Beide artikelen zullen daarvoor eerst kort worden toegelicht waarna de meest relevante factoren van beide artikelen ingedeeld worden in nieuwe categorieën die passen bij de context van de invoeringstoets, namelijk: factoren met betrekking tot het evaluatieproces; het evaluatierapport en de presentatie en nazorg ervan; de evaluator en ontvanger van de evaluatie; overige stakeholders

en de grotere politiek-bestuurlijke context. In paragraaf 3.3.2. wordt uiteengezet hoe bepaald wordt welke factoren voor dit onderzoek relevant zijn.

Allereerst geven Johnson et al. (2009) een overzicht van 41 Engelstalige studies uitgevoerd in de periode 1986 tot en met 2005. Uit deze 41 studies hebben zij in totaal 22 factoren gehaald die van invloed zijn op evaluatiegebruik, verdeeld over de categorieën: evaluatie implementatie, beslissings- of beleidsomgeving en stakeholderbetrokkenheid. Alle factoren zijn onderverdeeld in deze categorieën en vervolgens gerangschikt op basis van hoe vaak ze zijn voorgekomen in de bestudeerde artikelen. Daarbij wordt de frequentie, een beschrijving van de factoren en de relatie met evaluatiegebruik gegeven. De bijbehorende tabel is opgenomen in bijlage 1. Ten tweede richt Bekkers et al. (2004) zich in het onderzoek op het verwerven van inzicht in factoren die van invloed zijn op de doorwerking van de beleidsadviesering door Nederlandse adviesraden (p. 5). Deze factoren (bijlage 2) zijn dus specifiek gericht op de Nederlandse context en daarmee ook zeer bruikbaar voor het opstellen van randvoorwaarden voor het borgen van de benutting van de invoeringstoets.

#### *Factoren samenvoegen en vertalen naar randvoorwaarden*

Voor beide artikelen zijn de factoren onderverdeeld in verschillende categorieën. Deze sluiten echter niet goed bij elkaar aan en zijn daarmee lastig samen te voegen. Ter illustratie zijn de factoren van Johnson et al. (2009) ingedeeld in de categorieën: evaluatie implementatie, beslissings- of beleidsomgeving en stakeholderbetrokkenheid; terwijl de factoren van Bekkers et al. (2004) zijn ingedeeld in de categorieën: kenmerken adviesproduct, behoeften ontvangers, organisatie adviesraad, interactie met doelgroep, presentatie en nazorg en institutionele kenmerken. De focus is dus zeer verschillend.

Om de factoren echter toch bruikbaar te maken en te kunnen vertalen naar randvoorwaarden voor het borgen van benutting van de invoeringstoets, is gekozen om de factoren samen te voegen door ze in te delen in nieuwe categorieën die aansluiten bij de invoeringstoets zoals die tot nu toe in dit onderzoek is omschreven. Dit zijn factoren met betrekking tot het evaluatieproces; het evaluatierapport en de presentatie en nazorg ervan; de evaluator en de ontvanger van de evaluatie; andere stakeholders en de grotere politiek-bestuurlijke context.

Vervolgens is gekeken naar welke factoren relevant lijken te zijn voor dit onderzoek en in welke categorie deze thuishoren. Voor Johnson et al. (2009) is de relevantie bepaald op basis van de frequentie waarin deze voorkwamen in de bestudeerde studies en het verband dat ze hadden met evaluatiegebruik (in sommige gevallen was de factor wel genoemd maar had het geen invloed op evaluatiegebruik). Voor Bekkers et al. (2004) was dit ingewikkelder omdat hierbij geen prioritering of andere waarde-indicatie is gegeven. Daardoor is enerzijds gekeken naar welke factoren ook in Johnson et al. (2009) worden genoemd en anderzijds naar bij welke factoren het aannemelijk is dat zij ook in de context van de invoeringstoets een substantiële invloed hebben op de benutting. Na het volgen van deze stappen blijft onderstaande tabel over.

Perspectieven	Categorieën	Belangrijkste factoren:
Rationeel	Het evaluatieproces	- Kwaliteit van het uitgevoerde proces - Eigenschappen van het beleidsprobleem/besluit - (Vroege) afstemming van het proces tussen kerndepartement en uitvoeringsorganisatie
	Het evaluatierapport	- Communicatiekwaliteit

		- Timing van afronding en verspreiding
Organisch	De relatie tussen de evaluator en ontvanger van de evaluatie	- Persoonlijke eigenschappen van de ontvanger - Betrokkenheid en/of ontvankelijkheid naar evaluatie
Politiek	Overige stakeholders	- Betrokkenheid van stakeholders
	De overkoepelende politiek-bestuurlijke context	- Politieke klimaat - Mate van politieke belangstelling

### 3.4.2 Randvoorwaarden voor het borgen van benutting

De factoren uit bovenstaande tabel zullen nu per categorie worden besproken en vertaald naar randvoorwaarden voor het borgen van benutting van de invoeringstoets. Ten slotte zal aan de hand van bovenstaande factoren antwoord worden gegeven op deze laatste theoretische deelvraag waarbij ook een verwachting wordt geformuleerd voor het empirische deel van het onderzoek.

#### *Het evaluatieproces*

Rondom het evaluatieproces is een drietal factoren van invloed die Johnson et al. (2009) en Bekkers et al. (2004) noemen en die ook van invloed zouden kunnen zijn op de benutting van de invoeringstoets. Allereerst is dit de kwaliteit van het uitgevoerde proces. Johnson et al. (2009) benoemt hierbij de eigenschappen van het evaluatieproces, de verfijning van de methoden en het type evaluatiemodel.

Ook de eigenschappen van het beleidsprobleem/besluit zijn van invloed. Deze factor is een combinatie van de factor 'besluit eigenschappen' van Johnson et al. (2009) en 'type beleidsprobleem' van Bekkers et al. (2004, p. 38). Die eerste staat omschreven als: het belang van het besluit of het geëvalueerde probleem, het type te maken besluit, de mate waarin het programmagebied nog nieuw is. Bekkers et al. (2004) gaan hier verder op in en stellen dat de 'kenmerken van het advies' van grote invloed zijn op de benutting van adviezen. Het advies zou vooral invloed kunnen hebben bij beleidsproblemen waarvan de ontvangers van de evaluatie nog geen mening hebben gevormd of waar juist de meningsverschillen onoverbrugbaar zijn en de evaluatie hierin een uitweg kan bieden.

Tot slot kan ook de afstemming van het proces tussen kerndepartement en uitvoeringsorganisatie van belang zijn. Deze factor is vooral gebaseerd op Bekkers et al. (2009) die de 'afstemming over het werkprogramma' noemt. Hiermee wordt bedoeld dat er een goede afstemming over het werkprogramma moet zijn met zowel het departement als de uitvoeringsorganisatie. Hierbij wordt benoemd dat van belang is dat goed inzicht wordt gekregen in de behoeften van het departement, de verwachtingen omtrent het advies en de stand van zaken in het beleidsproces en de datum van oplevering van het advies.

Op basis van deze omschrijvingen kunnen de factoren als volgt worden vertaald naar randvoorwaarden:

1. *Het evaluatieproces zit kwalitatief goed in elkaar, volgens verfijnde methoden en relevant evaluatiemodel.*
2. *Het type beleidsprobleem/besluit is relatief nieuw en er zijn nog weinig meningen over gevormd of de meningsverschillen over het onderwerp zijn onoverbrugbaar waardoor een invoeringstoets juist een oplossing biedt.*



3. *Het evaluatieproces wordt goed en duidelijk afgestemd tussen het departement en de uitvoeringsorganisatie.*

*Het evaluatierapport*

Bij het resultaat van de evaluatie: het evaluatierapport, zijn twee factoren van invloed. Te beginnen met communicatiekwaliteit. In het artikel van Johnson et al. (2009) komt deze factor met 11 keer naar voren als de meest voorkomende uit alle onderzochte studies. Het wordt omschreven als de duidelijkheid en frequentie van de rapportage van de resultaten, de belangenbehartiging van de resultaten door de evaluator en het verspreidingsgebied. Het gaat er dus vooral om dat over de rapportage ook voldoende wordt gecommuniceerd en het wordt verspreid. Ook Bekkers et al. (2004) benoemt deze factor en stelt dat de aanwezigheid van een communicatiestrategie van invloed is op de benutting. Hierbij wordt de stelregel '*naarmate het advies onder de aandacht van meer mensen komt, neemt de kans op doorwerking toe*' genoemd (p. 47). Daar wordt wel nog aan toegevoegd dat in de strategie moet worden afgewogen waar doorwerking kan optreden en waar en met welk doel op gericht moet worden voor een goede presentatie van het rapport.

Daarnaast wordt ook de timing van de afronding en verspreiding van de rapportage van invloed op de benutting. Johnson et al. (2009) omschrijven dit als de timing van de evaluatie in grotere context, de tijdigheid van de rapportage en de timing van de verspreiding. Bekkers et al. (2004) gaat hier verder op in en stelt dat het gaat om het gebruik maken van de eerder genoemde 'beleidsramen' waar politiek ondernemerschap en een fijn gevoel voor timing voor nodig is (p. 49).

De factoren kunnen op basis van bovenstaande als volgt worden vertaald naar randvoorwaarden:

4. *De resultaten worden duidelijk en wijd verspreid onder relevante actoren.*
5. *De evaluatie wordt afgerond en verspreid op een tactisch moment.*

*De relatie tussen de evaluator en ontvanger van de evaluatie*

De volgende categorie gaat over de evaluator, de ontvanger en de relatie tussen die twee, waaronder de factoren 'persoonlijke eigenschappen van de ontvanger' en 'betrokkenheid en/of ontvankelijkheid naar de evaluatie' vallen. Die eerste is geheel afgeleid uit Johnson et al. (2009) aangezien Bekkers et al. (2004) het heeft over adviesraden en hier persoonlijke eigenschappen niet in meegenomen zijn. In 9 verschillende onderzochte studies is een positief verband gevonden met benutting. De factor wordt omschreven als de eigenschappen van de evaluatiegebruiker zoals de organisatorische rol van de besluitvormer, de informatieverwerkingsstijl en sociale eigenschappen.

Ook de betrokkenheid en/of ontvankelijkheid naar de evaluatie komt van het artikel van Johnson et al. (2009) en is in 8 studies genoemd waarvan 7 als van invloed op benutting. In de beschrijving staat dat het gaat om de houding ten opzichte van de evaluatie en de toewijding voor het uitvoeren van de evaluatie door gebruikers. Ook de mate waarin de organisatie weerstand biedt tegen de evaluatie en de mate waarin stakeholders open staan voor de evaluatie worden hieronder geschaard.

De factoren kunnen als volgt worden vertaald naar randvoorwaarden:

6. *De ontvanger heeft persoonlijke eigenschappen die gunstig zijn voor benutting (organisatorische rol, informatieverwerkingsstijl, sociale eigenschappen).*

### *7. De ontvanger is betrokken en ontvankelijk naar de evaluatie*

#### *Overige stakeholders*

In de categorie overige stakeholders is slechts één factor opgenomen, namelijk: stakeholderbetrokkenheid. Deze factor is opgesteld uit een volledige categorie van Johnson et al. (2009) en twee factoren uit Bekkers et al. (2004). Johnson et al. (2009) hebben één van de drie categorieën volledig gewijd aan factoren over stakeholders. Hieronder vallen factoren als: betrokkenheid bij commitment of ontvankelijkheid voor de evaluatie; betrokkenheid bij communicatiekwaliteit; directe betrokkenheid stakeholders; betrokkenheid bij geloofwaardigheid; betrokkenheid bij bevindingen etc. Aangezien het lastig is te schatten welke stakeholders en welke aspecten van betrokkenheid bij de invoeringstoets van invloed gaan zijn is gekozen om deze factoren als één te beschouwen. Daarnaast is de factor afstemming over werkprogramma van Bekkers et al (2004) hierin meegenomen. Dit gaat over het betrekken en laten meedenken van relevante actoren als departementale vertegenwoordigers, representanten uit het beleidsveld of andere betrokkenen. Dit heeft twee voordelen: in de eerste plaats wordt de betrokkenheid van de doelgroep bij het advies vergroot, in de tweede plaats wordt ervoor gezorgd dat de Adviesraad een uitwerking kiest die aansluit bij de behoeften van de afnemers (p. 45).

De factor betrokkenheid van stakeholders wordt als volgt vertaald naar een randvoorwaarde:

#### *8. Overige stakeholders worden betrokken in het (opzetten van het) proces van de evaluatie.*

#### *Overkoepelende politiek-bestuurlijke context*

De laatste en minst beïnvloedbare categorie is de overkoepelende politiek-bestuurlijke context waaronder de factoren politieke klimaat en de mate van politieke belangstelling gelden. De factor politieke klimaat is opgesteld vanuit de gelijknamige factor van Johnson et al. (2009) die ook in 6 van de onderzochte studie voorkwam. Hieronder wordt verstaan 'de politieke oriëntatie van de mensen die opdracht gaven tot de evaluatie, de mate waarin de besluitvormer afhankelijk is van externe sponsors, interne rivaliteit, strijd om budgetten en machtsstrijd'. De factor politieke belangstelling komt tot slot juist van de gelijknamige factor van Bekkers et al. (2004). Zij stellen dat zodra een thema in het hart staat van de politieke belangstelling het doorgaans ook relatief veel aandacht van verschillende actoren krijgt, zoals belangengroepen en de media. Dit zou de instrumentele en conceptuele doorwerking positief kunnen beïnvloeden. Ook kan het juist tot politiek-strategische doorwerking leiden zonder dat dit uiteindelijk leidt tot instrumentele of conceptuele doorwerking. Bij onderwerpen waar weinig aandacht naar uit gaat kan de instrumentele of conceptuele doorwerking juist groot zijn (p. 50).

De factoren kunnen tot slot als volgt worden vertaald naar randvoorwaarden voor benutting:

#### *9. De evaluatie vindt plaats en wordt afgerond en verspreid in een gunstig politiek klimaat.*

#### *10. De evaluatie krijgt voldoende (maar niet te veel) politieke belangstelling.*

### 3.4.3 Conclusie en verwachtingen

Uit het artikel van Bekkers et al (2004) en de literatuurstudie van Johnson et al. (2009) is een aantal factoren afgeleid waarbij het aannemelijk is dat ze ook van invloed zijn op de benutting van de invoeringstoets. Deze zijn vertaald naar randvoorwaarden en staan weergegeven in

onderstaande tabel. De verwachting voor het empirische deel van het onderzoek is dus dat deze randvoorwaarden gelden voor het borgen van de benutting van de invoeringstoets, maar zullen nog moeten worden getoetst aan de empirie.

<b>Perspectieven</b>	<b>Categorieën</b>	<b>Belangrijkste randvoorwaarden</b>
<b>Rationeel</b>	Het evaluatieproces	<ol style="list-style-type: none"> <li>1. Het evaluatieproces zit kwalitatief goed in elkaar, volgens verfijnde methoden en een relevant evaluatiemodel.</li> <li>2. Het type beleidsprobleem/besluit is relatief nieuw en er zijn nog weinig meningen over gevormd of de meningsverschillen over het onderwerp zijn onoverbrugbaar waardoor een invoeringstoets juist een oplossing biedt.</li> <li>3. Het evaluatieproces wordt goed en duidelijk afgestemd tussen het departement en de uitvoeringsorganisatie.</li> </ol>
	Het evaluatierapport	<ol style="list-style-type: none"> <li>4. De resultaten worden duidelijk en wijd verspreid onder relevante actoren.</li> <li>5. De evaluatie wordt afgerond en verspreid op een tactisch moment.</li> </ol>
<b>Organisch</b>	De relatie tussen de evaluator en ontvanger van de evaluatie	<ol style="list-style-type: none"> <li>6. De ontvanger heeft persoonlijke eigenschappen die gunstig zijn voor benutting (organisatorische rol, informatieverwerkingsstijl, sociale eigenschappen).</li> <li>7. De ontvanger is betrokken en ontvankelijk naar de evaluatie</li> </ol>
<b>Politiek</b>	Overige stakeholders	<ol style="list-style-type: none"> <li>8. Overige stakeholders worden betrokken in het (opzetten van het) proces van de evaluatie.</li> </ol>
	De overkoepelende politiek-bestuurlijke context	<ol style="list-style-type: none"> <li>9. De evaluatie vindt plaats en wordt afgerond en verspreid in een gunstig politiek klimaat.</li> <li>10. De evaluatie krijgt voldoende (maar niet te veel) politieke belangstelling.</li> </ol>

## 4. Het empirisch onderzoek

### 4.1 Inleiding

In dit hoofdstuk wordt verder ingegaan op de aanpak en de kwaliteit van het onderzoek. Hierbij wordt allereerst de keuze voor kwalitatief onderzoek met een inductieve benadering gemotiveerd. Vervolgens wordt in paragraaf 4.3 de onderzoeksopzet uiteengezet en wordt verder toegelicht hoe de verschillende interviews vorm krijgen. In paragraaf 4.4 wordt besproken hoe de verschillende casussen en respondenten geselecteerd zijn, waarna in 4.5 tot een operationalisatie van het concept benutting wordt gekomen. Paragraaf 4.6 gaat daarna in op hoe de data verzameld en geanalyseerd wordt, waarna tot slot in 4.7 de kwaliteit van het onderzoek wordt besproken.

### 4.2 Kwalitatief onderzoek

Dit onderzoek gebruikt kwalitatieve onderzoeksmethoden. Het doel hiervan is namelijk om concepten te ontwikkelen die ons helpen om sociale verschijnselen in hun context te begrijpen met de nadruk op betekenissen, ervaringen en gezichtspunten van alle betrokkenen. De gedachte hierachter is dat de mens en zijn overwegingen, argumenten, ervaringen en motieven bestudeerd en begrepen kan worden (Boeije, 2019, p. 21). Hiermee sluit het kwalitatief onderzoek zeer goed aan bij dit onderzoek. Om tot de randvoorwaarden voor het borgen van benutting van de invoeringstoets te komen, is een diepgaand begrip nodig van de context waarin deze terecht komt. Beleidsprocessen bij de Rijksoverheid zijn ontzettend ingewikkeld omdat ze zich afspelen in een uitzonderlijk dynamische omgeving met een veelvoud aan betrokken actoren die elk weer eigen perspectieven en belangen hebben. Dit zijn allemaal factoren die van invloed zouden kunnen zijn op hoe de invoeringstoets (structureel) benut kan worden. De context is dus niet alleen te begrijpen vanuit tekstuele bronnen, maar moet ook blijken uit betekenisgeving, ervaringen en perspectieven van de mensen die al een diepgaand begrip van de context hebben en zelf ook werken met de invoeringstoets. Vooral ook omdat je pas met verschillende - misschien wel botsende - perspectieven en ervaringen écht een goed beeld kunt krijgen van de praktijk, en niet alleen door eenduidige tekstuele bronnen te bestuderen.

Een andere reden dat kwalitatief onderzoek zeer geschikt is voor dit onderzoek is dat de invloed van de onderzoeker bij zowel de dataverzameling als de data-analyse groot is (Boeije, 2019, p. 27). Zeker omdat de invoeringstoets een nieuw instrument is en het lastig is te voorspellen hoe deze in de praktijk werkt, is het vooraf lastig in te schatten wat er in de dataverzameling naar voren komt. Om tot meer diepgang te komen is het dus zeer nuttig om vervolgens in te kunnen spelen en door te kunnen vragen op relevante onderwerpen. Het is daarbij niet erg om ook minder relevante onderwerpen te bespreken omdat deze er in de data-analyse weer uitgefilterd kunnen worden. Tot slot is de kwalitatieve methode nuttig omdat concepten kunnen worden ontwikkeld voor het begrijpen van sociale verschijnselen (Boeije, 2019, p. 28), wat goed aansluit bij het doel van dit onderzoek: het vinden van randvoorwaarden voor het borgen van benutting van de invoeringstoets.

Binnen deze kwalitatieve methoden is een deductieve benadering gebruikt. In hoofdstuk 3 kwam naar voren dat vanuit verschillende perspectieven gekeken kan worden naar beleidsprocessen en dus ook de wijze waarop de invoeringstoets hierbinnen past. Daarbij is vanuit een rationele, organische en politieke bril gekeken naar hoe de invoeringstoets kan gaan werken. Dezelfde perspectieven zijn ook van toepassing in het onderscheid tussen verschillende vormen van benutting in 3.2 en in de voorwaarden voor benutting uit paragraaf 3.3.3. De drie perspectieven fungeren daarom ook als *sensitizing concepts* en zijn als dusdanig meegenomen in de dataverzamelingsfase. Ook de

opgestelde randvoorwaarden voor benutting zullen worden gebruikt als basis voor de dataverzameling. Er zal niet direct worden gevraagd of de randvoorwaarden inderdaad in de praktijk gelden, maar er kan in de gesprekken wel op gestuurd of op teruggevallen worden. Zodra echter blijkt dat het andere (rand)voorwaarden zijn dan kan daar juist verder op in worden gegaan in plaats van te fixeren op de randvoorwaarden uit hoofdstuk 3.

### 4.3 Onderzoeksdesign

Voor het onderzoeksdesign zal gebruik worden gemaakt van een combinatie van een vergelijkende en een meervoudige casestudie. Er zijn namelijk verschillende casussen onderzocht met zowel invoeringstoetsen als andere evaluaties met vergelijkbare kenmerken.

Meervoudige casestudies zijn om twee redenen nuttig voor theorievorming. Ten eerste kan door twee of mee cases te vergelijken de onderzoeker beter de omstandigheden bloot leggen waarin een theorie wel of niet standhoudt. Daarnaast kan de vergelijking zelf concepten suggereren die relevant zijn voor een opkomende theorie (Bryman, 2012, p. 67). In dit onderzoek wordt gezocht naar de randvoorwaarden voor het borgen van de benutting van de invoeringstoets, waarvoor in 3.3 al een aantal randvoorwaarden vanuit de literatuur zijn voorgelegd. Door cases van uitgevoerde invoeringstoetsen te vergelijken met cases van andere evaluaties kan beter worden getoetst onder welke omstandigheden welke voorwaarden gelden. Daarnaast kunnen door te kijken naar andere evaluaties ook nieuwe randvoorwaarden gevonden worden die (wellicht) ook voor de invoeringstoets gelden

Een andere reden voor de keuze voor een vergelijkende meervoudige casestudie is van een meer praktische aard. Aangezien de invoeringstoets nog nauwelijks bestaat en op het moment van uitvoeren slechts op enkele plekken binnen Rijksoverheid wordt geëxperimenteerd met de invoeringstoets is het onrealistisch om tot verregaande verdieping te komen van één casus van de invoeringstoets. In de meeste gevallen is namelijk maar één persoon voldoende inhoudelijk betrokken geweest om echt een interview te geven over die invoeringstoets. Sterker nog: er is überhaupt maar een enkele invoeringstoets gedaan waarbij men bereid was hier een interview over te geven. Om die reden was het niet mogelijk om puur data te verzamelen over de invoeringstoets en moest gezocht worden naar een andere benadering. Hierdoor zijn ook casussen gebruikt van evaluaties met een soortgelijk karakter als de invoeringstoets om lessen te kunnen trekken uit de ervaringen hiermee.

#### 4.3.1. Semigestructureerde interviews

Er zijn in totaal 12 interviews afgenomen met drie verschillende opzetten:

- Interviews over uitgevoerde invoeringstoetsen
- Interviews over evaluatiecasussen met een soortgelijk karakter als de invoeringstoets
- Interviews over de (benutting van de) invoeringstoets in het algemeen

##### *Interviews over uitgevoerde invoeringstoetsen*

5 interviews zijn afgenomen met verantwoordelijken voor een uitgevoerde invoeringstoets. Deze interviews hadden als doel om zo goed mogelijk in kaart te brengen hoe het proces van de invoeringstoets in de praktijk in zijn werk gaat. Hierbij werd onder andere gekeken naar het handelen van betrokkenen en de ervaringen, opvattingen en houdingen ten aanzien van de invoeringstoets. De ondervraagde respondenten

waren allemaal beleidsmedewerkers die de getoetste wet of regeling in hun portefeuille hadden en die dus ook genoeg konden vertellen over enerzijds de betreffende wet/regeling en anderzijds de uitgevoerde invoeringstoets.

Deze interviews hadden een semigestructureerd karakter waarbij van tevoren een topiclijst is opgesteld (zie bijlage 3). Alleen voor de derde empirische deelvraag *'in hoeverre en hoe wordt in de praktijk voldaan aan de randvoorwaarden voor het borgen van de benutting van de invoeringstoets?'* wordt deels deductief beredeneerd en wordt er dus theorie meegenomen in de interviews. Alleen voor de randvoorwaarden voor het borgen van benutting zijn de theoretische randvoorwaarden uit 3.3.2 meegenomen in de interviews. Hier kon extra naar worden gevraagd in het geval dat hier gedurende de gesprekken aanleiding voor was. Bijvoorbeeld als de respondent een soortgelijk punt maakte of juist iets tegenstrijdigs zei. Verder wordt vooral inductief beredeneerd waardoor relatief weinig structuur gewenst is. De respondenten weten immers wat belangrijk en relevant is en zij hoeven alleen enigszins gestuurd te worden naar het juiste onderwerp.

#### *Interviews over evaluatiecasussen met een soortgelijk karakter als de invoeringstoets*

Met 6 personen zijn interviews afgenomen over een evaluatiecasus met een soortgelijk karakter als de invoeringstoets. Dit zijn voornamelijk casussen van wetten en/of regelingen waarover ook een invoeringstoets gedaan had kunnen worden omdat het onderwerp zich hiervoor leent. Dit zouden onderwerpen kunnen zijn waarbij bijvoorbeeld de doelgroep kwetsbaar is of de uitvoering lastig zou kunnen zijn. Daarnaast zijn het vooral casussen waarbij het evaluatie-instrument zelf ook gelijkenissen heeft met de invoeringstoets, bijvoorbeeld omdat het evaluatiemoment hetzelfde is of het ook gaat om een relatief lichtvoetige evaluatie. Ook hier waren de geïnterviewde respondenten portefeuillehouder van de betreffende wet/regeling, waarbij voor één casus zowel iemand van de uitvoeringsorganisatie als het kerndepartement is gesproken.

Deze interviews hadden als doel om lessen te trekken voor de invoeringstoets uit soortgelijke, reeds bestaande evaluaties. Hierbij lag de focus voornamelijk op de randvoorwaarden voor de benutting van evaluaties en dus de derde empirische deelvraag. Deelvraag 1 en 2 gaan namelijk specifiek over de invoeringstoets waarbij niet per se lessen getrokken kunnen worden uit andere evaluaties. Ook bij deze interviews is gekozen voor een semigestructureerde opzet met een topiclijst (zie bijlage 4). In deze gevallen was toch enige mate van structuur nodig omdat een vergelijking getrokken moest worden met de invoeringstoets. Hiervoor was het wel van belang dat eerst de benodigde informatie over de betreffende evaluatie werd vergaard voordat over kon worden gegaan op de invoeringstoets en de vergelijking tussen de twee. Tegelijkertijd was het ook hier van belang dat niet te veel sturing werd gegeven aan het interview zodat ingesprongen kon worden op relevante zaken, bijvoorbeeld zodra de respondent een complicatie bij de evaluatie noemt. Hier kan dan verder op worden ingegaan zodat goed in kaart kan worden gebracht of dit ook voor de invoeringstoets een relevante randvoorwaarde zou kunnen zijn.

#### *Interviews over de (benutting van de) invoeringstoets in het algemeen*

Tot slot zijn nog 3 interviews gehouden met 'experts' van de invoeringstoets. Dit waren interviews met de huidige projectmanager van de rijksbrede implementatie van de invoeringstoets, diens voorganger en iemand uit de kopgroep voor de implementatie van de invoeringstoets. Ook deze interviews waren semigestructureerd en hadden een tweeledig doel. Enerzijds moesten de interviews antwoord geven op de eerste twee empirische deelvragen. Als (oud-)projectmanager hebben zij het beste een beeld van hoe de invoeringstoets in theorie en praktijk werkt en hoe en waar deze tot nu toe benut worden. Daarnaast zijn de interviews gebruikt om de randvoorwaarden voor te leggen die uit de rest van de interviews zijn gekomen en te valideren of dit inderdaad randvoorwaarden zijn en in hoeverre en hoe hieraan wordt voldaan binnen de overheid. Met name voor het eerste doel was het handig om niet te veel structuur in het gesprek te hebben en vooral de respondenten te laten

vertellen, terwijl het voor de randvoorwaarden juist fijn was om dit meer te structureren. Daarbij is eerst gevraagd welke randvoorwaarden zij zelf dachten dat gelden om vervolgens de randvoorwaarden uit de andere interviews voor te leggen.

#### 4.4 Casus- en respondentselectie

Zoals bij 3.2.2 al is aangegeven zijn in totaal 12 interviews afgenomen met 14 respondenten over in totaal 9 cases (3 interviews gingen niet specifiek over een case maar over de invoeringstoets in het algemeen). Hierbij is gebruik gemaakt van een combinatie van *purposive sampling* – een steekproef op basis van relevantie voor de onderzoeksvraag - en *convenience sampling* – een steekproef op basis van gelegenheid (Bryman, 2016, p. 408). De cases over de invoeringstoets zijn puur en alleen geselecteerd om de reden dat ze over een invoeringstoets gaan (*convenience*). Zoals al vaker is aangegeven is op het moment van uitvoeren slechts een enkele invoeringstoets gedaan waarvan het ook nog eens niet altijd mogelijk was om een interview te organiseren. Voor deze cases zijn in totaal 12 interviews gehouden waarvan twee met twee respondenten tegelijk (omdat zij geen toegevoegde waarde zagen in twee losse interviews).

Voor de selectie van cases met soortgelijke evaluaties is gepoogd om zoveel mogelijk een verscheidenheid aan betrokken instanties te krijgen (*purposive*). Dat wil zeggen dat de evaluaties zijn uitgevoerd door verschillende uitvoeringsorganisaties met verschillende kerndepartementen. Hiervoor is gekozen omdat verschillende ministeries en uitvoeringsorganisatie momenteel los bezig zijn met het ontwikkelen van een methode voor de invoeringstoets en deze dus ook anders benut kunnen worden. Uiteindelijk zijn 8 cases onderzocht waarvoor 9 interviews zijn gehouden. Bij één case is zowel iemand van de uitvoeringsorganisatie als het kerndepartement bevroegd. Tot slot zijn twee interviews gehouden met 'experts'. Zij zijn geselecteerd op basis van hun expertise als (oud-)projectmanager. De respondenten werkten bij onderstaande organisaties:

- IND
- Rijksdienst voor Ondernemend Nederland
- Nederlandse Voedsel- en Warenautoriteit
- VNG
- Toeslagen
- Ministerie van Economische Zaken en Werkgelegenheid
- Ministerie van Justitie en Veiligheid
- Ministerie van Sociale Zaken en Werkgelegenheid

#### 4.5 Operationalisatie

Bij een inductieve benadering is het in de dataverzameling en -analyse noodzakelijk dat de belangrijkste concepten uit het onderzoek ook meetbaar worden gemaakt (Boeije, 2019, p. 101). Aangezien dit onderzoek poogt te komen tot de randvoorwaarden voor het borgen van de benutting van de invoeringstoets is het enige centrale concept benutting. Daarbij is de operationalisatie van het concept alleen voor empirische deelvraag 2 van belang, omdat alleen hiervoor de benutting gemeten hoeft te worden. Immers gaat het bij deelvraag 1 om de werking van de invoeringstoets en bij deelvraag 3 om hoe de benutting van de invoeringstoets geborgd kan worden en hoeft de benutting zelf dus niet gemeten

te worden. De basis voor de operationalisatie van het concept benutting is in hoofdstuk 3 gelegd, waarbij het onderscheid in verschillende vormen van doorwerking centraal staat.

Om het concept benutting meetbaar te maken zal het uiteengegafd moeten worden in verschillende dimensies waarin het concept voor kan komen. Hiervoor zullen de verschillende perspectieven en de verschillende vormen van doorwerking van Bekkers et al. (2004, pp. 24-26) worden gebruikt. Aangezien deze dimensies nog niet concreet waargenomen kunnen worden moet nog een extra stap worden gezet. Om tot dergelijke indicatoren te komen zal wederom gebruik worden gemaakt van het artikel van Bekkers et al. (2004). Zij hebben namelijk meetpunten geïdentificeerd voor de verschillende vormen van doorwerking in 6 verschillende domeinen (p. 33), waarvan in dit onderzoek alleen departement, politiek en beleidsveld worden meegenomen. Aangezien de focus van het onderzoek van Bekkers et al. (2004) ligt op de benutting van adviezen van adviesraden is een kleine vertaalslag gemaakt zodat de meetpunten gebruikt kunnen worden als indicatoren voor benutting van de invoeringstoets.

Benutting	Dimensies	Meetbare indicatoren	
1. Rationeel	Instrumentele doorwerking: de resultaten van de evaluatie leiden tot een directe verandering van gedrag van individuen of organisaties in lijn met de aanbevelingen van het advies.	Departement	- Aanwezigheid Kabinetsstandpunt - Aantal vervolgacties
		Tweede Kamer	- Aantal vragen of moties van de Tweede Kamer in vragenuurtje, commissie of plenaire zitting
		Uitvoering	- Aantal vragen en verzoeken om nadere informatie - Aantal officiële reacties van koepelorganisaties, uitvoeringsorganisaties of professionals
2. Organisch	Conceptuele doorwerking: de resultaten van de evaluatie hebben een verandering in de kennis, opvattingen of causale redeneringen van individuen of organisaties tot gevolg.	Kerndepartement	- Aantal verwijzingen in strategische beleidsdocumenten
		Tweede Kamer	- Aantal vervolgacties van Tweede Kamer
		Uitvoering	- Aantal verwijzingen in beleidsplannen en jaarverslagen van koepelorganisaties, uitvoeringsorganisaties of professionals
3. Politiek	Agenderende doorwerking: de resultaten van de evaluatie worden gebruikt om een nieuw onderwerp of beleidsthema onderwerp van het	Kerndepartement	- Aantal opgestarte, nieuwe beleidsinitiatieven op aangedragen nieuw thema
		Tweede Kamer	- Aantal toelichtingen in commissievergaderingen of hoorzittingen



	maatschappelijke of politieke debat te maken.		- Aantal geagendeerde issues
		Uitvoering	- Aantal keren thema onderwerp van studiedagen, symposia en congressen
	Politiek-strategische doorwerking: de resultaten van de evaluatie worden gebruikt om de machtspositie van één of enkele spelers te versterken om op deze wijze de politieke doelen beter te kunnen bereiken.	Kerndepartement	- Legitimering van bestaand beleid
		Tweede Kamer	- Aantal keer genoemd als gelegenheidsargument ter versterking van eigen positie
		Uitvoering	- Aantal keer genoemd als gelegenheidsargument ter versterking van eigen positie

#### 4.6 Dataverzameling en -analyse

Om de verschillende empirische deelvragen te beantwoorden zijn verschillende methoden gebruikt bij het verzamelen en analyseren van data. Zoals in 4.3 al is beschreven zijn de 12 interviews uitgevoerd met 3 verschillende opzetten, omdat nog niet voldoende invoeringstoetsen zijn uitgevoerd en tot benutting hebben kunnen leiden. Hiervan hebben de interviews over uitgevoerde invoeringstoetsen en evaluatiecasussen met een soortgelijk karakter dezelfde functie ten behoeve van de beantwoording van de theoretische deelvragen. Deze zijn namelijk gebruikt om te achterhalen hoe de invoeringstoets binnen de beleidscyclus van de Rijksoverheid past, hoe en waar de invoeringstoets wordt benut binnen de Rijksoverheid en wat de randvoorwaarden voor het borgen van de benutting van de invoeringstoets zijn. De drie expertinterviews zijn vervolgens gebruikt om de bevindingen uit de voorgaande interviews voor te leggen en te achterhalen hoe en in hoeverre in de praktijk wordt voldaan aan de randvoorwaarden voor het borgen van de invoeringstoets.

Na de dataverzamelingsfase zijn de opnames van de interviews getranscribeerd, geanonimiseerd en met behulp van het programma Nvivo 12 pro gecodeerd. Wederom is hiervoor een deductieve benadering toegepast. Dat wil zeggen dat met behulp van de literatuur en de operationalisatie uit de vorige paragraaf een codeboom is opgesteld (bijlage 5). Deze bestaat in de basis uit drie onderdelen die gelinkt zijn aan de drie empirische deelvragen. De data is vervolgens eerst gecodeerd aan de hand van indicatoren. Voor de eerste deelvraag zijn hierbij de *sensitizing concepts* als indicatoren gebruikt. Voor de tweede deelvraag zijn de indicatoren uit paragraaf 4.4 gebruikt en voor de laatste deelvraag zijn de theoretische randvoorwaarden uit 3.3 als indicatoren gebruikt. Na het coderen in indicatoren is de operationalisering bijgesteld waarbij vooral de randvoorwaarden zijn herzien. Tot slot is nogmaals gecodeerd met dimensies en is ook de categorisering aangepast om zo tot een onderbouwde analyse van de data te komen.

## 4.7 Kwaliteit van het onderzoek

Tot slot zal dit hoofdstuk stilstaan bij de kwaliteit van het onderzoek. Hierbij wordt ingegaan op de externe validiteit, de interne validiteit en de betrouwbaarheid, waarna tot slot wordt besproken hoe gedurende het onderzoek wordt omgegaan met ethiek.

### *Externe validiteit*

Allereerst zal worden ingegaan op de vraag of de resultaten generaliseerbaar zijn buiten de context van het onderzoek. Aan dit onderzoek hebben in totaal 14 respondenten meegewerkt van in totaal 9 verschillende organisaties die met de invoeringstoets moeten gaan werken. Deze organisaties vormen samen een mooie mix van kerndepartementen en uitvoeringsorganisaties uit verschillende hoeken van Rijksoverheid. Gezien de grootte van het onderzoek was het niet mogelijk om alle organisaties die met de toets gaan werken in het onderzoek te betrekken. De huidige selectie is daarmee een mooie steekproef waardoor de resultaten voldoende generaliseerbaar zijn voor de overige organisaties.

Helaas heeft zich wel een andere beperking voorgedaan voor de externe validiteit. Het doel van dit onderzoek is om te komen tot de randvoorwaarden voor het borgen van de benutting van de invoeringstoets. Doordat nog heel weinig is gewerkt met de invoeringstoets en hierin veel van die gevallen nog niet over is gerapporteerd was het lastig om geschikte casussen te vinden. In het onderzoek zijn daarom slechts vier invoeringstoetsen onderzocht en is gekozen om ook andere evaluaties mee te nemen. Daarmee is in feiten dus niet alleen onderzoek gedaan naar de invoeringstoets, maar ook naar andere evaluaties. Hierbij is gepoogd om evaluaties te vinden met veel gelijkenissen met de invoeringstoets, maar het is lastig te zeggen welke kenmerken van die evaluaties voor het onderzoek van belang zijn. Dit doet af aan de generaliseerbaarheid van het onderzoek. De respondenten die betrokken waren bij soortgelijke evaluaties bleken uiteindelijk prima in staat om vanuit hun ervaringen met bestuur en politiek de vertaalslag te maken naar de invoeringstoets. De andere evaluaties zijn dus niet zozeer meegenomen in de data, maar vooral de ervaringen van de respondenten over evaluaties in het algemeen. Die ervaringen zijn weer toegespitst op de invoeringstoets. Hierdoor zijn de conclusies alsnog voldoende generaliseerbaar.

### *Interne validiteit*

Ten tweede komt de interne validiteit aan bod. Om dit te waarborgen zijn interviews gehouden met mensen die zich bezighouden met ofwel de invoeringstoets ofwel soortgelijke evaluaties. Vanuit hun ervaring kunnen zij dus goed inschatten wat wel en niet bijdraagt aan de benutting van evaluaties en specifiek de invoeringstoets. Daarnaast is de literatuur gebruikt om een onderbouwd uitgangspunt te hebben voor de interviews. Hierdoor konden de interviews waar nodig worden bijgestuurd zodat vooral de relevante onderwerpen besproken werden en het gesprek niet teveel afdwaalde.

Helaas heeft het onderzoek ook een belangrijke beperking op het gebied van interne validiteit. Om benutting te kunnen meten is het concept in paragraaf 4.4 geoperationaliseerd. Hiervoor zijn in de literatuur indicatoren gevonden om de verschillende categorieën waar het concept in ingedeeld kan worden te meten. Dergelijke indicatoren waren 'het aantal geagendeerde issues', 'het aantal keer genoemd als gelegenheidsargument ter versterking van eigen positie' en 'het aantal vragen en verzoeken om nadere informatie'. Doordat tot nu toe heel weinig resultaten van invoeringstoetsen in de praktijk hebben kunnen doorwerken was het heel moeilijk om het concept aan de hand van de indicatoren uit de literatuur te meten. Daarom is in de dataverwerking gekozen om gebruik te maken van uitspraken over hoe en waar de

resultaten volgens de verwachting van respondenten gaan doorwerken. Hiermee is geen concrete doorwerking gemeten maar wordt middels de ervaringen van respondenten uit bestuur en beleid wel een goed beeld geschetst van waar die doorwerking plaatsvindt.

#### *Betrouwbaarheid*

Verder staan we stil bij de betrouwbaarheid van het onderzoek. Het grootste probleem hierbij zit hem in het moment waarop het onderzoek wordt uitgevoerd. Gedurende het onderzoek heeft de invoeringstoets namelijk al een flinke ontwikkeling doorgemaakt en de komende jaren zullen nog vele ontwikkelingen volgen. Naar mate het instrument meer vorm krijgt zullen ook de conclusies veranderen. Daarnaast hadden de interviews een vrij open karakter waarbij de onderzoeker veel ruimte krijgt om het onderzoek te sturen. Het gevaar is daarbij dat tunnelvisie ontstaat over wat van belang is voor het onderzoek. Daarom is geprobeerd om zoveel mogelijk de respondenten te laten vertellen en alleen bij te sturen zodra het gesprek echt afdreef. Zodra dus nu het zelfde onderzoek uitgevoerd zou worden zouden ongeveer dezelfde resultaten naar voren moeten komen.

#### *Ethiek*

Tot slot staan we kort stil bij de ethiek van dit onderzoek. Alle respondenten hebben voorafgaand aan de interviews toestemming gegeven om het gesprek op te nemen en eventueel quotes op te nemen in het rapport. De opnames zijn vervolgens weer verwijderd en de opgestelde transcripten zijn samen met een concept van het resultatenhoofdstuk teruggestuurd naar de respondenten. Zij konden de stukken doorlezen om te controleren of wat in het rapport is opgenomen ook echt is wat zij hebben gezegd. Daarnaast konden zij controleren of geen van de uitspraken (ongewenst) herleidbaar waren en dat geen schadelijke informatie was opgenomen die niet openbaar mocht zijn. Op een kleine aanpassing na was dit niet het geval.

## 4.8 Conclusie

Dit onderzoek is volgens kwalitatieve methoden uitgevoerd en heeft daarbij een inductieve benadering gehanteerd. Voor het empirische deel zijn 12 semigestructureerde interviews afgenomen met 14 respondenten van 8 verschillende organisaties, waarbij 9 casussen zijn onderzocht. Hiervan gingen 4 casussen over een invoeringstoets en 4 casussen over een evaluatie met vergelijkbare kenmerken als de invoeringstoets. De interviews zijn vervolgens getranscribeerd en aan de hand van de codeboom uit bijlage 5 gecodeerd.

## 5. Resultaten

### 5.1 inleiding

Dit hoofdstuk presenteert de belangrijkste resultaten waarmee antwoord wordt gegeven op de verschillende empirische deelvragen. Allereerst wordt besproken hoe de invoeringstoets binnen de beleidscyclus van de Rijksoverheid past, waarbij net als in het theoretisch kader vanuit de drie perspectieven wordt gekeken. Paragraaf 5.3 bespreekt hoe en waar de invoeringstoets wordt benut. Hierbij zullen de vier vormen van doorwerking van Bekkers et al. (2004) centraal staan. In 5.4 wordt ingegaan op de laatste empirische deelvraag en worden de randvoorwaarden voor het borgen van de benutting van de invoeringstoets gepresenteerd. Tot slot geeft paragraaf 5.5 een conclusie van de bevindingen van het empirische deel van dit onderzoek. Deze wordt in hoofdstuk 6 verder besproken en geduid.

### 5.2 Hoe past de invoeringstoets binnen de beleidscyclus van de Rijksoverheid?

Tijdens de interviews is veelvuldig besproken wat de invoeringstoets precies is en hoe deze in de praktijk werkt. Dit onderwerp kwam niet alleen naar boven naar aanleiding van expliciete vragen ernaar, maar het onderwerp is ook regelmatig op natuurlijke wijze ter sprake gekomen of tussen de thema's door kort aangekaart. Hierbij viel op dat met name zodra het gesprek specifiek over de werking van het instrument ging op een voornamelijk rationele wijze werd gekeken, terwijl gedurende de rest van de gesprekken ook veel aspecten van de andere twee perspectieven terugkwamen. De werking van de toets volgens de respondenten zal nu vanuit elk van de perspectieven worden omschreven waarna een antwoord zal worden gegeven op de eerste empirische deelvraag.

#### 5.2.1 Rationeel

Zodra aan de respondenten werd gevraagd om het instrument invoeringstoets te omschrijven kwam er in de meeste gevallen een beschrijving uit die overeenkomt met hoe het geformuleerd is in de Kamerbrief van Weerwind (2022). Een aantal kenmerken komt hier veelvuldig in terug. Allereerst werd het door verschillende respondenten een *'lichtvoetige'* of een *'wat kleinere'* toets genoemd (R2, R3, R6, R11 en R12), waarbij het ook verschillende keren werd vergeleken met een tussentijdse thermometer. Daarnaast vindt de toets volgens de meeste respondenten relatief vroeg na inwerkingtreding van wet- en regelgeving plaats (R1a, R2, R3, R4, R6, R7, R11, R12). De projectmanager (R11) en oud-projectmanager (R12) van de rijksbrede implementatie van de invoeringstoets gebruikten hierbij ook echt de verwoording die in de Kamerbrief staat, namelijk: *'het vroegst mogelijke moment waarop iets gezegd kan worden over de werking van een nieuwe wet of regeling'*. Andere respondenten waren hier minder concreet over, maar hadden het wel telkens over 'een vroeg stadium' en 'niet pas na 4/5 jaar' of iets in die richting.

Respondenten beschreven ook vaak welke aspecten van wet- en regelgeving precies onderzocht of geëvalueerd (moeten) worden in een invoeringstoets. Over het algemeen focust de toets op vragen als: hoe werkt de wet of regeling in de praktijk, zijn er onvoorziene (neven)effecten, zijn er knelpunten en waar zitten die dan en zijn er dingen die je kunt verbeteren? Deze vragen zijn telkens gericht op enerzijds de uitvoering en anderzijds op burgers en bedrijven die samen de doelgroep vormen. Het uiteindelijke doel van het instrument is volgens de respondenten dat je het beleid waar nodig op tijd kunt bijsturen (R1a, R3, R4, R6, R7, R11, R12). Uit de interviews komt echter niet volledig consistent naar voren wat deze bijsturing precies inhoudt. De respondenten lijken het unaniem eens te zijn over dat wijzigingen kunnen worden aangebracht in de manier waarop invulling wordt gegeven aan beleid. Daarbij zouden regelingen in de meeste gevallen ook gewoon aangepast

kunnen worden. De vraag is alleen nog of ook het aanpassen van wetgeving een gevolg zou moeten kunnen zijn. Zo zeggen respondenten 2, 11 en 12 dat de invoeringstoets niet bedoeld is om gelijk wetswijzigingen door te gaan voeren, omdat beleid wel de kans moet krijgen om in te klinken. Respondenten 1a en 8 stellen juist dat het wél een manier is om (ook) wetgeving terug te kunnen leggen bij de politiek. Ook respondent 12 benoemt dit maar stelt wel dat de wet ook echt grote onbedoelde gevolgen moet hebben. Hij zegt hierover het volgende: *'Stel je, moet een vergunning weigeren omdat het nou eenmaal zo in de wet staat, maar je ziet eigenlijk na een jaar bij invoering dat je in die situatie toch ook wel graag een vergunning zou willen geven, omdat het eigenlijk wel past in de bedoeling en geest van de wet, maar naar de letter mag het niet. Ja, dan moet je toch ook wel echt terug naar de juridische afdeling van het ministerie, of misschien ook wel terug naar de Tweede Kamer? En gaan voorstellen om die regels dan ook aan te passen. Want zoiets kan dus gewoon onbedoeld zijn'* (respondent 12).

Als dus rationeel wordt gekeken naar de invoeringstoets dan is het een relatief lichte vorm van ex durante evaluatie met de focus op de gevolgen voor de uitvoering en doelgroep. Relatief vroeg na inwerkingtreding wordt gekeken naar de (onbedoelde) gevolgen van wet- en regelgeving voor de uitvoering en burger. Het doel hierbij is om op tijd aanpassingen te kunnen doen aan de invulling en in sommige gevallen ook de wet of regeling zelf.

### 5.2.2 Organisch

Hoewel de respondenten zodra gevraagd wordt naar de werking van de invoeringstoets vooral een rationele bril op lijken te zetten, blijkt deze in de belevingswereld van de respondenten ook afhankelijk te zijn van de omstandigheden. Gedurende de interviews is een veelvoud aan aspecten van de invoeringstoets genoemd die in hoge mate bepaald worden door de omstandigheden waarin de wet/regeling wordt ingevoerd en die dus van invloed is op de werking van het instrument. Momenteel zijn de ministeries en uitvoeringsorganisaties individueel bezig met uitzoeken hoe ze de invoeringstoets gaan aanpakken (R11). Hiermee wordt erkend dat de werking van de toets afhankelijk is van de context en wordt er ook op ingespeeld door decentraal tot een invulling te komen. Het idee erachter is dan ook dat de invoeringstoets in één standaard format voor veel organisaties niet bruikbaar is en dat de organisaties dus ook zelf moeten kijken welke vorm in de eigen context werkbaar is. Een aantal van de genoemde aspecten wordt besproken.

Allereerst geeft de formulering 'het vroegst mogelijke moment waarop iets kan worden gezegd over de werking' eigenlijk al aan dat het dus sterk afhankelijk is van de context en wet of regeling op welk moment die toets uitgevoerd wordt. Hierbij ligt de afweging vooral bij op welk moment je inderdaad voldoende kunt zeggen over de werking van een wet of regeling, maar je niet dusdanig lang wacht dat het ook daadwerkelijk tot (grote) problemen kan leiden. Zo geeft respondent 1a een voorbeeld van een regeling die in een hectische omgeving is ingevoerd, waardoor de invoeringstoets al na 3 maanden is uitgevoerd. Hier merkte hij ook dat verschillende betrokken partijen nog niet goed konden zeggen wat de gevolgen waren. Tegelijkertijd kon ook niet pas na 1 of 2 jaar een invoeringstoets worden gedaan omdat de risico's op knelpunten te groot waren en deze dan te lang zouden doormodderen. Ook respondent 7 benoemt dit punt. Hierop aansluitend is het nog maar de vraag in hoeverre die afweging door ministerie en uitvoeringsorganisatie op dezelfde manier gemaakt wordt. Respondent 5b vertelt bijvoorbeeld over een situatie waarin door de uitvoeringsorganisatie nog werd gewerkt aan een evaluatie toen het departement het beleid al had aangepast.

Een ander genoemd punt is dat niet iedereen ervoor open staat om tijd en energie in de evaluatie te steken (R7, R12). Respondent 12 omschreef dit als dat je lang en intensief gewerkt hebt om in een korte tijd nieuwe wetgeving te creëren en dat als je dan eindelijk klaar bent en

iets ander wil gaan doen, je nog lastiggevallen wordt met zo'n invoeringstoets. Zeker omdat men volgens de respondent binnen de overheid allergisch is voor 'bureaucratische hoepeltjes'. Daarbij werkt ook het gebrek aan capaciteit niet mee (respondent 1a, 7, 8, 10, 12), zowel bij de uitvoering als bij de opvolging van een evaluatie. Respondent 7 stelt om die reden dat aan de voorkant gekeken moet worden naar of een invoeringstoets nuttig en noodzakelijk is en of mensen daarop zitten te wachten. Tegelijkertijd blijkt uit de interviews met de projectmanager en oud projectmanager (R11 en R12) dat bij de implementatie wel rekening gehouden wordt met het feit dat de methoden van het instrument afhankelijk zijn van de omstandigheden waarin deze gebruikt wordt. Respondent 11 zegt bijvoorbeeld het volgende: *'ik kan me voorstellen, ik weet niet of dat zo is hoor, maar dat de manier waarop je de invoeringstoets inricht, dat het natuurlijk ook wel een beetje afhangt van het dossier. Wat is je doelgroep bijvoorbeeld of wat voor uitvoeringsorganisatie is het? Er is dus niet één juiste manier'*. Ook respondent 12 erkent veel van de omstandigheden die van invloed zijn. Hij legt daarom de bal ook nadrukkelijk bij de verschillende organisaties en probeert daarbij de nadruk te leggen op de houding ten opzichte van de toets. Men zou namelijk nieuwsgierig moeten worden naar het effect van hun werk in de praktijk. Het uiteindelijke doel van de invoeringstoets is volgens hem dan ook om een lerende overheid te creëren.

### 5.2.3 Politiek

Tot slot benoemt een groot deel van de respondenten dat ook politiek invloed heeft op de werking van de toets in de praktijk (R1a, R1b, R2, R6, R8, R9, R10, R11, R12). Dergelijke invloed vindt op verschillende momenten en bij verschillende aspecten van het proces plaats. Om te beginnen benoemen respondenten 7, 11 en 12 dat de Kamer steeds meer hoogte begint te krijgen van de plannen van de invoeringstoets en hier ook steeds vaker naar begint te vragen. Respondent 7 gaf hierbij aan dat één van de Kamercommissies een brief gestuurd heeft naar verschillende ministeries met de vraag waar de invoeringstoetsen blijven en wat de aanpak hiervan wordt. Daarnaast is ook al een motie ingediend om een invoeringstoets op een specifiek onderwerp te doen (R2). Respondent 8 spreekt hierover juist de angst uit dat het een politieke overweging wordt om een invoeringstoets uit te voeren in plaats van de selectie op inhoud te laten plaatsvinden.

Politieke belangstelling rondom een beleidsdossier lijkt ook mee te spelen bij het uitvoeren van een invoeringstoets. Respondent 1b merkte bijvoorbeeld dat de politieke en maatschappelijke belangstelling die er voor het onderwerp was in positieve zin invloed had op hoe snel de toets afgerond kon worden en hoe behulpzaam de verschillende stakeholders waren in het proces. Zij zegt hierover het volgende: *'ik denk, misschien dat als je een onderwerp zou hebben dat wat minder prominent speelt, dan vraag ik me af of we hem dan bijvoorbeeld in die 3 weken hadden kunnen uitvoeren'*. Daarmee komt ook de vraag op in hoeverre de toets goed uitgevoerd kan worden zodra die aandacht ontbreekt.

Verder zitten er achter beleid nou eenmaal toch vaak bepaalde politieke belangen. Zo benoemen respondenten 1a en 9 dat beleid soms ook onderdeel kan zijn van de politieke koers van een bewindspersoon of politieke partij. Beide respondenten denken dat het in die gevallen moeilijk wordt om te overtuigen dat het een slecht idee is. In bepaalde gevallen zal het beleid ook juist het resultaat zijn van een hevige politieke strijd. Respondenten 2 en 12 denken dat in die gevallen een invoeringstoets juist ook heel gevoelig kan liggen omdat je dan misschien dezelfde argumenten terug op tafel brengt waar al zolang over is gediscussieerd.

De invoeringstoets staat dus zeker niet los van de politieke omgeving waar die zich in bevindt. Vrijwel alle respondenten hebben benoemd dat het op zijn minst mogelijk is dat er bepaalde belangen van invloed zijn op het proces van de invoeringstoets, zoals bij de selectie van casussen, hoe meewerkend stakeholders zijn, en de mate waarin geluisterd wordt naar de uitkomsten. Respondenten 8, 9 en 12 vinden dan ook dat de

toets eigenlijk zo veel mogelijk gedepolitiseerd moet worden en dat de mate waarop invloed uitgeoefend kan worden geminimaliseerd moet worden.

#### 5.2.4 Conclusie

In officiële rapporten en Kamerbrieven wordt de werking van de invoeringstoets voornamelijk op rationele werking omschreven, maar in de praktijk ligt het toch echt wel ingewikkelder. De meeste aspecten van het instrument die genoemd zijn in de officiële berichtgeving werden ook veelvuldig door de respondenten genoemd. De invoeringstoets is een lichtvoetige ex durante evaluatie die relatief vroeg in het beleidsproces plaatsvindt. Met het instrument wordt gefocust op de werking van wet- en regelgeving in de praktijk en de gevolgen die het heeft voor de doelgroep (burgers en bedrijven) en uitvoering. Doordat de evaluatie relatief vroeg plaatsvindt kunnen op tijd aanpassingen worden gedaan aan de invulling van het beleid en in sommige gevallen zelfs aan het beleid zelf.

In de praktijk zijn veel aspecten echter nog onduidelijk of contextafhankelijk. Zo is het soms lastig vast te stellen op welk moment de toets moet plaatsvinden omdat deze op een te vroeg moment te weinig oplevert en er op een te laat moment al te veel schade aangericht kan zijn. Daarnaast zal om verschillende redenen niet iedereen even veel energie willen steken in het instrument of de resultaten ervan. Tot slot spelen er zeker ook politieke krachten die de werking van de toets beïnvloeden. Zo zullen niet altijd alle partijen even blij zijn zodra een invoeringstoets wordt uitgevoerd. Bijvoorbeeld omdat er al eerder politieke onderhandelingen zijn geweest waarvan de wet/regeling een onderdeel was. Hoewel de toets dus vooral rationeel wordt omschreven zal deze in de praktijk regelmatig worden verstoord door de omstandigheden waarbinnen deze wordt uitgevoerd en de persoonlijke belangen van de betrokken partijen.

### 5.3 Hoe en waar wordt de invoeringstoets benut binnen de Rijksoverheid?

Voor de tweede empirische deelvraag is bevestigd hoe en waar de resultaten van de onderzochte invoeringstoetsen (tot nu toe) zijn benut. In aanvulling daarop is gekeken naar de verwachtingen over in welke vorm en waar resultaten van de invoeringstoets zullen doorwerken. De respondenten werken allemaal al langer bij Rijksoverheid en hebben dus ook een eigen perspectief op beleid en politiek. De verwachtingen die hieruit voortvloeien zijn daarom een mooie aanvulling op de feitelijke doorwerking van de vier onderzochte invoeringstoetsen waarmee gezamenlijk de deelvraag beantwoord kan worden.

#### 5.3.1 Instrumentele doorwerking

Uit de gesprekken blijkt dat de benutting van de invoeringstoets voornamelijk voorkomt in de vorm van instrumentele doorwerking. Hiervan zijn twee voorbeelden uit de onderzochte invoeringstoetsen naar voren gekomen. Respondenten 1a en 1b zijn betrokken geweest bij een invoeringstoets van een uitvoeringsorganisatie waarbij de resultaten voor een deel ook zijn meegenomen in een latere wetswijziging. Zij hebben in de invoeringstoets een aantal oplossingsrichtingen opgenomen voor de geconstateerde knelpunten. Later is teruggekoppeld dat de wetswijziging deels kwam door de conclusies uit de invoeringstoets en deels door signalen die het ministerie elders opgevangen had. De respondenten benadrukten ook dat een deel van de conclusies dus níét is meegenomen. *'Tegelijkertijd met de invoeringstoets werden ook andere oplossingsrichtingen door het ministerie uitgedacht. De oplossingsrichting die werd aangedragen in de invoeringstoets is daardoor later*

*beperkt verder verkend en niet opgenomen in de wetswijziging'* (r1b). Dit had volgens de respondent te maken met de snelle ontwikkelingen op het thema van de invoeringstoets. Een andere invoeringstoets uitgevoerd door een ministerie (R2 was hierbij betrokken) had als conclusie dat een aspect van de geëvalueerde wet voor een deel van de doelgroep nog niet helemaal duidelijk was. Hierop hebben het ministerie en de betrokken uitvoeringsorganisatie afgesproken om hier beide beter over te communiceren.

Meer in het algemeen blijkt de invoeringstoets verschillende vormen van instrumentele doorwerking als doel te hebben. De meest voor de hand liggende vorm is een aanpassing in beleid. Volgens respondenten 11 en 12 kan dit, als het kleine aanpassingen zijn aan de invulling ervan, simpelweg worden gedaan door de betrokken persoon of personen bij de uitvoeringsorganisatie. Als er echt een aanpassing aan de regeling bij komt kijken moet dit volgens respondenten 3, 7, 9 en 12 echt belegd worden bij het ministerie. In zeldzame gevallen zal de invoeringstoets volgens respondenten 7, 8, 9 en 12 echt aanleiding zijn voor de wijziging van een wet en dan zal het ook echt door de Tweede Kamer moeten worden behandeld. Het is dus afhankelijk van wat uit de evaluatie komt, wat de vervolgstappen zijn en wie daar mee aan de slag moet.

### 5.3.2 Conceptuele doorwerking

Naast instrumentele doorwerking komt ook de conceptuele vorm van doorwerking voor. In de meeste gevallen betekent dit simpelweg dat andere partijen worden geïnformeerd over de resultaten van de toets. Onder conceptuele doorwerking valt immers ook een verandering in kennis van individuen of organisaties als gevolg van de resultaten van een evaluatie. Uitvoeringsorganisaties lijken vooral over de resultaten te rapporteren naar het betreffende ministerie (R1a, R3, R9) en het ministerie aan de Tweede Kamer (R2, R6, R7, R10). Daarbij benoemen een aantal respondenten dat er een informatieplicht is aan de Kamer en dat dit dus ook geldt voor de invoeringstoets (R6, R7, R10). Zo zegt respondent 7 bijvoorbeeld: *'Nou ja, kijk sowieso van alles wat wij doen. Ja daarover informeren we de Kamer. We hebben allerlei standen van de uitvoering, voortgangsrapportages etc. en in dat soort documenten zal je ook iets moeten zeggen over de uitkomsten of het verloop van de invoeringstoetsen, hè? Dus je hebt sowieso vanuit je informatievoorziening aan de Kamer een soort algemene verplichting om de Kamer te informeren over de voortgang van beleid of nieuw beleid.'* Een concreet voorbeeld uit een invoeringstoets is de Kamerbrief die respondent 2 naar de Kamer heeft gestuurd om te informeren wat de resultaten zijn en hoe het betreffende ministerie en de betreffende uitvoeringsorganisatie hierop acteren.

Naast pure informatievoorziening vindt ook intern conceptuele doorwerking plaats volgens respondent 10. Zij constateerde dat een recente evaluatie ook echt door veel collega's is gelezen en ook hier en daar besproken werd. Ook respondent 8 stelt dat zij intern lessen hebben getrokken uit een invoeringstoets en waar zij de volgende keer aandacht aan gaan besteden. Conceptuele doorwerking komt dus voornamelijk voor in de vorm van publicatie en verspreiding van de resultaten, maar eventuele lessen kunnen zeker ook intern worden besproken.

### 5.3.3 Agenderende doorwerking

De publicatie en verspreiding van resultaten kan ook nog een extra lading krijgen door dit met een agenderende intentie te doen. In dat geval is het dus ook de bedoeling dat de resultaten onderwerp van gesprek of zelfs debat worden. Deze agenderende vorm van doorwerking is minder vaak genoemd in de interviews. Onder andere in de invoeringstoets van respondent 2 is deze vorm voorgekomen. Hierin is geconstateerd dat de betreffende uitvoeringsorganisatie bij een bepaald aspect van de regeling te veel ruimte voor eigen invulling van de doelgroep had gelaten. De doelgroep heeft laten weten hierdoor juist niet te weten hoe ze het aan moeten pakken en dat ze hier richtlijnen voor willen. Het ministerie heeft dit vervolgens geagendeerd bij de uitvoeringsorganisatie die ermee aan de slag is gegaan. Daarnaast geven ook respondenten 1a, 5a, 8,



9 en 11 voorbeelden waarbij zij met evaluaties bepaalde zaken – al dan niet met die intentie – agenderen. Respondent 8 geeft aan agenderen ook echt als een functie van de invoeringstoets te zien. Hij zegt hierover: *'We zien het echt als een attenderingsmiddel inderdaad. [...] Ja en we sturen het natuurlijk ook naar de Kamer om het ook via die route in principe op de agenda te krijgen.'*

#### 5.3.4 Politiek-strategische doorwerking

Tot slot kan ook politiek-strategische doorwerking voorkomen. Dit lijkt wel de minst voorkomende vorm te zijn. Uit de gesprekken blijkt vooral dat de verschillende organisaties soms ook wel een eigen belang nastreven en hier ook naar handelen wat betreft de resultaten van een invoeringstoets of andere evaluaties. Respondent 9 zegt bijvoorbeeld het volgende: *'Uitvoeringsorganisaties hebben het op dit moment heel moeilijk. Ze hebben allemaal te maken met personeelstekorten, ingewikkelde wetgeving en continue aanpassing daarvan... Ze moeten ervoor zorgen dat ze als organisatie gewoon goed in het nieuws komen, ook al is het uiteindelijk de minister die verantwoording moet afleggen in de Kamer.'* Respondent 2 geeft aan dat uitvoeringsorganisaties soms politieke zaken opnemen in evaluaties die zij beleidsmatig graag willen zien. Het ministerie gaat daar volgens hem niet (altijd) in mee. Daarnaast geeft respondent 5b nog aan dat het in ieder geval al goed is om zaken in evaluaties aan te kaarten, al is het maar zodat zij er later niet op afgerekend kunnen worden. Ook hier is het dus vooral tactisch bedoeld en wordt niet altijd verwacht dat er daadwerkelijk aanpassing worden gedaan. Ook het ministerie gaat wel vaker 'tactisch' om met de resultaten van evaluaties. Respondent 8 noemt dat resultaten vaak intern vast met de opdrachtgevers besproken worden voordat ze naar de Kamer worden gestuurd zodat mogelijke oplossingsrichtingen vast besproken kunnen worden en het ministerie vast een reactie kan voorbereiden. Tot slot geven respondent 11 en 12 aan bang te zijn dat het instrument te gepolitiseerd zal worden: *'Ik denk dat zeker hoe de Kamer nu functioneert het vooral snel gepolitiseerd zal worden. Dat ze dus de minister erop af gaan rekenen dat hij of de uitvoering het niet goed gedaan heeft en dat de minister dan in het beklagdenbankje komt te zitten en ja, met een motie wordt opgedragen er zo snel mogelijk er iets mee te doen.'* (R12).

#### 5.3.5 Conclusie

Voor de benutting van de invoeringstoets kunnen vergelijkbare conclusies worden getrokken als voor de werking van de toets. De verwachting is dat deze vooral rationeel worden benut en instrumenteel doorwerken, terwijl deze in werkelijkheid ook op andere manieren (zullen) doorwerken. In eerste instantie zullen de resultaten van de invoeringstoets altijd worden aangeboden bij het ministerie/de uitvoeringsorganisatie (afhankelijk van wie de toets uitvoert) en/of de Tweede Kamer, waarmee het dus conceptueel doorwerkt omdat het onderdeel wordt van de kennis van de ontvangers. Dit kan in sommige gevallen ook door de uitvoeringsorganisatie of het ministerie worden gedaan met een agenderende intentie waarbij de bedoeling is dat het onderdeel wordt van gesprek. De meest concrete vorm van doorwerking is instrumentele doorwerking waarbij ook concrete actie wordt ondernomen naar aanleiding van de invoeringstoets. Het is niet altijd zo dat conclusies uit een invoeringstoets ook leiden tot actie. Het kan bijvoorbeeld ook zo zijn dat elders informatie is opgehaald die meegenomen wordt in hoe vervolg wordt gegeven aan een wet of regeling. Tot slot wordt door de respondenten aangegeven dat ook individuele belangen van betrokken partijen van invloed kunnen zijn op hoe de invoeringstoets benut wordt. Hoewel de focus op instrumentele doorwerking te begrijpen valt, komen de andere vormen dus ook zeker regelmatig voor.

## 5.4 In hoeverre en hoe wordt in de praktijk voldaan aan de randvoorwaarden voor het borgen van de benutting van de invoeringstoets?

Om de hoofdvraag van dit onderzoek te kunnen beantwoorden zal tot slot worden ingegaan op de derde en laatste empirische deelvraag. In paragraaf 3.3 is op basis van de literatuur een aantal randvoorwaarden opgesteld die naar verwachting gelden voor het borgen van de benutting van de invoeringstoets. Een groot deel van deze randvoorwaarden is regelmatig teruggekomen in de interviews, maar er is ook een aantal randvoorwaarden naar voren gekomen die helemaal niet uit de literatuur kwamen. Alle randvoorwaarden die uit de gesprekken zijn gekomen zullen nu per categorie individueel worden besproken.

### 5.4.1 Organische randvoorwaarden

#### **Bekendheid van de invoeringstoets**

*De invoeringstoets en diens functie moeten rijksbreed bekend zijn*

Verschillende respondenten benoemden in de interviews dat het belangrijk is dat men rijksbreed bekend is met de invoeringstoets (R1a, R2, R4, R6, R10, R11 en R12). In eerste instantie is dit van belang zodat de invoeringstoetsen überhaupt worden uitgevoerd, maar vervolgens ook zodat deze daadwerkelijk benut worden. Respondent 4 trekt hierbij de vergelijking met de *quick scan* en uitvoeringstoets die ook ooit geïntroduceerd zijn en vrij onbekend waren, maar die nu ook iedereen kent en gebruikt. Ze stelt dat als genoeg mensen weten dat het bestaat en het een nuttig instrument is dat het ook vanzelf benut wordt. Respondenten 10, 11 en 12 voegen hieraan toe dat niet alleen de invoeringstoets bekend moet zijn, maar dat vooral ook bekend moet zijn hoe die werkt en welk 'doel' het wel (en niet) heeft. Respondent 10 zegt hierover het volgende: *'Toch wel iets van verwachtingsmanagement rondom de impact van zo'n invoeringstoets (...) En ik denk dat het belangrijk is om bij zo'n invoeringstoets heel helder te hebben, wat zo een invoeringstoets wel teweeg kan brengen en wat niet.'* Die bekendheid is er inmiddels wel redelijk, maar dat kan zeker nog wel meer volgens respondent 11. Hier wordt wel aan gewerkt door middel van netwerkbijeenkomsten, kopgroep-overleggen en het geven van presentaties als daarom gevraagd wordt. Dat kan denk ze wel nog beter door bijvoorbeeld ook proactief organisaties te gaan benaderen.

*Men moet binnen Rijksoverheid overtuigd zijn van het nut van de invoeringstoets*

In aanvulling op de bekendheid van het instrument moet men ook overtuigd zijn van het nut ervan (R2, R4, R6, R7, R8, R11 en R12). Respondent 7 zegt hierover dat als het departement zelf ook het nut ervan inziet dat ook binnen de organisatie van het ministerie en de uitvoeringsorganisatie het idee ontstaat dat het een mooi meetinstrument is om beleid mee te volgen. Ook zou het volgens hem helpen om de Kamer te enthousiasmeren voor de eerste resultaten van beleid. De vraag is alleen in hoeverre het nut van de invoeringstoets inderdaad ingezien wordt binnen Rijksoverheid. De respondenten lijken hier namelijk nog niet helemaal van overtuigd. Respondent 12 gaat hier uitgebreid op in en stelt dat die overtuiging er nog nauwelijks is: *'Ja in mijn wetgevingservaring heb ik eigenlijk nooit iemand heel enthousiast gezien over evaluaties (...) Ja, we laten dat vaak aan externe partijen over, en we zijn er zelf niet echt bij betrokken. Ik denk dat op dat punt nog heel veel te winnen is. Dat is een randvoorwaarde. Zolang dat idee er is dan, is het echt een probleem voor de invoeringstoets'*. Meerdere respondenten waren wel bezig met de vraag hoe die overtuiging van het nut breder gedragen kan worden. Respondent 2 stelt bijvoorbeeld voor om via het

Rijkspitaal succesverhalen over de invoeringstoets te delen en respondent 6 had recent een Webinar gehad met als doel om collega's te inspireren ook met invoeringstoetsen aan de slag te gaan.

### **Relatie evaluator en ontvanger**

*De evaluator heeft persoonlijke eigenschappen die anderen enthousiasmeren*

Respondenten 10 en 12 noemden dat de persoonlijke eigenschappen van degene die de invoeringstoets leidt ook heel erg belangrijk zijn voor of zo'n toets ook benut wordt. Respondent 10 stelt dat er nog steeds met mensen gewerkt wordt en dat het dus wel helpt als iemand anderen mee kan krijgen, weet te verbinden en zich gecommitteerd voelt. Respondent 12 stelt dat diegene enthousiasme en een goede uitstraling moet hebben en ook respondent 11 erkent dat persoonlijkheid altijd een rol speelt. Zij vindt wel dat een invoeringstoets eigenlijk nooit door één iemand gedragen moet worden en dat er meerdere paar ogen mee moeten kijken.

*De uitvoeringorganisatie en het departement werken goed samen*

Verder is een randvoorwaarde voor de benutting van de invoeringstoets dat de uitvoeringsorganisatie en het ministerie die het dossier in de portefeuille hebben over het algemeen goed samen werken (R1a R3, R5b, R11, R12). Zodra twee organisaties of afdelingen hiervan vaker (nauw) samen werken zal ook de kans groter zijn dat open wordt gestaan voor de resultaten en dat daar wat mee gedaan zal worden. Respondent 12 stelt hierover dat de uitvoering zich gesteund moet weten door de beleidspraktijk omdat zij toestemming moeten krijgen om het op een andere manier te gaan doen. Bij een goede samenwerking zullen ze er sneller beide van doordrongen zijn dat het belangrijk is.

## **5.4.2 Politieke randvoorwaarden**

### **Overige Stakeholders**

*De relevante stakeholders zijn betrokken bij het proces en het uitvoeren van de toets*

Een andere randvoorwaarde die zeer belangrijk is voor de benutting van de invoeringstoets en ook door veel respondenten is genoemd is dat de relevante stakeholders ook zijn betrokken bij de toets (respondenten 1a, 1b, 2, 8, 9, 11 en 12). Dat geldt voor zowel het meenemen van de stakeholders in het proces als bij het uitvoeren van de toets zelf. Hierbij wordt voornamelijk genoemd dat het belangrijk is dat ook alle relevante stakeholders uit de doelgroep zijn meegenomen in de toets omdat dit cruciaal is voor representativiteit en daarmee de geloofwaardigheid ervan. Respondent 9 zegt hierover dat de evaluatie waar hij mee bezig is werkt met een begeleidingscommissie waar alle partijen die bij de wet/regeling betrokken zijn inzitten. Via die weg probeert hij voor draagvlak en acceptatie te zorgen van de uitkomsten, omdat die partijen dat anders eerder in hadden moeten brengen. Respondent 2 zegt over de betrokkenheid bij het proces het volgende: *'Maar dat is wel belangrijk. Dat je in ieder geval iedereen ook goed meeneemt, dus dat je van tevoren aankondigt van: oké we gaan zo'n invoeringstoets doen. We zijn van plan om jullie mee te nemen, maar ook die, die en die organisaties zodat ze dat van elkaar weten. En dat je zegt we komen over een maand terug met een enquête of met een interview zodat ze zich erop kunnen voorbereiden en dat iedereen goed op de hoogte is van het plan. Ja, ik denk dat dat ook wel belangrijk is.'*

## **Politiek-bestuurlijke context**

### *Het politieke klimaat werkt niet verhinderend voor het betreffende onderwerp*

Een randvoorwaarde die wellicht minder beïnvloedbaar is, maar wel geldt is dat het politieke klimaat de benutting van de resultaten niet tegenwerkt. Respondenten 1a, 4, 9, 11 en 12 stellen dat het politieke klimaat van invloed is op of en hoe de resultaten benut worden. Respondent 4 zegt hierover bijvoorbeeld dat of en welke aanbevelingen uit een evaluatie overgenomen worden beïnvloed wordt door geluiden uit de maatschappij en politiek. Volgens respondent 1a komt het ook vaker voor dat bepaalde aanbevelingen uit evaluaties niet worden overgenomen omdat deze niet in lijn liggen met de politieke koers van een bewindspersoon of bewindspersonen. Tot slot trekt respondent 11 het zelfs nog breder. Zij stelt dat de laatste jaren een focus ligt op de burger en het effect van regelgeving waardoor ook de invoeringstoets sneller benut wordt.

### *Het onderwerp heeft voldoende prioriteit*

Een veelgenoemde factor die bijdraagt aan de benutting van de invoeringstoets is prioriteit (R1a, R1b, R2, R4, R5b, R7, R10, R11, R12). Een deel van de respondenten gaf aan dat een invoeringstoets sneller benut wordt zodra het onderwerp voldoende prioriteit heeft (R1a, R1b, R2, R5b, R7 en R11). Zo is het volgens respondent 7 belangrijk dat de Kamer naar de invoeringstoets blijft vragen en er structurele belangstelling voor toont. Zodra zij alleen maar aandacht heeft voor de waan van de dag dan krijgt het instrument onvoldoende gewicht. Het andere deel van de respondenten gaf juist aan dat een toets juist minder snel wordt benut zodra er ook minder prioriteit voor is (R4, R7, R10). Respondent 4 zegt hierover het volgende: *'Als een uitvoeringsorganisatie zegt: wij willen dit graag onderzoeken door middel van een invoeringstoets. Dan moet dat wel goed worden afgestemd. Want als een beleidsdirectie dan al zegt van: oh ja, maar dat willen wij niet of gezien de prioriteit of politieke prioriteit kunnen we hier niks mee... dan is het weer zonde om die moeite in zo'n invoeringstoets te steken.'* Verder denkt ook respondent 12 dat het goed is voor de benutting als een onderwerp voldoende prioriteit heeft, maar dat het onderwerp ook weer niet té veel prioriteit moet hebben. Hij denkt namelijk dat als een onderwerp te veel aandacht en prioriteit heeft, er juist te snel wijzigingen aan het beleid worden aangebracht waardoor er geen ruimte meer is voor een invoeringstoets.

## **5.4.3 Rationele randvoorwaarden**

### **Het beleidsproces**

#### *De invoeringstoets wordt een standaard onderdeel van het beleidsproces bij bepaalde wet- en regelgeving*

Het aspect van de invoeringstoets waar in de interviews het meest over gesproken is, is de vraag of en in welke vorm de invoeringstoets een vast onderdeel wordt van de beleidscyclus. Alle respondenten bij wie het ter sprake is gekomen vonden dat de invoeringstoets op de een of andere manier een vast onderdeel moet worden van de beleidscyclus (R1a, R1b, R2, R3, R4, R6, R7, R8, R11, R12). De discussie zat hierbij vooral in de vraag op welke manier: verplicht of vrijblijvend, voor alle wet- en regelgeving of slechts bij een selectie en met een vast format of juist niet? Respondenten 1a, 2 en 3 zeiden hierover alleen dat ze vonden dat standaard relatief kort na inwerkingtreding een invoeringstoets plaats moet vinden en nuanceerden dit verder ook niet meer. De andere respondenten deden dit wel, maar een consensus over onder welke voorwaarden de invoeringstoets een vast onderdeel moet worden ontbrak (R1b, R4, R6, R7, R8, R11, R12). Respondent 1b dacht dat als je op alle nieuwe wet- en regelgeving een invoeringstoets zou dan, het er ook voor zou zorgen dat het een kwestie van een 'vinkje' wordt en het dus niet benut wordt. Hoe dat in een vast format gegoten kan worden wist ze alleen nog niet. Ook de andere respondenten die vonden dat niet altijd

een invoeringstoets moet worden gedaan gaven aan dat dit vooral alleen bij wet- en regelgeving moet worden waarbij het nuttig is. Voor welke wet- en regelgeving dan wel, noemden de respondenten vooral criteria die vaker terugkomen: het moet grote maatschappelijke impact hebben voor burgers, het moet een grote kans op knelpunten hebben of de uitvoeringstoets moet voor veel twijfels hebben gezorgd. Tot slot hadden verschillende respondenten nog ideeën om de invoeringstoetsen op een centrale plek te organiseren, zodat een partij hier waar nodig ook achteraan kan gaan. Respondent 1a zei bijvoorbeeld dat het ministerie van BZK hier een rol in zou kunnen spelen, *'net zoals ze dat eigenlijk bij de uitvoeringstoets hebben of zouden kunnen hebben en door gewoon ook, nou ja, collegaministeries erop aan te spreken van hé heb je al een invoeringstoets gedaan en wat doen jullie met de uitkomsten?'* Een voorwaarde is in ieder geval wel dat op zijn minst voor een deel van nieuwe wet- en regelgeving standaard een invoeringstoets wordt gedaan. De vraag is alleen nog hoe dit precies vorm zou kunnen en moeten krijgen.

### **Het evaluatieproces**

#### *De uitvoeringsorganisatie en het ministerie stemmen het evaluatieproces goed af*

De volgende genoemde randvoorwaarde ligt in het verlengde van de randvoorwaarde dat de uitvoeringsorganisatie en het ministerie over het algemeen goed samenwerken, namelijk dat zij ook het proces van de invoeringstoets goed afstemmen (R1, R4, R7, R8, R11, R12). Hierbij wordt vooral ook benoemd dat goed moet worden afgestemd dat überhaupt een invoeringstoets gedaan wordt, maar ook dat het kan helpen om de invoeringstoets echt helemaal in samenwerking uit te voeren. Daarnaast blijkt bij dit onderwerp dat er vooral ook nog onduidelijkheid is over bij wie de invoeringstoets belegd moet worden: de uitvoeringsorganisatie of het ministerie. Enerzijds is vaak genoemd dat het initiatief juist moet liggen bij de uitvoering omdat zij de signalen binnenkrijgen en dus ook beter kunnen bepalen wanneer een invoeringstoets gedaan moet worden (R7). Zo kan het resultaat van zo'n toets ook meer waarde krijgen en daarom misschien ook sneller benut worden. Anderzijds zeggen meerdere respondenten juist dat als het ministerie het initiatief neemt er ook sneller iets met de resultaten gebeurt, omdat zij toch vooral over het beleid gaan en er dan blijkbaar al aandacht voor hebben (R2, R4 en R12). Respondenten 7 en 11 denken dat dit weer per dossier het beste kan worden gezien. Bij wie het initiatief ook ligt, het moet in ieder geval voor beide partijen duidelijk zijn en idealiter ook geaccepteerd worden, voor de resultaten om benut te worden.

#### *De kenmerken van de wet of regeling lenen zich voor een invoeringstoets*

Een randvoorwaarden die ook vaker bij andere randvoorwaarden van invloed is, is dat de kenmerken van de betreffende wet of regeling zich moeten lenen voor een invoeringstoets (R4, R7, R9, R11, R12). In dit onderzoek is al vaker beargumenteerd dat de invulling van de invoeringstoets in hoge mate afhangt van de omstandigheden en context. Dit wordt met deze randvoorwaarde ook onderkent. Respondent 9 geeft het voorbeeld dat het afhangt van om wat voor beleid het gaat. *'Een wet aanpassen – ook al zijn het kleine dingen - daar zit veel looptijd achter... dat moet ook weer via de Tweede Kamer en alle andere stappen die bij een wet komen kijken. Daarnaast heb je besluiten en die zijn iets makkelijker aan te passen, en daaronder heb je regelingen en die zijn nóg makkelijker aan te passen.'* (R9). Respondent 12 noemt ook de vraag of het beleid heel strikt is geformuleerd of dat er ruimte is voor invulling, als factor. Bij strikt beleid is vaker ook heel bewust nagedacht over de contouren binnen het beleid waardoor veel afwegingen ook al vooraf zijn gemaakt, terwijl bij lossier beleid ook meer ruimte is voor knelpunten om op te spelen, die je dan met een invoeringstoets weer zou kunnen oplossen.

#### *Het evaluatieproces is kwalitatief goed uitgevoerd*

Een andere randvoorwaarde die door een groot deel van de respondenten genoemd is, is dat het evaluatieproces kwalitatief goed uitgevoerd

moet zijn (R5b, R7, R8, R9, R10, R11, R12). Deze voorwaarde is heel breed geformuleerd, omdat heel veel verschillende aspecten van het proces door de respondenten zijn genoemd die volgens de respondenten goed moeten zijn uitgevoerd voor de invoeringstoets om voldoende serieus genomen te laten worden. De volgende aspecten zijn voorbijgekomen: het proces moet transparant, duidelijk, objectief en zo veel mogelijk gedepolitizeerd zijn, het proces moet met voldoende deskundigheid zijn opgezet en er moet voldoende tijd worden genomen om terug te kijken en relevante data te verzamelen. Respondent 9 omschreef deze randvoorwaarde op pakkende wijzen: *'maar uiteindelijk gaat het erom dat die burgers goed bediend worden en dat perspectief en belang moet volgens mij vooral in zo'n invoeringstoets meegenomen worden, omdat dat de ultieme toets is of het echt werkt. En als dat onderzoek goed uitgevoerd is, dan kan niemand daar echt omheen.'*

Er schuilt echter wel nog een belangrijke tendens achter deze voorwaarde, namelijk wanneer de toets voldoende lichtvoetig is maar ook nog kwalitatief goed genoeg is opgezet. Respondent 11 denkt dat het goed mogelijk is om een toets lichtvoetig en kwalitatief goed op te zetten, maar erkent ook dat het een uitdaging wordt om het onderzoek niet te groot te maken. Ook respondenten 5b en 8 benoemen deze afweging, terwijl respondenten 5a en 12 juist met een oplossing komen. Hij stelt voor om er een 'tweetrapsraket' van te maken waarbij eerst een lichtvoetige toets wordt gedaan en extra onderzoek wordt gedaan als knelpunten naar voren komen om te kijken hoe groot die knelpunten zijn. Tot slot gaven respondenten 4, 6 en 12 aan dat het goed zou zijn om ook een kwantitatief aspect toe te voegen aan de invoeringstoets. Hiermee zouden de resultaten beter geduid kunnen worden en wordt het ook overzichtelijk hoe breed bepaalde knelpunten spelen. Dit zou het ook makkelijker maken om de toets te benutten.

## **Het resultaat**

*Het resultaat of de uitkomsten worden naar de juiste actoren verspreid.*

Tot slot benoemden nog verschillende respondenten dat de uitkomsten ook moeten worden verspreid naar de juiste actoren (R2, R8, R10, R12). Het klinkt misschien logisch, maar als je wil dat iets door iemand benut wordt dan moet diegene het wel ontvangen. *'Ik denk dat daarnaast gremium ook belangrijk is, dus los van het al dan niet onverkort delen met de Kamer, dat je zelf ook nog moet kijken naar van welke andere gremia je gebruik kunt maken om een onderzoeksrapport anderszins te agenderen'* (R10). De vraag is vooral ook naar wie het dan verspreid moet worden. Respondent 10 gaat hier verder op in en stelt dat er heel veel overleggen, commissies en andere samenwerkingsverbanden zijn waarbij je tijdens het proces input op kunt halen, maar waar je vervolgens ook over de resultaten in gesprek kan gaan en kunt bepalen wat je het best met die resultaten kan doen. Ook respondent 12 beschouwt dit als een randvoorwaarde, zeker ook met het oog op de knelpunten uit 'Werk aan Uitvoering'. Daarin werd geconstateerd dat vaker tegen een soort plafond opgelopen werd en dat onderwerpen belangrijk genoeg moesten zijn om bij de directeur-generaal onder de aandacht te brengen en die moest het dan weer belangrijk genoeg vinden om onder de aandacht te brengen bij de beleidsdirecteur. *'Dus er zit altijd een soort filter op. Dus als jij niet in staat bent om bepaalde knelpunten uit je uitvoeringspraktijk in de invoeringstoets te krijgen of als het in de invoeringstoets staat, maar de invoeringstoets wordt als te onbelangrijk gezien om onder de aandacht te brengen bij de betrokken beleidsdirecteur, directeur-generaal of minister. Dan blijft het hangen en dan krijgt het niet de aandacht die het nodig heeft om zo'n knelpunt op te lossen. En datzelfde geldt ook met ja binnenskamers houden of openbaar maken. Als dingen openbaar gemaakt worden heb je toch de kans dat iemand het oppikt en daar een vraag over gaat stellen. En dan zorgt het op die manier voor een soort externe druk, waardoor het dan toch ook weer bij de eigenaar terecht gaat komen'* (R12).

Het zou volgens respondent 12 dus ook kunnen helpen om het standaard openbaar te maken en wijst er daarbij op dat dat sowieso is hoe de overheid zou moeten functioneren. Het zou wellicht dus kunnen helpen om de invoeringstoetsen standaard op een centrale plek te publiceren.

## 5.5 Conclusie

Na alle randvoorwaarden te hebben besproken kan nu eindelijk ook de laatste empirische deelvraag worden beantwoord: *'In hoeverre en hoe wordt in de praktijk voldaan aan de randvoorwaarden voor het borgen van de benutting van de invoeringstoets?'* Alle randvoorwaarden voor het borgen van de benutting van de invoeringstoets staan in de tabel hieronder weergegeven. De meeste van deze randvoorwaarden zijn casusafhankelijk of hebben nog geen invulling gekregen binnen de toets, waardoor nu geen antwoord kan worden geformuleerd op in hoeverre en hoe eraan wordt voldaan. Voor de overige voorwaarden kan over het algemeen worden gesteld dat er nog niet goed aan wordt voldaan. Uit de interviews blijkt dat de invoeringstoets en met name diens functie nog niet bij iedereen bekend is. Daarnaast is vooral ook niet iedereen overtuigd van het nut ervan. Zo werd regelmatig in twijfel getrokken of men het wel zag zitten om er energie in te steken. Eén van de redenen hiervoor is het gebrek aan capaciteit. Men is al druk genoeg met de taken die ze al hebben en dan moeten ze ook nog aan de slag met een evaluatie. Dit is ook één van de meest genoemde argumenten om niet zomaar overal een invoeringstoets op te doen, maar dit op de één of andere manier te selecteren. Verder werd ook genoemd dat betrokkenen geen zin hebben om een wet achteraf weer in twijfel te trekken wanneer deze net na veel discussie eindelijk tot stand is gekomen. Dan zou de discussie weer aangewakkerd kunnen worden waar men dan niet op zit te wachten.

Een voorwaarde waar momenteel wel aan wordt voldaan is dat van een gunstig politiek klimaat. In de interviews kwam namelijk naar voren dat er momenteel veel aandacht uitgaat naar de invloed van wetgeving op burgers en dus ook de invoeringstoets zeer relevant is. Dit blijkt ook uit de verschillende opmerkingen dat de Kamer momenteel veel oog heeft voor het instrument. De vraag is natuurlijk wel of die aandacht in de toekomst aanhoudt. Voor de overige randvoorwaarden geldt dat deze afhankelijk zijn van de verdere invulling van de invoeringstoets, waar dus ook in het licht van deze randvoorwaarden naar gekeken moet worden.

Perspectieven	Categorieën	Randvoorwaarden
<b>Organisch</b>	De bekendheid van de invoeringstoets	1. De invoeringstoets en diens functie moeten rijksbreed bekend zijn
		2. Men moet binnen Rijksoverheid overtuigd zijn van het nut van de invoeringstoets
	De relatie tussen de evaluator en ontvanger	3. De evaluator heeft persoonlijke eigenschappen die anderen enthousiasmeren
		4. De uitvoeringorganisatie en het departement werken goed samen
<b>Politiek</b>	Overige stakeholders	5. De relevante stakeholders zijn betrokken bij het proces en het uitvoeren van de toets
	De politiek-bestuurlijke context	6. Het politieke klimaat werkt niet verhinderend voor het betreffende onderwerp

		7. Het onderwerp heeft voldoende prioriteit
<b>Rationeel</b>	Het beleidsproces	8. De invoeringstoets wordt een standaard onderdeel van het beleidsproces bij bepaalde wet- en regelgeving
	Het evaluatieproces	9. De uitvoeringsorganisatie en het ministerie stemmen het evaluatieproces goed af
		10. De kenmerken van de wet of regeling lenen zich voor een invoeringstoets
		11. Het evaluatieproces is kwalitatief goed uitgevoerd
Het resultaat	12. Het resultaat of de uitkomsten worden naar de juiste actoren verspreid.	



## 6. Conclusies en aanbevelingen

### 6.1 Inleiding

Dit hoofdstuk reflecteert op de resultaten en beargumenteert wat deze impliceren. Allereerst wordt aan de hand van de verschillende deelconclusies uit het theoretisch kader en het voorgaande hoofdstuk antwoord gegeven op de verschillende deelvragen. Deze zullen vervolgens samen worden gebracht om te komen tot een antwoord op de hoofdvraag uit dit onderzoek. Hierbij wordt ook besproken wat deze conclusies betekenen, zowel voor de implementatie van de invoeringstoets als voor de evaluatieliteratuur. Op basis hiervan wordt een aantal aanbevelingen gedaan, waarna een slotwoord volgt.

### 6.2 Deelvragen beantwoord

Nu alle deelvragen beantwoord zijn kunnen deze bij elkaar worden gebracht om uiteindelijk de hoofdvraag te beantwoorden. Elke deelvraag wordt nu eerst los beantwoord en besproken om vervolgens in paragraaf 6.3 ook de hoofdvraag te beantwoorden.

#### 6.2.1 Wat is de invoeringstoets en hoe is deze te positioneren in de beleidscyclus?

De invoeringstoets is volgens officiële berichtgeving een nieuw ex durante evaluatie-instrument waarmee ongeveer een jaar na inwerkingtreding van wet- en regelgeving teruggeblikt wordt op de werking ervan in de praktijk. Het instrument moet knelpunten vanuit de beleidsuitvoering direct terugleggen bij de beleidsvoorbereiding. Hiermee kan het beleid of de wet- en regelgeving ook direct aangepast worden om zo de knelpunten te verhelpen. Op papier wordt de toets echter zeer rationeel omschreven, terwijl het vanuit een organische of politieke blik een heel ander aanzicht krijgt. De invoeringstoets bevindt zich namelijk in een dynamische omgeving waar veel belangen een rol spelen. De werking van het instrument is daardoor enerzijds sterk afhankelijk van de context en omstandigheden waarin het zich bevindt. Hierbij spelen bijvoorbeeld andere beleidsdossiers, politieke commitment en momentum een rol en is het maar de vraag of er een jaar na inwerkingtreding nog voldoende momentum is of momentum kan worden gevonden om beleid nog te wijzigen. Anderzijds is de werking van het instrument afhankelijk van het handelen van de verschillende actoren en de doelen die zij nastreven. Een puur rationele focus geeft al met al geen realistisch beeld.

Het idee achter het instrument is echter wel passend. Juist door beleid en wet- en regelgeving structureel te evalueren en ook in te bouwen dat de conclusies hiervan structureel teruggelegd worden bij de beleidsmaker, kan structurele herziening worden bewerkstelligd. Indien nodig kan de wet of regeling dan worden aangepast om de knelpunten te verhelpen.

#### 6.2.2 Wat is benutting van evaluatie en welke benaderingen voor benutting zijn te onderscheiden?

Uit de literatuur blijkt dat voor het concept benutting geen eenduidige definitie bestaat, maar dat doorgaans de zeven modellen van Weis (1979) worden gebruikt om het concept te definiëren. Daarnaast kan een onderscheid worden gemaakt in verschillende vormen van doorwerking (Bekkers et al., 2004, pp. 24-26) om de benutting beter te analyseren, namelijk in instrumentele, conceptuele, agenderende en politiek-strategische doorwerking. Op basis van de Kamerbrief van Weerwind (2022a) en andere officiële berichtgeving over de invoeringstoets lijkt het hoofddoel voornamelijk instrumentele doorwerking te zijn. De brief benoemt namelijk vooral hoe de toets moet zorgen voor herziening en wijziging van wet- en regelgeving en dat daarnaast de uitvoering signalen van knelpunten terug kan leggen bij beleid- en besluitmakers. Daarbij

wordt wel het risico op een té instrumentele benadering erkend, maar daarmee wordt bedoeld op de gebruikte methode voor het onderzoek en niet de werking van de toets zelf. Conceptuele en politiek-strategische doorwerking worden dan ook volledig achterwegen gelaten, terwijl deze vormen op het eerste oog - gezien de dynamische omgeving en de vele belangen die spelen - ook zeker niet nutteloos en onwaarschijnlijk lijken.

### 6.2.3 Welke randvoorwaarden voor het borgen van benutting van de invoeringstoets zijn te onderscheiden?

Voor de laatste theoretische deelvraag is een aantal factoren die waarschijnlijk ook gelden voor de invoeringstoets, afgeleid uit de artikelen van Bekkers et al. (2004) en Johnson et al. (2009). Deze zijn vertaald naar randvoorwaarden en staan weergegeven in onderstaande tabel.

<b>Perspectieven</b>	<b>Categorieën</b>	<b>Belangrijkste randvoorwaarden</b>
<b>Rationeel</b>	Het evaluatieproces	1. Het evaluatieproces zit kwalitatief goed in elkaar, volgens verfijnde methoden en een relevant evaluatiemodel. 2. Het type beleidsprobleem/besluit is relatief nieuw en er zijn nog weinig meningen over gevormd of de meningsverschillen over het onderwerp zijn onoverbrugbaar waardoor een invoeringstoets juist een oplossing biedt. 3. Het evaluatieproces wordt goed en duidelijk afgestemd tussen het departement en de uitvoeringsorganisatie.
	Het evaluatierapport	4. De resultaten worden duidelijk en wijd verspreid onder relevante actoren. 5. De evaluatie wordt afgerond en verspreid op een tactisch moment.
<b>Organisch</b>	De relatie tussen de evaluator en ontvanger van de evaluatie	6. De ontvanger heeft persoonlijke eigenschappen die gunstig zijn voor benutting (organisatorische rol, informatieverwerkingsstijl, sociale eigenschappen). 7. De ontvanger is betrokken en ontvankelijk naar de evaluatie
<b>Politiek</b>	Overige stakeholders	8. Overige stakeholders worden betrokken in het (opzetten van het) proces van de evaluatie.
	De overkoepelende politiek-bestuurlijke context	9. De evaluatie vindt plaats en wordt afgerond en verspreid in een gunstig politiek klimaat. 10. De evaluatie krijgt voldoende (maar niet te veel) politieke belangstelling.

#### 6.2.4 Hoe past de invoeringstoets binnen de beleidscyclus van de Rijksoverheid?

In officiële rapporten en Kamerbrieven wordt de werking van de invoeringstoets voornamelijk met een rationele werking omschreven, maar in de praktijk ligt het ingewikkelder. De meeste aspecten van het instrument die op papier genoemd zijn werden ook veelvuldig door de respondenten genoemd: 'lichtvoetig', 'ex durante', 'focus op werking in de praktijk' etc. In werkelijkheid zijn veel onderdelen van de toets nog onduidelijk of contextafhankelijk. Zo is het soms lastig vast te stellen op welk moment de toets moet plaatsvinden omdat deze op een te vroeg moment te weinig oplevert en op een te laat moment al te veel schade aangericht kan zijn. Daarnaast zal om verschillende redenen niet iedereen even veel energie willen steken in het instrument of de resultaten ervan. Tot slot spelen er ook politieke krachten die de werking van de toets beïnvloeden. Zo zullen niet altijd alle partijen even blij zijn zodra een invoeringstoets wordt uitgevoerd. Bijvoorbeeld omdat er al eerder politieke onderhandelingen zijn geweest waarvan de wet/regeling een onderdeel was. Hoewel de toets dus vooral rationeel wordt omschreven zal deze in de praktijk regelmatig worden 'verstoord' door de omstandigheden waarbinnen deze wordt uitgevoerd en de persoonlijke belangen van de betrokken partijen.

#### 6.2.5 Hoe en waar wordt de invoeringstoets benut binnen de Rijksoverheid?

Voor de benutting van de invoeringstoets kunnen vergelijkbare conclusies worden getrokken als voor de werking van de toets. De verwachting is dat deze vooral rationeel wordt benut en instrumenteel doorwerkt, terwijl deze in werkelijkheid waarschijnlijk ook op andere manieren doorwerkt. In eerste instantie zullen de resultaten van de invoeringstoets altijd worden aangeboden bij het ministerie/de uitvoeringsorganisatie (afhankelijk van wie de toets uitvoert) en/of de Tweede Kamer. Hiermee werkt het dus conceptueel door, omdat het onderdeel wordt van de kennis van de ontvangers. Dit kan in sommige gevallen ook door de uitvoeringsorganisatie of het ministerie worden gedaan met een agenderende intentie waarbij het onderwerp moet worden van gesprek. De meest concrete vorm van doorwerking is instrumentele doorwerking waarbij ook concrete actie wordt ondernomen naar aanleiding van de invoeringstoets. Het is niet altijd zo dat conclusies uit een invoeringstoets ook leiden tot actie. Het kan bijvoorbeeld ook zo zijn dat elders informatie is opgehaald die meegenomen wordt in hoe vervolg wordt gegeven aan een wet of regeling. Tot slot wordt door de respondenten aangegeven dat ook individuele belangen van betrokken partijen van invloed kunnen zijn op hoe de invoeringstoets benut wordt. Hoewel de focus op instrumentele doorwerking te begrijpen valt, komen de andere vormen dus ook regelmatig voor.

#### 6.2.6 In hoeverre en hoe wordt in de praktijk voldaan aan de randvoorwaarden voor het borgen van de benutting van de invoeringstoets?

Alle randvoorwaarden voor het borgen van de benutting van de invoeringstoets zijn weergegeven in paragraaf 5.5. De meeste van deze randvoorwaarden zijn casusafhankelijk of hebben nog geen invulling gekregen binnen de toets, waardoor nu geen antwoord kan worden geformuleerd op in hoeverre en hoe eraan wordt voldaan. Voor de overige voorwaarden kan over het algemeen worden gesteld dat er nog niet goed aan wordt voldaan. Uit de interviews blijkt dat de invoeringstoets en met name diens functie nog niet rijksbreed bekend zijn. Daarnaast is vooral ook niet iedereen overtuigd van het nut ervan. Zo werd regelmatig in twijfel getrokken of men het wel zag zitten om energie in de toets te steken. Eén van de redenen hiervoor is het gebrek aan capaciteit: ambtenaren zijn al druk genoeg met de taken die ze reeds hebben en dan moeten ze ook nog aan de slag met een evaluatie. Dit is ook één van de meest genoemde argumenten om niet zomaar op alle wet- en regelgeving een invoeringstoets te doen, maar op de één of andere manier een selectie te maken. Verder werd benadrukt dat betrokkenen

geen 'zin' hebben om een wet achteraf weer in twijfel te trekken wanneer deze net na veel discussie eindelijk is ingevoerd. Dan zou de discussie weer aangewakkerd kunnen worden en daar zit men niet altijd op te wachten.

Een voorwaarde waar momenteel wel aan wordt voldaan is dat van een gunstig politiek klimaat. In de interviews kwam namelijk naar voren dat momenteel veel aandacht uitgaat naar de invloed van wetgeving op burgers en dus ook de invoeringstoets zeer relevant is. Dit blijkt ook uit de verschillende opmerkingen dat de Kamer momenteel veel oog heeft voor het instrument. De vraag is natuurlijk wel of die aandacht in de toekomst aanhoudt. Voor de overige randvoorwaarden geldt dat deze afhankelijk zijn van de verdere invulling van de invoeringstoets, waar dus ook in het licht van deze randvoorwaarden naar gekeken moet worden.

### 6.3 Hoofdvraag beantwoord

Na het beantwoorden van alle deelvragen kan ook antwoord worden gegeven op de hoofdvraag: *Onder welke randvoorwaarden kan de benutting van de invoeringstoets worden geborgd?*

De invoeringstoets is op papier een ex durante evaluatie met een zeer rationeel-analytische gedachtegang: door vroeg in het beleidsproces te kijken naar de werking van wet- en regelgeving kunnen knelpunten al snel worden gesignaleerd en worden teruggelegd bij de beleid- en besluitmakers om ze op te lossen. Wanneer de respondenten het over de invoeringstoets hadden, verwezen zij vooral naar de formuleringen zoals die in de officiële berichtgeving van onder andere de Kamerbrief van Weerwind (2022) worden gebruikt. Hierbij werden voornamelijk de rationele/instrumentele en agenderende (door)werking omschreven. De genoemde doelen van het instrument zijn dan ook 1) het kabinet in staat stellen om eerder en beter in te spelen op signalen uit de doelgroep over de werking van nieuwe regelgeving in de praktijk en 2) het makkelijker maken voor de uitvoering om knelpunten uit de uitvoeringspraktijk terug te leggen op de beleidstafel. De respondenten vertelden echter tussen de onderwerpen door dat ook de organische/politieke werking en andere vormen van doorwerking regelmatig voorkomen. Sterker, de respondenten zijn zich ook bewust dat de omgeving een sterke invloed heeft, dat veel actoren de toets voor hun eigen doelen gebruiken of misbruiken en dat de toets ook politiek-tactisch benut wordt. Bij de vertaling van papier naar praktijk lijkt een soort *rationaliseringsreflex* te zitten waarbij bijna krampachtig op het rationele wordt gefocust.

Deze rationaliseringsreflex heeft een paradox als gevolg die er juist voor zorgt dat de toets moeilijk van de grond komt. De implementatie van de invoeringstoets is namelijk zeer ingewikkeld omdat deze werkbaar moet worden gemaakt voor verschillende contexten, verschillende wet- en regelgeving en verschillende organisaties. Doordat het zo ingewikkeld en complex is heeft men de neiging om de implementatie heel rationeel te bekijken en aan te pakken. Tegelijkertijd wordt het juist ingewikkeld dóór deze rationele aanpak, omdat juist daardoor de dynamische omgeving, politieke belangen, omstandigheden etc. de kans krijgen om op te spelen. De werkelijkheid is namelijk - zoals de respondenten zelf ook laten blijken - niet zo rationeel en de toets werkt nou eenmaal niet precies zoals deze op papier staat omschreven. De realiteit is dat de rationele werking van de invoeringstoets in grote mate wordt beïnvloed en verstoord door de context en politieke belangen. Zo gaven verschillende respondenten bijvoorbeeld aan dat niet iedereen energie in de toets wil steken omdat 1) ze het al ontzettend druk hebben met de rest van het werk, en/of 2) omdat het beleid na lang onderhandelen staat en 'stabiel' is, en/of 3) ze discussies niet opnieuw of onnodig willen aanwakkeren. Ofwel, meer gevoelsmatig, 1) 'laat me met rust', 2) 'het beleid staat', en 3) 'geen slapende honden wakker maken'. Er speelt dus een rationaliseringsparadox die de werking van de toets juist in de weg zit.

Bij de implementatie van het instrument wordt echter niet helemaal puur rationeel gekeken. In de vertaling van de werking op papier naar de praktijk wordt wel degelijk rekening gehouden met het feit dat de invoeringstoets niet in elke context hetzelfde zal werken. Bij het rijksbreed implementeren van het instrument is gekozen om dit middels een decentrale aanpak te doen waarbij alle departementen en uitvoeringsorganisaties zelf methoden ontwikkelen die in de eigen context werkbaar zijn. Hierbij speelt echter een andere paradox op. Deze decentrale aanpak heeft namelijk als gevolg dat de invoeringstoets juist een heel abstract, vaag en voor alles generiek idee blijft. Een duidelijke aanpak ontbreekt, het is niet duidelijk bij welke wet- en regelgeving de toets uitgevoerd wordt en het is niet duidelijk wanneer de toets precies met de Kamer gedeeld moet worden. Daarnaast geven respondenten aan dat bij veel collega's – maar ook bewindspersonen – het precieze doel en het 'nut' van de toets niet duidelijk is. Hierdoor lijkt het vooral bij betrokkenen te leven waardoor het instrument meer een soort algemeen ritueel is dat symbool staat voor de nieuwe bestuurscultuur, en waar decentraal weinig handelingspotentieel aan te koppelen is. Juist omdat die toets decentraal wordt ingevoerd krijgt hij dus lastig gestalte en blijft het niet uitnodigend om met het instrument aan de slag te gaan. Een centrale aanpak zou tegelijkertijd juist zorgen voor een slechte aansluiting op de uitvoeringspraktijk, waarmee dus sprake is van een handelingsparadox.

Grof gezegd zijn er twee opties voor het vervolg van de invoeringstoets. Of het wordt geaccepteerd dat de toets een symbolisch instrument is dat past in zijn tijd en onderdeel is van de Werk aan Uitvoering-'hype', maar dat over een aantal jaar ook weer vervlogen is. Of er wordt meer nadruk gelegd op de organische en politieke kanten, vooral ook decentraal, om er 'echt wat' van te maken. Door juist ook de werking vanuit de andere perspectieven te erkennen en hier meer focus op te leggen kunnen namelijk de invloeden van factoren als persoonlijke belangen, de relaties van evaluatoren en ontvangers en persoonlijke eigenschappen van de evaluator worden beperkt. Dan is het alleen wel ook noodzakelijk dat er middelen aan gewijd worden zoals tijd, energie en capaciteit en – nog belangrijker – dat *routines* worden opgebouwd. Dan is duidelijkheid vereist over wanneer een invoeringstoets gedaan wordt, volgens welke methoden en wanneer de resultaten met wie gedeeld worden.

Om de toets echt tot een succes te maken moet het voor iedereen duidelijk zijn wat de toets precies is en hoe deze werkt. Op dit moment krijgt de toets, zo kunnen we concluderen, nog niet de structurele aandacht die geambieerd wordt. De toets heeft vooralsnog te weinig gewicht, in politieke en ambtelijke werelden waarin al meer dan genoeg om aandacht wordt gestreden.

## 6.4 Aanbevelingen

Nu de belangrijkste conclusies zijn gepresenteerd wordt nog een aantal aanbevelingen gedaan. Hierbij worden eerst vooral maatschappelijke en meer praktische aanbevelingen gedaan, waarna ook wordt besproken welk vervolgonderzoek nog kan worden gedaan.

### 6.4.1 Maatschappelijke aanbevelingen

De invoeringstoets is een instrument met veel potentie die momenteel onvoldoende wordt benut. Om de toets echt te laten werken zoals die is bedoeld is het van belang om uit de genoemde paradoxen te stappen. Voor de verdere implementatie moet de nadruk juist minder liggen bij het rationele, maar bij het organische en politieke. Ga niet kijken hoe het instrument van papier naar praktijk kan worden vertaald, maar kijk juist naar hoe papier en praktijk van elkaar verschillen. Focus op waarom het instrument en het nut ervan nu nog onvoldoende bekend zijn binnen de Rijksoverheid, waarom men geen energie wil steken in een nieuwe evaluatie, en waarom de Kamer bij bepaalde onderwerpen

geïnteresseerder is dan bij anderen. Juist dan kun je die problemen ook gaan oplossen. Meer concreet kan een tweetal aanbevelingen worden gedaan om de benutting te vergroten.

Allereerst is het van belang dat de toets concreter vorm krijgt. Door het instrument verder in te kleuren wordt het voor organisaties interessanter om het instrument te gebruiken en wordt het vooral ook tastbaarder. Zodra iets tastbaar wordt zal het ook sneller aan bekendheid winnen en meer wind vangen. Zodra de toets vervolgens vaker wordt gebruikt komen ook sneller de beperkingen van de invulling naar voren en kan hier makkelijker op worden ingespeeld. Het is dus wel van belang dat een open blik wordt gehanteerd en dat men zich niet blind staart op de gekozen aanpak. Als de toets vaker wordt gebruikt krijgt men ook de kans om meer routines en gebruiken op te bouwen waardoor de toets ook rationeler zal gaan werken. Natuurlijk is de toets in hoge mate afhankelijk van de context, maar juist door deze concreet in te vullen wordt het makkelijker om ervan af te wijken zodra de omstandigheden hierom vragen. Het zou dus goed zijn als de projectmanager met de kopgroep voor de invoeringstoets aan de slag gaat met een concrete invulling van de invoeringstoets die enerzijds wel redelijk algemeen werkbaar is, maar tegelijkertijd ook voldoende gedetailleerd is om houvast te bieden.

Daarnaast is het van belang dat de focus verschuift van het rationele naar het organische en politieke. Juist wanneer de chaos beperkt wordt krijgt de rede een kans. Voor de verdere inkleuring van het instrument moet de projectmanager - maar ook de vele organisaties die aan het werken zijn aan passende methoden – juist opzoek gaan naar de factoren die de rationele werking van het instrument verstoren. Ga juist aan de mensen vragen die het nut er niet van in zien, waarom zij het nut er niet van in zien; ga aan de Kamer vragen waarom zij de resultaten van een invoeringstoets niét hebben benut in plaats van wel; ga aan ministeries vragen waarom zij bepaalde aanbevelingen niét over hebben genomen van uitvoeringorganisaties in plaats van wél. De nadruk moet dus niet liggen op hoe het goed kan gaan, maar op hoe het fout kan gaan. Zo kunnen de beperkingen van de toets worden aangepakt en kan deze juist gaan werken zoals die is bedoeld.

#### 6.4.2 Wetenschappelijke aanbevelingen

Naast maatschappelijke aanbevelingen kan ook een aantal wetenschappelijke aanbevelingen worden gedaan. Dit onderzoek geeft inzicht in de verschillende vormen van doorwerking en welke vormen van doorwerking voorkomen bij de invoeringstoets. Het doel van de invoeringstoets is voornamelijk instrumentele en voor een deel agenderende doorwerking. In de Kamerbrief van Weerwind (2022-a) wordt daarnaast duidelijk dat men voornemens is om goed na te denken over wanneer een invoeringstoets zinvol is en wanneer niet. Bij teveel toetsen bestaat namelijk het risico dat het te veel capaciteit vraagt en eerder een kwestie van afvinken wordt. De vraag die dan opspeelt is wanneer een invoeringstoets zinvol is. Aangezien de focus op instrumentele en agenderende doorwerking ligt zou het dus relevant zijn om te onderzoeken bij welke casussen deze vormen van doorwerking voorkomen en welke niet. Op die manier kan worden onderzocht welke factoren ervoor zorgen dat een invoeringstoets inderdaad instrumenteel of agenderend doorwerkt en bij welke factoren juist niet. Middels een *casestudy* zouden dus verschillende casussen van invoeringstoetsen kunnen worden onderzocht en vergeleken. Hierbij zou bijvoorbeeld een politiekgevoelige casus vergeleken kunnen worden met een casus die juist niet is. Of een casus met een invoeringstoets die gedaan is door een ministerie met een casus van een invoeringstoets die gedaan is door een uitvoeringsorganisatie. Op die manier kan dus worden achterhaald welke factoren bijdragen aan de verschillende vormen van doorwerking en hoe de gewenste vorm geoptimaliseerd kan worden.

## 6.5 Tot slot

Dit onderzoek begon met een korte introductie over hoe de invoeringstoets werd aangekondigd als reactie op het rapport over de kinderopvangtoeslagaffaire. Iets meer dan een jaar later startte dit onderzoek terwijl nog niet veel over de invoeringstoets bekend was. Hierdoor werd juist gedurende het proces steeds meer bekend over de implementatie en vormgeving van het instrument. Bij aanvang waren er amper invoeringstoetsen gedaan, stonden er ook slechts een aantal op de planning en had ook de Kamer het woord invoeringstoets pas een enkele keer laten vallen. Inmiddels zijn er verschillende Kamerbrieven over gestuurd, staan er tientallen invoeringstoetsen op de planning en vraagt ook de Tweede Kamer regelmatig naar de voortgang. De invoeringstoets begint dus wel degelijk te leven en de implementatie begint steeds meer wind te vangen. Juist daarom is het ook zaak om nu door te pakken en juist ook de politieke en organische kanten te adresseren, met nadruk op het gewicht geven aan de toets, in werelden die er gewichtig over spreken.

## Bronnenlijst

Bekkers, V., Fenger, M., Homburg, V. & Putters, K. (2004). Doorwerking van Strategische Beleidsadviesing. Geraadpleegd van: <https://repub.eur.nl/pub/1594/BSK-CPG-2004-001.pdf>

Bressers, J. T. A., & Hoogerwerf, A. (1995). *Beleidsevaluatie* (3<sup>e</sup> dr, Ser. Serie maatschappijbeelden). Samsom H.D. Tjeenk Willink.

Cousins, J. B., & Leithwood, K. A. (1986). Current Empirical Research on Evaluation Utilization. *Review of Educational Research*, 56(3), 331–364.

Derksen, W. & Fleurke, F. (2010). Benutting van bestuurskundig advies en de onafhankelijkheid van de onderzoeker. In Aardema, H., Derksen, W., Herweijer, M., de Jong, P. (Eds.), *Meerwaarde van de bestuurskunde* (pp. 165-174). Den Haag: Boom Lemma uitgevers.

Heerschop, D. & van der Vlist, H. (2020a). *Werk aan Uitvoering: Fase 1: Probleemanalyse*. Den Haag: Algemene bestuursdienst, Ministerie van Binnenlandse Zaken en Koninkrijksrelaties.

Heerschop, D. & van der Vlist, H. (2020b). *Werk aan Uitvoering: Fase 2: Handelingsperspectieven en samenvatting analyse*. Den Haag: Algemene bestuursdienst, Ministerie van Binnenlandse Zaken en Koninkrijksrelaties.

Hoogerwerf, A. (2014). *Beleid, Processen en effecten*. Hoogerwerf, A., Herweijer, M. (EDS), *Overheidsbeleid. Een inleiding in de beleidswetenschap* (15-28). Alphen aan den Rijn: Kluwer.

Janssen, G.S.H., Hellendoorn, J.C. (1999). *Heeft beleid effect? Enkele benaderingen voor de evaluatie van bestaand beleid*. Den Haag. SDU.

Johnson, k., Greenesid, L.O., Toal, S.A., King, J.A., Lawrenz, F. & Volkov, K. (2009). Research on Evaluation Use: A Review of the Empirical Literature from 1986 to 2005. *American Journal of Evaluation*. 30(3), 377-410

Kenniscentrum voor beleid en regelgeving. (2022, 29 april). Ex post evaluatie. Geraadpleegd van: <https://www.kcbr.nl/beleid-en-regelgeving-ontwikkelen/integraal-afwegingskader-voor-beleid-en-regelgeving/7-wat-zijn-de-gevolgen/77-evalueren-en-monitoren-van-beleid/ex-post-evaluatie>

Kenniscentrum voor beleid en regelgeving. (2022, 29 april). Uitvoerbaarheid. Geraadpleegd van: <https://www.kcbr.nl/beleid-en-regelgeving-ontwikkelen/integraal-afwegingskader-voor-beleid-en-regelgeving/6-wat-het-beste-instrument/65-uitvoerbaarheid>

Keultjes, H., & van Ast, M. (2021, 2 april). Rutte na motie van afkeuring door als premier, gaat 'hard werken' aan terugwinnen vertrouwen. *Het Parool*. Geraadpleegd van: <https://www.parool.nl/nederland/rutte-na-motie-van-afkeuring-door-als-premier-gaat-hard-werken-aan-terugwinnen-vertrouwen~ba0d3fdc/>

Kingdon, J. W. (2002). *Agendas, alternatives, and public policies* (2nd [upd.], Ser. Longman classics in political science). Longman.



- Korsten, A.F.A. (1983). *Wat is goed genoeg? Benutting van onderzoek naar overheidsbeleid*. Amsterdam: uitgeverij Kobra.
- Korsten, A.F.A. (2013a). *Evaluaties benutten. Een studie naar benuttingsmogelijkheden van evaluatiestudies*. Geraadpleegd van: <https://arnokorsten.nl/PDF/Adviezen%20algemeen/Evaluaties%20benutten.pdf>
- Korsten, A.F.A. (2013b). *Wat ervan terecht komt: zicht op beleidsevaluatie*. Geraadpleegd van: <https://arnokorsten.nl/PDF/Onderzoek/ABC%20van%20beleidsevaluatie.pdf>
- Leeuw, F. & van Gils, G. (2010). Over benutting van beleidsevaluaties bij het ministerie van Buitenlandse Zaken. In Aardema, H., Derksen, W., Herweijer, M., de Jong, P. (Eds.), *Meerwaarde van de bestuurskunde* (pp. 165-174). Den Haag: Boom Lemma uitgevers.
- Pattyn, V. (2020). *Beleidsevaluatie*. In Wayenberg, E. (EDS), *Bouwen aan beleid. Het proces van de overheid* (pp.182-204). Gent: Owl press.
- Polder, K. (1996). *Implementatie van wetgeving in onderwijs: Beleidsevaluatie-onderzoek in basisonderwijs en volwasseneneducatie*. Leuven-Apeldoorn: Garant.
- Rutte, M. (2021, 15 januari). *Kamerbrief met reactie kabinet op rapport Ongekend onrecht [Kamerbrief]*. Geraadpleegd van: <https://www.rijksoverheid.nl/documenten/kamerstukken/2021/01/15/kamerbrief-met-reactie-kabinet-op-rapport-ongekend-onrecht>
- Smet, V. (2013). Evidence-based policy en de doorwerking van sociaalwetenschappelijk onderzoek: are we there yet?. *Beleidsonderzoek Online*, 12.
- Teisman, G. (1993). De reconstructie van beleidsprocessen: over fasen, stromen en ronden. Van Heffen, O., van Twist, M.J.W. (EDS), *Beleid en wetenschap* (18-32). Alphen aan den Rijn.
- Tweede Kamer. (2020a). *Ongekend onrecht. Verslag – Verslag parlementaire ondervragingscommissie kinderopvangtoeslag*. Tweede Kamer der Staten-Generaal. Geraadpleegd van: [https://www.tweedekamer.nl/sites/default/files/atoms/files/20201217\\_eindverslag\\_parlementaire\\_ondervragingscommissie\\_kinderopvangtoeslag.pdf](https://www.tweedekamer.nl/sites/default/files/atoms/files/20201217_eindverslag_parlementaire_ondervragingscommissie_kinderopvangtoeslag.pdf)
- Tweede Kamer. (2020b). *Brief van het Presidium over een onderzoek naar de oorzaken van problemen bij uitvoeringsorganisaties en instelling van een tijdelijke commissie parlementair onderzoek uitvoeringsorganisaties*. Tweede Kamer der Staten-Generaal. Geraadpleegd van: <https://www.tweedekamer.nl/kamerstukken/detail?id=2020D04621&did=2020D04621>
- Tweede Kamer. (2021). *Klem Tussen Balie en Beleid. Rapport: Tijdelijke commissie Uitvoeringsorganisaties*. Tweede Kamer der Staten-Generaal. Geraadpleegd van: [https://www.tweedekamer.nl/sites/default/files/atoms/files/20210225\\_eindrapport\\_tijdelijke\\_commissie\\_uitvoeringsorganisaties.pdf](https://www.tweedekamer.nl/sites/default/files/atoms/files/20210225_eindrapport_tijdelijke_commissie_uitvoeringsorganisaties.pdf)
- Van der Berg, C.F., van der Steen, M.A., Tummers, L.G. (2021). *Openbaar bestuur. Beleid, organisatie en politiek*. Deventer: Wolters Kluwer.

- Van der Vlist, A.J., Bunte, F.H.J., van Galen, M.A. (2007). *Beleidsevaluatie ex-post Methodiek en illustratie*. Den Haag: LEI.
- Van Hattem, J., & van der Horst, M. (2022). *Innovatieve beleidsevaluatie. Een beeld van de mogelijkheden voor innovatie in beleidsevaluatiemethoden bij mobiliteitsbeleid*. Kennisinstituut voor Mobiliteitsbeleid.
- VVD, D66, CDA & Christenunie. (2021). *Omzien naar elkaar, vooruitkijken naar de toekomst – Coalitieakkoord 2021-2025*. Geraadpleegd van: [Coalitieakkoord 'Omzien naar elkaar, vooruitkijken naar de toekomst' | Publicatie | Kabinetsformatie \(kabinetsformatie2021.nl\)](#)
- Weerwind, F.M. (2022-a, 21 april). *Kamerbrief Inkleuring en stand van zaken invoeringstoets* [Kamerbrief]. Geraadpleegd van: <https://www.rijksoverheid.nl/documenten/rapporten/2022/04/21/ek-bijlage-tk-inkleuring-en-stand-van-zaken-invoeringstoets>
- Weerwind, F.M. (2022-b, 13 september). *Reactie op aanvullende verzoeken inzake de invoeringstoets* [Kamerbrief]. Geraadpleegd van: <https://open.overheid.nl/documenten/ronl-219219d9e808955d5174c84e6b88de350dec8187/pdf>
- Weiss, C. H. (1986). The many meanings of research utilization. In *Social science and social policy* (pp. 31-40). Routledge.
- Willink, T. (2017). *Over de uitvoerbaarheid en uitvoering van nieuw beleid (regeerakkoord)* [bijlage *Eindverslag Informatie Tjeenk Willink*]. Den Haag: Tweede Kamer.
- Willink, T. (2021). *Eindverslag informatie Tjeenk Willink*. Geraadpleegd van: <https://www.kabinetsformatie2021.nl/documenten/verslagen/2021/04/30/eindverslag-informateur-tjeenk-willink>.
- Winter, H.B., Scheltema, M. & Herweijer, M. (1990). *Evaluatie van wetgeving*. Deventer: Kluwer.

# Bijlagen

## Bijlage 1: Factoren van invloed op evaluatiegebruik (Johnson et al., 2009)

+ = positief verband met evaluatiegebruik  
 - = negatief verband met evaluatiegebruik  
 X = geen verband met evaluatiegebruik

Categorie	Eigenschap	Beschrijving van de eigenschap	Frequentie	Relatie met evaluatiegebruik	Verspreiding met evaluatiegebruik
Evaluatie Implementatie	Communicatie kwaliteit	Duidelijkheid en frequentie van rapportage resultaten, belangenbehartiging van de resultaten door beoordelaar, verspreidingsgebied. Bevat ook het type aanbevelingen in het rapport en het proces van communicatie tussen beoordelaars en klanten	11	Vaak een van de meest belangrijke elementen met betrekking tot evaluatiegebruik. Gedetailleerd, uitvoerbaar, <i>evidence-based</i> aanbevelingen verhogen gebruik. In twee studies werd daarnaast geen relatie met gebruik gevonden	9+ 2X
	Tijdigheid	Timing van de evaluatie in grotere context; tijdigheid van rapportage wanneer de evaluatie is voltooid; timing van verspreiding naar besluitmakers	7	Voornamelijk een positieve relatie tussen timing en evaluatiegebruik gevonden. Eén studie concludeerde dat tijdigheid niet belangrijk was in het bepalen van gebruik	6+ 1-
	Evaluator competentie	Persoonlijke eigenschappen van de beoordelaar buiten het evaluatieproces, niveau van culturele competentie, leiderschapsstijl van beoordelaar	6	De meeste studies suggereren dat competentie van de beoordelaar niet belangrijk is voor evaluatiegebruik	6-
	Evaluatie kwaliteit	Eigenschappen van het evaluatieproces, verfijning van methoden, striktheid, type evaluatiemodel	6	Sommige studies vonden een link tussen kwaliteit en gebruik, hoewel minder belangrijk dan advies en communicatie. Eén studie vond geen relatie tussen kwaliteit en gebruik	5+ 1X
	Bevindingen	Aard van bevindingen (bv. positief of negatief), mate van congruentie met de verwachtingen van het publiek, waarde van bevindingen voor besluitvorming	6	Gemengde conclusies. In twee onderzoeken waren bevindingen belangrijk voor het gebruik, hoewel minder dan communicatie, tijdigheid en evaluatiekwaliteit	2+ 4-

	Relevantie	Mate waarin de informatie uit de evaluatie relevant is voor de ontvanger, en de organisatorische locatie van de beoordelaar	6	Gemengde conclusies. Twee studies vonden relevantie niet belangrijk voor gebruik, maar twee studies kwamen tot sterkere relaties tussen informatierelevantie en gebruik	2- 2+
	Geloofwaardigheid	De objectiviteit, geloofwaardigheid en geschiktheid van het evaluatieproces en/of van de activiteiten van de beoordelaar	4	Gemengde conclusies. Twee studies vonden een sterke relatie met evaluatiegebruik; twee studies vonden geen dergelijke relatie	2+ 2-
Beslissing s- of beleids- omgeving	Persoonlijke eigenschappen	Eigenschappen van de evaluatiegebruiker, bijvoorbeeld organisatorische rol van de besluitvormer, informatieverwerkingsstijl, sociale eigenschappen etc.	9	Verschillen in leerstijlen, functies, managementlaag en ervaringsniveau hebben invloed op het gebruik van evaluaties	9+
	Betrokkenheid en/of ontvankelijkheid naar evaluatie	Houding ten opzichte van de evaluatie, en toewijding voor het uitvoeren van de evaluatie door de gebruikers; de mate waarin de organisatie weerstand biedt tegen de evaluatie; de mate waarin stakeholders open staan voor de evaluatie	8	Sommige studies concludeerden dat commitment, actieve organisatie-inspanningen, en (financiële) ondersteuners gebruik versterken. Eén studie concludeerde dat de houding ten opzichte van evaluatie geen invloed heeft op het gebruik ervan	7+ 1-
	Politieke klimaat	De politieke oriëntatie van de mensen die opdracht gaven tot de evaluatie, de mate waarin de besluitvormer afhankelijk is van externe sponsors, interne rivaliteit, strijd om budgetten en machtsstrijd	6	Over het algemeen werd gevonden dat aandacht voor het politieke klimaat een positief effect heeft op gebruik	6+
	Besluit eigenschappen	Het belang van het besluit of het geëvalueerde probleem, het type te maken besluit, de mate waarin het programmagebied nog nieuw is	5	Elk van de vijf onderzoeken rapporteerden verbanden tussen de kenmerken van de beslissing en evaluatiegebruik	5+
	Concurrerende informatie	Informatie met betrekking tot het evaluatieonderwerp die beschikbaar is voor stakeholders van buiten het evaluatieproces, dat wil zeggen via persoonlijke observatie, die concurreert met evaluatiegegevens	3	Tegenstrijdige bevindingen. Eén onderzoek wees uit dat een grote hoeveelheid van concurrerende informatie geen invloed heeft op instrumenteel gebruik, terwijl een ander onderzoek ontdekte dat hoge beleidsambtenaren alleen de evaluatieresultaten gebruikten die werden ondersteund door andere informatiebronnen	1- 1+

	Informatie-behoefte van de evaluatie-doelgroep	Informatiebehoefte van de evaluatiedoelgroep, de soorten informatie, het aantal doelgroepen met verschillende informatiebehoefte, tijdsdruk en waargenomen behoefte aan evaluatie	2	Uit beide onderzoeken bleek dat het in acht nemen van de informatie- behoeften van de evaluatie- doelgroep een positieve invloed had op het gebruik van evaluatieresultaten	2+
Stakehold erbetrokke nheid	Betrokkenheid bij commit-ment of ontvanke-lijkheid voor evaluatie	Het betrekken van stakeholders bij de evaluatie creëert een commitment of ontvankelijkheid voor evaluatie	14	Voor het grootste deel, had commitment versterkt door betrokkenheid bij de evaluatie een positieve invloed op het gebruik. In één studie was het betrekken van een toegewijde uitvoeringsambtenaar zelfs essentieel voor het implementeren van de evaluatieresultaten	14+
	Betrokkenheid bij commu-nicatiekwaliteit	Betrokkenheid van belanghebbenden bevordert verbeterde communicatie	5	Alle vijf onderzoeken identificeerden manieren waarbij stakeholderbetrokkenheid leidde tot meer gebruik	5+
	Directe betrokkenheid stakeholders	De directe relatie tussen betrokkenheid en evaluatie gebruik	4	Alle onderzoeken rapporteerden een positieve invloed van betrokkenheid op verschillende soorten gebruik	4+
	Betrokkenheid bij geloof-waardigheid	Betrokkenheid van stakeholders leidde tot verhoogde geloofwaardigheid van het evaluatieproces en/of de beoordelaar	4	Drie van de vier geobserveerde onderzoeken constateerden een sterke relatie met gebruik	3+ 1x
	Betrokkenheid bij bevin-dingen	Het betrekken van stakeholders in het kennen en begrijpen van de evaluatieresultaten	4	Drie studies benadrukten dat betrokkenheid bij de bevindingen belangrijk is voor evaluatiegebruik	3+ 1x
	Betrokkenheid bij relevantie	Stakeholderparticipatie om belangrijke organisatorische zorgen te betrekken in het evaluatieontwerp	4	Beter contact met stakeholders bevorderde grotendeels de relevantie waardoor de resultaten meer werden gebruikt	4+
	Betrokkenheid bij persoonlijke	Betrokkenheid van evaluatiestakeholders bij verschillende organisatieniveaus	2	Bevindingen van één onderzoek suggereren dat betrokkenheid van managers het gebruik meer beïnvloedt	1+ 1x

	eigenschappen			dan het betrekken van ander personeel	
	Betrokkenheid bij Eigenschappen besluit	Het betrekken van een scala aan belanghebbenden in verschillende omgevingen afhankelijk van de kenmerken van de te maken beslissing	1	Deze studie constateerde een positieve relatie tussen evaluatiegebruik en betrokkenheid van individuen in niet-traditionele bureaucratieën waar de besluitvorming input van mensen op alle niveaus in de organisatie omvat	1+
	Betrokkenheid bij informatie-behoeften	De betrokkenheid van stakeholders faciliteert de introductie van hun informatiebehoeften	1	De wens van stakeholders voor informatie en de tijdigheid van de evaluatie bevordert eigenaarschap van informatie, wat weer een positieve relatie heeft met gebruik	1+

Bijlage 2: Kritieke succesfactoren voor organisatie en proces van doorwerking (Bekkers et al., 2003, pp. 51-53)

<b>Aangrijpingspunt doorwerking</b>	<b>Factoren die doorwerking beïnvloeden</b>
<i>Kenmerken adviesproduct</i>	Kenmerken beleidsprobleem
	Aansluiten bij kenmerken beleidsprobleem
	Leesbaarheid rapport
	Vormgeving rapport
	Kwaliteit rapport
	Type aanbevelingen
	Bijdrage aan ontkokering
	Verhouding aanbevelingen tot huidige beleid
<i>Behoeften ontvangers</i>	Flexibiliteit adviescapaciteit
	Afstemming over werkprogramma
	Organisatiecultuur ontvanger
	Niveau van departementale contacten
<i>Organisatie adviesraad</i>	Reputatie adviesraad
	Samenstelling adviesraad
	Samenstelling secretariaat
	Werkwijze secretariaat
	Betrokkenheid en rol leden bij projecten
	Allianties met andere raden of kennisinstituten
	Betrokkenheid van derden in adviesproces
<i>Interactie met doelgroep</i>	Afstemming over werkprogramma
	Afstemming tijdens voorbereiding adviestraject
	Afstemming tijdens uitvoering adviestraject

	Structurele contacten met doelgroep
<i>Presentatie en nazorg</i>	Aanwezigheid communicatiestrategie
	Ondersteunende adviesproducten
	Organisatie presentatie
	Actieve opstelling richting media
	Organisatie van nazorg met betrekking tot formele afhandeling
	Organisatie van 'lange termijn' nazorg
<i>Institutionele kenmerken</i>	Timing advies
	Mate van (beleids- en kennis)concurrentie
	Mate van corporatisme
	Professionalisering veld
	Pluriformiteit/fragmentatie veld
	Politieke belangstelling



*Bijlage 3: Topiclijst interviews over invoeringstoetsen*

Opening:	<ul style="list-style-type: none"> <li>• Praktische info over opname en dataverwerking</li> <li>• Voorstellen</li> <li>• Korte introductie scriptieonderwerp</li> </ul>
1. De respondent(en) zelf	<ul style="list-style-type: none"> <li>• Zou je jezelf willen voorstellen?</li> <li>• Hoelang ben je al werkzaam in de organisatie?</li> <li>• Zou je wat meer kunnen vertellen over je functie?</li> <li>• Hoe ben je betrokken bij de invoeringstoets?</li> </ul>
2. De wet/regeling	<ul style="list-style-type: none"> <li>• Kan je wat vertellen over de geëvalueerde wet?</li> </ul>
3. De invoeringstoets in het algemeen. (rationeel, organisch, politiek)	<ul style="list-style-type: none"> <li>• Wat is volgens jou de invoeringstoets?</li> <li>• Kan je omschrijven hoe het instrument precies werkt?</li> <li>• Hoe is het instrument tot nu toe bij jouw organisatie uitgewerkt?</li> <li>• Hoe worden gemeenten betrokken bij zo'n invoeringstoets over de leefgeldregeling?</li> <li>• Hoe worden de resultaten van de invoeringstoets verwerkt binnen jouw organisatie?</li> <li>• Hoe worden deze resultaten benut? <ul style="list-style-type: none"> <li>○ Instrumenteel? (directe doorwerking – implicaties worden overgenomen)</li> <li>○ Conceptueel? (verandering in kennis of denkwijze)</li> <li>○ Agenderend? (aandacht trekkend)</li> <li>○ Politiek/strategisch (resultaten worden op een politieke of strategische manier gebruikt)</li> </ul> </li> </ul>
4. Complicaties bij de Invoeringstoets (rationeel, organisch, politiek)	<ul style="list-style-type: none"> <li>• Welke complicaties doen zich tot nu toe voor bij de invoeringstoets?</li> <li>• Hoe worden deze veroorzaakt? <ul style="list-style-type: none"> <li>○ Niet zoals beoogd (rationeel)</li> <li>○ Omstandigheden ongunstig (organisch)</li> <li>○ Bewuste beïnvloeding (politiek)</li> </ul> </li> <li>• Wie is (mede)veroorzaker van deze complicaties?</li> <li>• Zou je het karakter van deze complicaties omschrijven als rationeel, organisch of politiek? Waarom?</li> </ul>
5. Gevolgen van deze complicaties (rationeel, organisch, politiek)	<ul style="list-style-type: none"> <li>• Wat is het gevolg van deze complicaties?</li> <li>• Hoe hebben deze complicaties invloed op de benutting van de resultaten van de invoeringstoets?</li> </ul>

6. Leerpunten voor de invoeringstoets	<ul style="list-style-type: none"> <li>• Hoe zou je de invoeringstoets (anders) willen inrichten om de genoemde complicaties in de toekomst te voorkomen?</li> <li>• Hoe zou je de invoeringstoets überhaupt inrichten? (los van de genoemde complicaties)</li> <li>• Hoe zou je ervoor willen zorgen dat de benutting van de invoeringstoets wordt geborgd?</li> <li>• Welke factoren zijn voornamelijk van belang voor het benutten van een invoeringstoets?</li> </ul>
7. Wat zou je willen meegeven?	<ul style="list-style-type: none"> <li>• Zijn er nog dingen die je zou willen meegeven wat betreft de invoeringstoets?</li> <li>• Zijn er nog dingen die je zou willen meegeven wat betreft de benutting van de invoeringstoets?</li> <li>• Zijn er nog overige dingen je zou willen meegeven?</li> </ul>
Afsluiting	Heel erg bedankt voor de tijd en moeite. Zijn er misschien nog andere mensen of casussen die interessant zouden zijn voor dit onderzoek?

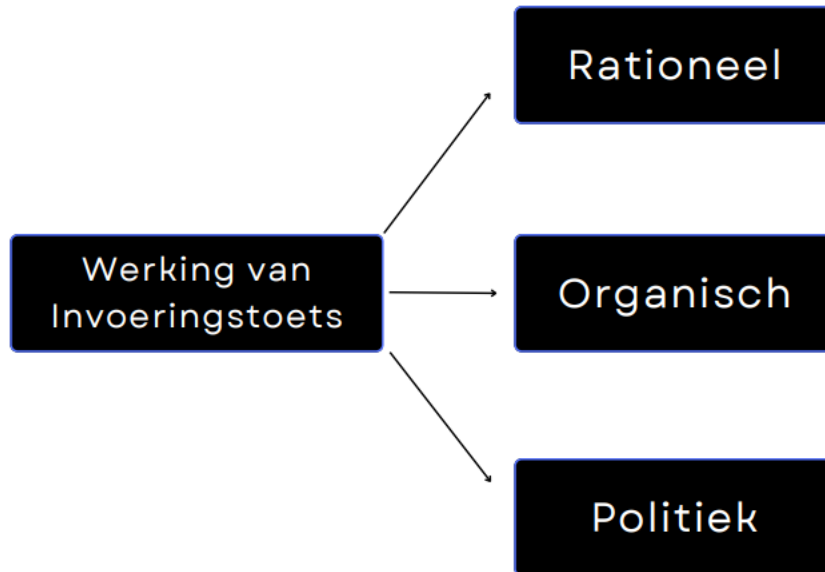
*Bijlage 4: Topiclijst interviews over evaluaties met gelijkenissen met de invoeringstoets*

Opening:	<ul style="list-style-type: none"> <li>• Praktische info over opname en dataverwerking</li> <li>• Voorstellen</li> <li>• Korte introductie scriptieonderwerp</li> </ul>
8. De respondent(en) zelf	<ul style="list-style-type: none"> <li>• Zou je jezelf willen voorstellen?</li> <li>• Hoelang ben je al werkzaam in de organisatie?</li> <li>• Zou je wat meer kunnen vertellen over je functie?</li> </ul>
9. De evaluatie	<ul style="list-style-type: none"> <li>• Zou je wat meer kunnen vertellen over de wet/regeling ...</li> <li>• Zou je wat meer kunnen vertellen over de evaluatie hiervan? <ul style="list-style-type: none"> <li>○ Welke methoden zijn toegepast?</li> <li>○ Door wie is de evaluatie uitgevoerd?</li> <li>○ Wie waren betrokken bij de evaluatie?</li> <li>○ Wat was het resultaat van de evaluatie?</li> <li>○ Aan wie zijn de resultaten aangeboden?</li> <li>○ Wat is er met de resultaten gebeurd?</li> </ul> </li> </ul>
10. Complicaties bij de evaluatie (rationeel, organisch, politiek)	<ul style="list-style-type: none"> <li>• Welke complicaties doen zich voor bij de evaluatie?</li> <li>• Hoe worden deze veroorzaakt?</li> <li>• Wie is (mede)veroorzaker van deze complicaties?</li> <li>• Zou je het karakter van deze complicaties omschrijven als rationeel, organisch of politiek? Waarom?</li> </ul>
11. Gevolgen van deze complicaties (rationeel, organisch, politiek)	<ul style="list-style-type: none"> <li>• Wat is het gevolg van deze complicaties?</li> <li>• Hoe hebben deze complicaties invloed op de benutting van de resultaten van de evaluatie?</li> </ul>

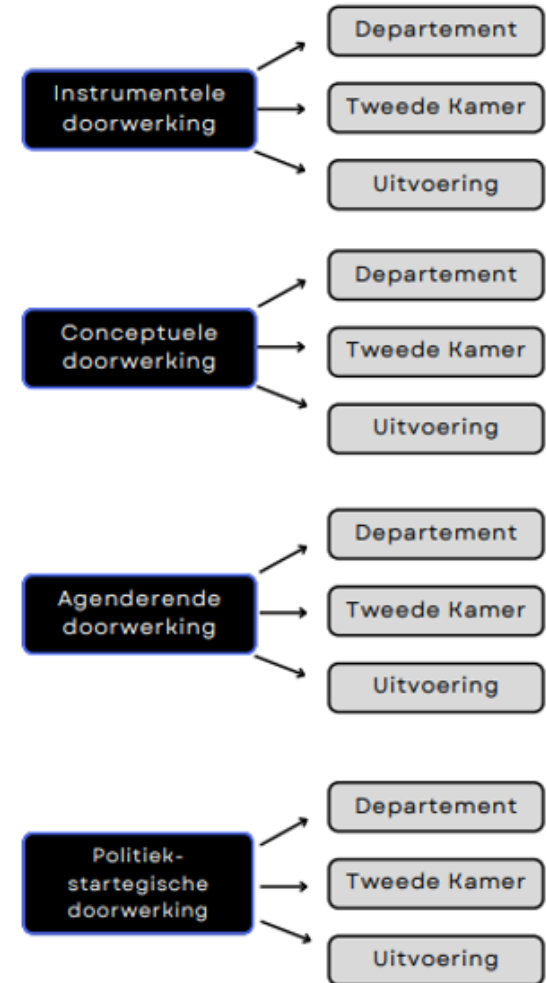
<p>12. De invoeringstoets in het algemeen. (rationeel, organisch, politiek)</p>	<ul style="list-style-type: none"> <li>• In hoeverre ben je bekend met de invoeringstoets? (niet? Dan uitleg geven en door naar volgende blok.)</li> <li>• Wat is volgens jou de invoeringstoets?</li> <li>• Kan je omschrijven hoe het instrument precies werkt?</li> <li>• Hoe is het instrument tot nu toe bij jouw organisatie uitgewerkt?</li> <li>• Hoe worden de resultaten van de invoeringstoets verwerkt binnen jouw organisatie?</li> <li>• Hoe worden deze resultaten benut?</li> </ul>
<p>13. Leerpunten voor de invoeringstoets</p>	<ul style="list-style-type: none"> <li>• Hoe zou je de invoeringstoets willen inrichten?</li> <li>• Hoe zou je de invoeringstoets (anders) willen inrichten om de genoemde complicaties van de evaluatie van X te voorkomen?</li> <li>• Hoe zou je ervoor willen zorgen dat de benutting van de invoeringstoets wordt geborgd?</li> </ul>
<p>14. Wat zou je willen meegeven?</p>	<ul style="list-style-type: none"> <li>• Zijn er nog dingen die je zou willen meegeven wat betreft de invoeringstoets?</li> <li>• Zijn er nog dingen die je zou willen meegeven wat betreft de benutting van de invoeringstoets?</li> <li>• Zijn er nog overige dingen je zou willen meegeven?</li> </ul>
<p>Afsluiting</p>	<p>Heel erg bedankt voor de tijd en moeite. Zijn er misschien nog andere mensen of casussen die interessant zouden zijn voor dit onderzoek?</p>

Bijlage 5: Codeboom per deelvraag:

Deelvraag 4) Hoe past de invoeringstoets binnen de beleidscyclus van de Rijksoverheid?



Deelvraag 5: Hoe en waar wordt de invoeringstoets benut binnen de Rijksoverheid?



Deelvraag 6: *In hoeverre en hoe wordt in de praktijk voldaan aan de randvoorwaarden voor het borgen van de benutting van de invoeringstoets?*

