

Rapportage Minimumniveau Toezicht Belastingdienst

2024-G4M2/5.1.2e |fv/gf

Disclaimer

Oktober 2024

Dit document is geschreven onder verantwoordelijkheid van PricewaterhouseCoopers Advisory N.V. met de Belastingdienst als opdrachtgever. Het betreft geen document opgesteld door accountants, belastingadviseurs of daaraan gelijk te stellen dienstverlening. Wij hebben ten aanzien van de aan ons aangereikte informatie (zowel schriftelijk als mondeling) geen juridische of fiscaal-inhoudelijke beoordeling uitgevoerd. Voor zover wij in dit rapport verwijzen naar relevante wet- en regelgeving kan dit niet beschouwd worden als het verstrekken van een juridische opinie of advies. U kunt contact over dit onderzoek opnemen via 5.1.2e@pwc.com. U blijft te allen tijde zelf volledig verantwoordelijk voor eventuele op het document gebaseerde besluitvorming en/of beslissing(en). PwC aanvaardt geen aansprakelijkheid (ook niet voor nalatigheid) voor de gevolgen van enig handelen of nalaten door u en/of derden op basis van (de inhoud van) het document, en wijst iedere verantwoordelijkheid, zorgplicht en/of aansprakelijkheid - contractueel, op basis van onrechtmatige daad (inclusief nalatigheid) of anderszins - af voor enig besluit en/of enige beslissing waaraan (de inhoud van) het document ten grondslag ligt. Wij stellen dit document uitsluitend op voor de Belastingdienst als opdrachtgever, in overeenstemming met de opdrachtbevestiging. In het geval u een verzoek ontvangt op grond van de Wet Open Overheid (hierna: WOO-verzoek) ter zake van schriftelijke uitingen van PwC, verzoeken wij u ons hierover onverwijld (in ieder geval voorafgaand aan de te nemen beslissing op het WOO-verzoek en daarom voorafgaand aan eventuele openbaarmaking) schriftelijk te informeren en in staat te stellen onze zienswijze te geven op het WOO-verzoek. Dit document alsmede enig geschil voortvloeiende uit of verband houdend met (de inhoud van) het document worden uitsluitend beheerst door Nederlands recht.

Managementsamenvatting

In 2021 wees intern onderzoek van de Belastingdienst uit dat expliciete niveaus van fiscaal toezicht nergens zijn bepaald en dat de bekostiging geheel of deels bepalend is voor het niveau van toezicht. Professor Speklé, als externe deskundige betrokken bij dat onderzoek, wees erop dat het bij het bepalen van een gewenst toezichtsniveau gaat om een 'intrinsiek politieke vraag', die niet intern binnen de Belastingdienst beantwoord kan worden maar 'onvermijdelijk een normatieve stellingname impliceert'.

De bepaling van een minimumniveau toezicht geniet onverminderd de belangstelling van de toezichthouders, de Auditdienst Rijk en de Algemene Rekenkamer. Daarom heeft de Belastingdienst eind 2023 aan PwC gevraagd om onderzoek uit te voeren ter beantwoording van de vraag: *"Is het mogelijk om een beheerst minimumniveau toezicht te bepalen ten behoeve van de handhavingstaken van de Belastingdienst en welke elementen daarbij een rol spelen, waarbij rekening wordt gehouden met acceptabele restrisico's."*

Ons onderzoek laat enerzijds zien dat de bepaling van een *absoluut* minimumniveau toezicht niet mogelijk is. Naast door de in 2021 genoemde redenen, komt dit ook doordat een minimumniveau van toezicht niet los gezien kan worden van keuzes in de uitvoering en andere vormen van handhaving dan toezicht.

Anderzijds toont ons onderzoek aan dat een *acceptabel* minimumniveau van toezicht wel mogelijk is. Het bestaansrecht van het adjectief *acceptabel* in deze interpretatie van de onderzoeksvraag is driedelig.

Ten eerste doet *acceptabel* recht aan het feit dat perfect toezicht per definitie onmogelijk is vanwege beperkingen in toezichtsinstrumenten en beschikbare data, onder andere vanwege privacywetgeving. De zinsnede *"waarbij rekening gehouden wordt met acceptabele restrisico's"* in de onderzoeksvraag verwijst hier ook naar: perfectie is niet nodig zolang de imperfectie beperkt is tot een acceptabel niveau. Ten tweede verlegt deze invalshoek de aandacht naar *acceptabel*, gegeven de keuzes in de uitvoering en andere vormen van handhaving dan toezicht. Ten derde verwijst *acceptabel* naar *politiek aanvaardbaar* en adresseert daarmee de opmerking van Professor Speklé dat een minimumniveau van toezicht ook deels een politieke keuze is.

De conclusie dat het mogelijk is om een acceptabel minimumniveau van toezicht vast te stellen roept de vraag op *hoe* dit vastgesteld kan worden. Dit rapport schetst een afweegkader om op basis van ex ante onderzoek en dialoog uit een scala aan mogelijke vormen van uitvoering en handhaving de vorm te selecteren met het grootste maatschappelijk draagvlak. Onder de premisse dat het beschouwde scala aan mogelijkheden voldoende groot is en bij benadering alle mogelijke vormen omvat, voldoet de geselecteerde vorm aan een acceptabel minimumniveau van toezicht. Daarmee is dat minimumniveau dus bepaald en *beheerst* in de zin dat de effecten en risico's vooraf onderzocht zijn. Zo'n afweegkader heeft zich op andere beleidsterreinen als de energietransitie en de accommodatie van vluchtelingen al bewezen.

Ons onderzoek laat zien dat sommige onderdelen binnen de Belastingdienst al volgens dit principe werken: voordat een toezichtsvorm wordt vastgesteld, is deze vorm al vergeleken met alternatieve vormen van toezicht en wenselijker bevonden dan die alternatieven. Echter, deze vergelijking vindt nu voornamelijk intern plaats.

Uit ons onderzoek blijkt ook dat nu nog niet voldaan wordt aan belangrijke randvoorwaarden om met het voorgestelde afweegkader te gaan werken. Ten eerste is meer data en onderzoek nodig om de effecten van verschillende vormen van uitvoering en handhaving te kunnen bepalen. Ten tweede is er behoefte aan kanalen via welke de dialoog met de maatschappij over de wenselijkheid van effecten van fiscaal toezichtsbeleid gevoerd kan worden. Tot slot is capaciteit nodig om met dit afweegkader te gaan werken.

Inhoudsopgave

Managementsamenvatting	1
Inhoudsopgave	2
Overzicht van figuren	4
Overzicht van tabellen	5
1. Achtergrond en aanleiding	6
1.1. Doelstelling van het onderzoek	6
1.2. Definities	7
1.3. Leeswijzer	8
2. Absoluut minimumniveau toezicht	9
2.1. Conceptueel model identificeren restrisico	9
2.1.1. Compliance: nalevingsbereidheid en robuustheid	9
2.1.2. Selectie voor onderzoek	10
2.1.3. Uitvoering onderzoek	10
2.1.4. Totale model	11
2.2. Onmogelijkheid bepalen absoluut minimumniveau toezicht	11
3. Acceptabel minimumniveau toezicht	13
3.1. Mogelijkheid acceptabel minimumniveau toezicht	13
3.2. Raamwerk ter bepaling acceptabel minimumniveau toezicht	15
3.2.1. Fase 1: Dialoog met belanghebbenden en experts uit verschillende disciplines	15
3.2.2. Fase 2: Bureau- en veldonderzoek	15
3.2.3. Fase 3: Afweging van strategische prioriteiten en bepaling risicobereidheid	16
3.2.4. Illustratie aan de hand van een sterk versimpeld voorbeeld	16
4. Toepassing afwegingskader	20
4.1. Fase 1: Dialoog met belanghebbenden en experts uit verschillende disciplines	20
4.1.1. Stap 1: Bepaling scope	20
4.1.2. Stap 2: Bepaling relevante effecten van toezicht	20
4.1.3. Stap 3: Inventarisatie van mogelijke uitvoerings- en handhavingsmixen	21
4.1.4. Stap 4: Inventarisatie van mogelijke scenario's	22
4.2. Fase 2: Bureau- en veldonderzoek	23
4.2.1. Stap 5: Ontwikkeling van simulatiemodel	23
4.3. Fase 3: Afweging van strategische prioriteiten en bepaling van risicobereidheid	24

4.3.1.	Stap 6: Afweging van strategische prioriteiten	24
4.3.2.	Stap 7: Bepaling van risicobereidheid	24
4.4.	Illustratie van toepassing binnen de horeca	25
4.4.1.	Stap 1: Bepaling scope	25
4.4.2.	Stap 2: Bepaling relevante effecten van toezicht	25
4.4.3.	Stap 3: Inventarisatie mogelijke uitvoerings- en handavingsmixen	25
4.4.4.	Stap 4: Inventarisatie van mogelijke scenario's	26
4.4.5.	Stap 5: Ontwikkeling van het simulatiemodel	27
4.4.6.	Stap 6: Afweging van strategische prioriteiten	27
4.4.7.	Stap 7: Bepaling van risicobereidheid	27
5.	Uitdagingen	28
5.1.	Fase 1: Dialoog met belanghebbenden en experts uit verschillende disciplines	28
5.2.	Fase 2: Bureau- en veldonderzoek	28
5.3.	Fase 3: Afweging van belangen en bepaling risicobereidheid	29
5.4.	Uitdagingen die niet fasespecifiek zijn	29
6.	Vervolgstappen	31
6.1.	Pilot	31
6.2.	Integratie van de twee raamwerken	31
6.3.	Initiëren van de verandering	32
6.4.	Samenvatting vervolgstappen	33
7.	Conclusies	34
A.	Geraadpleegde bronnen	35
A.1.	Gevoerde gesprekken	35
A.2.	Geraadpleegde documentatie	35
B.	Werkwijzen bij andere organisaties en eerder onderzoek	37
	Rapport analyse niveau toezicht achteraf	37
	Vergelijkbare organisaties Nederland	37
	Buitenlandse belastingdiensten	37
	OECD rapport	38
	Braithwaite	38
	Kirchler	39
	Bomb crater effect	40
	Overeenkomsten	41

Overzicht van figuren

Figuur 1. Verband tussen uitvoering, toezicht en handhaving	7
Figuur 2. Conceptueel model: nalevingsbereidheid en robuustheid.....	9
Figuur 3. Conceptueel model: selectie voor correcte (links) en foutieve aangiften (rechts).....	10
Figuur 4. Conceptueel model: uitvoering onderzoek	11
Figuur 5. Conceptueel model: totaaloverzicht.....	11
Figuur 6. Waardering van vier mogelijke handhavings- en uitvoeringsmixen	19
Figuur 7. Relatie tussen toezicht en effecten daarvan.....	21
Figuur 8. Voorbeeld van compliance model (Braithwaite, 2003)	39
Figuur 9. Kirchler's framework (Kirchler et al., 2008).....	40

Overzicht van tabellen

Tabel 1. Uitwerking fictieve determinanten	17
Tabel 2. Waardering van effecten op één schaal van 0 tot 100	17
Tabel 3. Mogelijke voorbeelden van endogene (stap 3) en exogene (stap 4) parameters	22
Tabel 4. Endogene determinanten horeca.....	25
Tabel 5. Exogene determinanten horeca.....	26

1. Achtergrond en aanleiding

De Algemene Rekenkamer (ARK) heeft geconstateerd¹ dat GO, MKB en Particulieren geen centraal kader hebben op basis waarvan zij kunnen bepalen welk niveau van fiscaal toezicht aanvaardbaar is. Hiermee doelt de ARK op een vorm van toezicht die leidt tot acceptabele resultaten en restrisico's. De ARK stelt zich op het standpunt dat de Belastingdienst een dergelijk minimumniveau van toezicht zou moeten vaststellen als kader voor de uitvoerende directies. In haar onderzoek in 2021 stelt de ARK dat het onderzoek naar het acceptabel niveau van toezicht achteraf nog niet geleid heeft tot een norm waaraan het gerealiseerde niveau van toezicht kan worden getoetst. Deze constatering onderstrepen het belang van een meer structurele aanpak en heldere normering binnen de Belastingdienst om de effectiviteit van toezicht en handhaving te verbeteren.

Dit onderwerp is eerder door de Belastingdienst onderzocht. In 2021 is in het kader van de analyse van het dalend niveau van toezicht (2021, zie bijlage) een uitgebreid onderzoek uitgevoerd naar de wijze waarop andere toezichthouders in binnen- en buitenland een "minimum" niveau van toezicht trachten vast te stellen. Dit onderzoek concludeerde dat expliciete niveaus van toezicht nergens bepaald zijn, waarbij als geïdentificeerde knelpunten de beperkte beschikbaarheid van data en het feit dat de beoogde effecten van toezicht niet expliciet vaststaan en dat deze ook niet altijd gemeten worden. De bij dit onderzoek geraadpleegde externe deskundige, professor Roland Spek, wees erop dat het bij het bepalen van een gewenst niveau gaat om een 'intrinsiek politieke vraag, die weliswaar kan worden ondersteund met goede diagnostische en evaluatieve informatie zoals het rapport die biedt, maar waarbij een uiteindelijke beleidskeuze niet logisch-dwingend kan worden afgeleid uit die informatie en onvermijdelijk een normatieve stellingname impliceert.'

Ondanks deze conclusie heeft het vraagstuk van het minimumniveau toezicht onverminderd de belangstelling van de toezichthouders, de Auditdienst Rijk en de ARK. Ook de politiek blijft onverminderd betrokken bij dit vraagstuk wegens de complexiteit ervan binnen de context van de maatschappij.

De Belastingdienst heeft PwC gevraagd om een onderzoek uit te voeren om duidelijkheid te verkrijgen of en zo ja, hoe, op consistente en effectieve wijze een beheerst minimumniveau van toezicht kan worden bepaald. Waar in het onderzoek uit 2021 de focus lag op toezicht achteraf, is zo'n focus nu niet benoemd. Verderop in dit rapport zal blijken dat toezicht in deze context ook niet los gezien kan worden van andere vormen van handhaving en zelfs uitvoering.

De Belastingdienst meet compliance aan de hand van vier factoren. In dit rapport ligt de focus van PwC op de derde compliancefactor: juist en volledig aangifte doen. De andere drie factoren blijven buiten beschouwing.

1.1. Doelstelling van het onderzoek

Het doel van dit onderzoek is de beantwoording van de volgende onderzoeksvraag:

Is het mogelijk om - gegeven de complexiteit en omvang van fiscale processen, het risicogericht werken en proportionaliteit bij de inzet van toezichtinstrumenten - een acceptabel minimumniveau toezicht te bepalen ten behoeve van de handhavingstaken van de Belastingdienst en welke elementen spelen daarbij een rol spelen, waarbij rekening wordt gehouden met acceptabele restrisico's.

Om de onderzoeksvraag te beantwoorden zijn de volgende uit te voeren acties geformuleerd:

- *Verkenning van of en zo ja, hoe, nationale en internationale toezichthoudende instanties komen tot een beheersbaar minimumniveau van toezicht.*

¹ Algemene Rekenkamer, Verantwoordingsonderzoeken 2020, 2021 en 2022

- *Identificatie van determinanten welke van belang zijn bij het komen tot een beheerst minimumniveau van toezicht.*
- *Identificatie van de belangrijkste uitdagingen en belemmeringen bij het vaststellen van een beheerst minimumniveau van toezicht met acceptabele restrisico's.*

1.2. Definities

In deze rapportage hanteren we onderstaande definities voor de relevante basisbegrippen. Deze definities zijn overgenomen uit uitvoerings- en handhavingsstrategie van de Belastingdienst².

- **Uitvoeren:** Het implementeren van wet- en regelgeving en het inrichten en uitvoeren van (semi-) massale processen om daarmee burgers en bedrijven in staat te stellen om wet- en regelgeving na te leven.
- **Handhaven:** Het 'doen' naleven; omvat het hele spectrum aan activiteiten dat hiervoor beschikbaar is en die onder de hieronder genoemde beleidsinstrumenten vallen. De Belastingdienst duidt dienstverlening, toezicht en opsporing als handhaving in brede zin. Handhaving bestaat uit meerdere instrumenten, die vaak in combinatie worden ingezet om compliance te bevorderen en borgen.
- **Dienstverlening:** Het zodanig ondersteunen van burgers en bedrijven dat zij hun eigen verantwoordelijkheid kunnen nemen om wet- en regelgeving na te leven en hun rechten geldend te maken.
- **Toezicht:** Het verzamelen van informatie over de vraag of een handeling of zaak voldoet aan de daaraan gestelde eisen, het zich daarna vormen van een oordeel daarover en het eventueel naar aanleiding daarvan interveniëren (handhaven in enge zin).
- **Opsporing:** Het doen van onderzoek in verband met strafbare feiten onder gezag van de officier van justitie met als doel het nemen van strafvorderlijke beslissingen.
- **Compliance:** Compliance is de bereidheid van burgers en bedrijven om hun wettelijke verplichtingen ten aanzien van de Belastingdienst na te komen
- **Nalevingsbereidheid:** de bereidheid van belastingplichtigen om een correcte aangifte in te dienen.
- **Robuustheid:** De robuustheid van een proces wordt vertaald in de foutgevoeligheid van aangiften.

Figuur 1 geeft weer hoe deze begrippen samenhangen en tot resultaten van toezicht leiden op hoog niveau (boven) en met een uitsplitsing van handhaven (onder), waarbij we opmerken dat opsporing buiten de scope van dit rapport valt. Resultaten geven het voldoen aan wet- en regelgeving weer.



Figuur 1. Verband tussen uitvoering, toezicht en handhaving

² Belastingdienst, Uitvoering- en handhavingsstrategie 2022

1.3. Leeswijzer

In hoofdstuk 2 wordt toegelicht waar een absoluut minimumniveau toezicht bij de Belastingdienst betrekking op heeft op basis van een conceptueel model. Ook bespreken we waar belastingdiensten en vergelijkbare organisaties tegenaanlopen bij het vaststellen van een acceptabel restrisico. In hoofdstuk 3 lichten we de mogelijkheid toe om tot een acceptabel minimumniveau toezicht te komen en zetten we op hoofdlijnen een raamwerk hiervoor uiteen. In hoofdstuk 4 wordt de toepassing van het raamwerk verder uitgewerkt en geïllustreerd aan de hand van een fictief voorbeeld binnen de context van de Belastingdienst. In hoofdstuk 5 worden de uitdagingen en belemmeringen voor de Belastingdienst bij het toepassen van het raamwerk om tot een acceptabel minimumniveau toezicht te komen uiteengezet. Hoofdstuk 6 bevat mogelijke vervolgstappen. Tot slot worden in hoofdstuk 7 de conclusies van het onderzoek benoemd en antwoord gegeven op de gestelde onderzoeksvragen.

2. Absoluut minimumniveau toezicht

In dit hoofdstuk introduceren we een conceptueel model om te identificeren waar een absoluut minimumniveau toezicht bij de Belastingdienst betrekking op heeft. Daarnaast lichten we toe waar belastingdiensten en vergelijkbare organisaties tegenaanlopen bij het vaststellen van acceptabele restrisico's³.

2.1. Conceptueel model identificeren restrisico

Het conceptuele model identificeert het absoluut minimumniveau toezicht aan de hand van verschillende aspecten. Deze aspecten, namelijk nalevingsbereidheid, robuustheid, selectie en uitvoering van het onderzoek worden in dit hoofdstuk nader belicht. Het conceptueel model omschrijft het verloop van een belastingaangifte en licht toe hoe dit zich verhoudt tot de uitvoerings- en handhavingsmix. Het minimumniveau heeft betrekking op alle handhavingsinstrumenten en wordt dus niet louter bepaald door bijvoorbeeld het aantal boekenonderzoeken.

Onderliggend aan het conceptuele model wordt rekening gehouden met de doelstelling om toezicht in te richten op de meest effectieve en efficiënte manier. Deze doelstelling kan bereikt worden met een mix aan handhavingsinstrumenten, waaronder dienstverlening en toezicht.

2.1.1. Compliance: nalevingsbereidheid en robuustheid

Het eerste deel van het conceptueel model (zoals weergegeven in figuur 2) richt zich op compliance, oftewel de mate waarin belastingplichtigen op vrijwillige basis voldoen aan de wet- en regelgeving. Hierbij onderscheiden wij compliance in twee onderdelen: nalevingsbereidheid en robuustheid. Nalevingsbereidheid omvat de bereidheid van belastingplichtigen om een correcte aangifte in te dienen. In praktijk blijkt dat belastingplichtigen die een correcte aangifte willen doen, mogelijk fouten maken en als gevolg een onbedoeld foutieve aangifte indienen. De kans op zulke fouten hangt af van de robuustheid van het proces. De robuustheid van een proces wordt vertaald in de foutgevoeligheid van aangiften. De Belastingdienst zet momenteel actief in om de robuustheid van het aangifteproces op diverse manieren te vergroten. Dit wordt gedaan door vereenvoudiging van de uitleg over belastingwetgeving, het gebruik van geavanceerde technologieën, zoals automatisering of de voorgevulde aangifte inkomstenbelasting voor particulieren. De Belastingdienst zet ook in op educatie en voorlichting van belastingplichtigen om hen bewust te maken van hun verplichtingen en rechten. Tevens biedt de Belastingdienst ondersteuning aan belastingplichtige bij het doen van de aangifte. Positieve prikkels voor naleving en strenge sancties voor niet-naleving vergroten in de handhavingsmix vergroten de nalevingsbereidheid.



Figuur 2. Conceptueel model: nalevingsbereidheid en robuustheid

³ Braithwaite, V., & Braithwaite, J. (2001). An evolving compliance model for tax enforcement. In Crimes of privilege: Readings in white-collar crime. Oxford University Press.

De Belastingdienst heeft kaders opgesteld in de uitvoerings- en handhavingsstrategie (UHS) door middel van een stappenplan waarmee handhavingsstrategieën geformeerd kunnen worden, waarbij deze strategieën voor specifieke doelgroepen ingezet kunnen worden. Deze stappen zijn uitgewerkt in de Leidraad. Gegeven de heterogeniteit van de doelgroepen en belastingmiddelen is dat ook logisch. De uitvoering- en handhavingsstrategie richt zich in haar strategische doelstelling op de verhoging van de compliance. De nalevingsbereidheid wordt verhoogd door belastingplichtigen te motiveren om compliant te zijn aan de gestelde wet- en regelgeving met betrekking tot het aangifteproces, zowel in het proces als de inhoudelijke invulling. De nalevingsbereidheid wordt eveneens verhoogd door corrigerend op te treden bij bewuste niet-naleving. De Belastingdienst verhoogt robuustheid door onder andere een (geautomatiseerde) omgeving te creëren waarin zaken zoveel mogelijk in één keer goed gaan en fouten zoveel mogelijk voorkomen worden. Het tijdig signaleren wanneer belastingplichtigen vastlopen in processen en daarop acteren maakt daar ook onderdeel van uit⁴.

2.1.2. Selectie voor onderzoek

Het tweede deel van het conceptueel model betreft de selectie van ingediende aangiften voor toezicht achteraf door de Belastingdienst, zoals weergegeven in figuur 3. Op het moment van selectie is de Belastingdienst niet op de hoogte of de aangiften voor controle foutief of correct zijn. Dit heeft tot gevolg dat de Belastingdienst zowel correcte als foutieve aangifte kan controleren voor onderzoek.

De selectie van onderzoeken door de Belastingdienst vindt op verschillende wijzen plaats, zoals de risicogerichte selectie, mede op basis van financiële en fiscale belangen, en steekproefsgewijze controles. Voor de toepassing van het conceptuele model wordt uitgegaan van alwetendheid.



Figuur 3. Conceptueel model: selectie voor correcte (links) en foutieve aangiften (rechts)

2.1.3. Uitvoering onderzoek

Het derde deel van het conceptueel model betreft de vorm van onderzoek op hiervoor geselecteerde aangiften, zoals weergegeven in figuur 4. Wij maken hierbij de aanname dat van de geselecteerde aangiften de correcte aangifte geen gevolgen ondervinden van het onderzoek en derhalve niet gecorrigeerd worden. Wij houden echter wel rekening met de mogelijkheid dat de Belastingdienst een foutieve, onderzochte aangifte mogelijk niet (volledig) als foutief identificeert en ten gevolge hiervan er geen correctie volgt.

Voor de uitvoering van onderzoek op een aangifte van een belastingplichtige maakt de Belastingdienst gebruik van verschillende vormen zoals de kantoortoets, een boekenonderzoek waarbij gebruik gemaakt wordt van een steekproef, en controles met behulp van contra-informatie⁵

⁴ Belastingdienst (2022) Uitvoerings- en Handhavingsstrategie: Uitvoering, dienstverlening, toezicht en opsporing bij de Belastingdienst

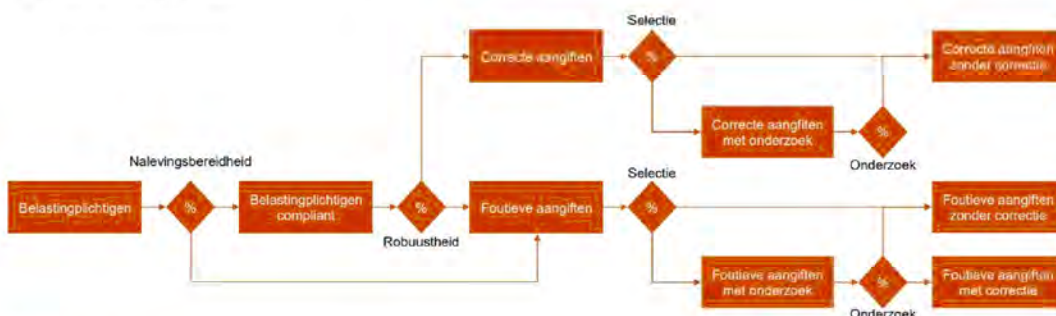
⁵ Belastingdienst (2017) Beleidsdoorlichting toezicht en opsporing en massale processen Belastingdienst



Figuur 4. Conceptueel model: uitvoering onderzoek

2.1.4. Totale model

Samengevoegd geven deze stappen in het conceptuele model weer, zoals weergegeven in figuur 5, op welke manieren de Belastingdienst op hoofdlijnen kan bijdragen om het aantal correcte aangiften te vergroten: door het verhogen van de nalevingsbereidheid en robuustheid, en door foutieve aangiften zo veel mogelijk te corrigeren via de selectie van aangiften voor onderzoek en de uitvoering van het onderzoek zelf.



Figuur 5. Conceptueel model: totaaloverzicht

Het aantal foutieve aangiften zonder correctie omvat het restrisico van de uitvoerings- en handavingsmix. De omvang van dit aantal is afhankelijk van externe invloeden en de strategie van de Belastingdienst. Dit aantal is naar nul te brengen door alle aangiften te selecteren voor verder onderzoek en alle onderzoeken in perfectie uit te voeren. Een perfecte uitvoering van een onderzoek zou bijvoorbeeld inhouden dat steekproeven voor de selectie van te controleren journaalposten geen mogelijkheid meer zijn. Echter, het terugbrengen tot nul zou te veel tijd en geld vragen van de Belastingdienst en van belastingplichtigen. Het gaat niet alleen om doeltreffendheid, maar ook om doelmatigheid. Een absoluut minimumniveau toezicht is daarom van belang om de benodigde inzet van de Belastingdienst te kunnen bepalen.

2.2. Onmogelijkheid bepalen absoluut minimumniveau toezicht

Zoals in de eerder beschreven aanleiding voor dit rapport aangegeven, heeft de Belastingdienst, na raadpleging van een externe expert, geconcludeerd dat het vaststellen van een absoluut minimumniveau van toezicht achteraf niet mogelijk is. Naast dat toezichtniveaus niet expliciet bepaald zijn en dat het bepalen van zo'n minimumniveau een intrinsiek politieke vraag is, blijkt ook uit figuur 1 en 5 dat als er al zo'n minimum niveau bepaald zou kunnen worden, dit niveau afhangt van welke keuzes er gemaakt wordt in de uitvoering en in de handhaving exclusief toezicht.

Op basis van bureaustudie hebben wij onderzocht hoe vergelijkbare organisaties en buitenlandse belastingdiensten hun uitvoerings- en handavingsmixen vormgeven en in hoeverre zij een minimumniveau van toezicht hanteren. In Bijlage B doen wij daar verslag van. In dit hoofdstuk beperken wij ons tot de belangrijkste inzichten daaruit.

Bij zowel vergelijkbare organisaties als buitenlandse belastingdiensten is een trend zichtbaar waarbij de nadruk meer ligt op het bevorderen van compliance, het eerste deel van het conceptueel model (Figuur 2). Dit gebeurt bijvoorbeeld via onlinedienstverlening (zoals in Zweden en Duitsland) en het

vooraf invullen van gegevens (zoals in Portugal). We concluderen dat bij geen van de buitenlandse belastingdiensten of vergelijkbare Nederlandse organisaties de uitvoerings- en handhavingsmix gebaseerd wordt op een absoluut minimumniveau van toezicht, maar dat de beschikbare financiering van de lokale belastingdienst eerder de bepalende factor is voor het niveau van toezicht.

3. Acceptabel minimumniveau toezicht

In voorliggend hoofdstuk wordt toegelicht dat het vaststellen van een absoluut minimumniveau toezicht niet mogelijk is en in plaats daarvan van een acceptabel niveau van minimumniveau toezicht bepaald moet worden. In dit hoofdstuk wordt toegelicht hoe een acceptabel niveau van minimumniveau van toezicht vastgesteld. Ook bespreken we wat we verstaan onder "acceptabel". Daarnaast wordt er een raamwerk geïntroduceerd ter bepaling van een acceptabel minimumniveau toezicht.

3.1. Mogelijkheid acceptabel minimumniveau toezicht

Professor Spekél wees erop dat het bij het bepalen van een gewenst niveau van toezicht gaat om een 'intrinsiek politieke vraag', die niet intern binnen de Belastingdienst beantwoord kan worden maar 'onvermijdelijk een normatieve stellingname impliceert'. Dit heeft tot gevolg dat het vaststellen van een absoluut minimumniveau toezicht is niet mogelijk, omdat het vaststellen van een acceptabel restrisiko een politieke vraag is.

Daarnaast gaat het niet om één minimumniveau van toezicht, bijvoorbeeld het aantal boekenonderzoeken, maar eerder om een minimumniveau van toezicht per inzetbaar instrument van het model, omdat die toezichtinstrumenten samen met de uitvoering en andere vormen van handhaving bepalen of de Belastingdienst haar strategische doelstelling haalt.

Maatschappelijke organisaties staan vaak voor de vraag hoe ze om moeten gaan met de intrinsiek politieke kwestie van de gepercipieerde rechtvaardigheid en proportionaliteit van uitvoeringsvormen. Een aantal van deze organisaties heeft een oplossing gevonden door verschillende potentiële strategieën breder te beoordelen op basis van de maatschappelijke waardering van uiteenlopende resultaten, zoals doelmatigheid en doeltreffendheid, of maatschappelijke kosten en baten. Deze benadering van het waarderen en vergelijken van strategieën wordt besproken in Roobeek et al. (2023). In dit onderzoek is bijvoorbeeld gekeken welke vormen van accommodatie voor statushouders de grootste maatschappelijke waarde genereert. Maatschappelijke waarde werd gedefinieerd als een combinatie van kosten, economische impact en kwaliteit van leven van zowel statushouders als oorspronkelijke bewoners. Een vergelijkbare methode is toegepast in de zorgsector om te bepalen welke mix van investeringen in zorgmedewerkers de beste balans oplevert in termen van kosten, capaciteitsvergroting, medewerkersbetrokkenheid en kwaliteit van zorg. Van Liere et al. (2023) bespreekt hoe het optimaliseren van een fraudedetectiesysteem draait om een afweging tussen de kosten van het systeem, de opbrengsten in de vorm van teruggevorderde bedragen, en de rechtvaardigheid, uitgedrukt in de kans dat fraudeurs worden ontmaskerd en de kans dat niet-fraudeurs valselijk worden beschuldigd. Ook in deze drie voorbeelden gaat het steeds om een mix van mogelijke instrumenten om doelstellingen te bereiken en is er daarom geen sprake van één minimumniveau per instrument om voldoende waarde te genereren.

Gegeven de eerder vastgestelde onmogelijkheid om een absoluut niveau van toezicht te definiëren, interpreteren we de vraag naar de vaststelling van een minimumniveau van toezicht als volgt: Welk minimumniveau van toezicht is acceptabel? Het adjectief acceptabel heeft een driedig bestaansrecht. Ten eerste doet het adjectief acceptabel recht aan het feit dat perfect toezicht per definitie onmogelijk is vanwege beperkingen in toezichtsinstrumenten en beschikbare data, onder andere vanwege privacywetgeving. De zinsnede "waarbij rekening gehouden wordt met acceptabele restrisiko's" in de onderzoeksvraag verwijst hier ook naar: perfectie is niet nodig zolang de imperfectie beperkt is tot een acceptabel niveau. Ten tweede verlegt deze invalshoek de aandacht naar acceptabel, gegeven de keuzes in de uitvoering en andere vormen van handhaving dan toezicht. Ten derde verwijst acceptabel naar politiek aanvaardbaar en adresseert daarmee de opmerking van Professor Spekél dat een minimumniveau van toezicht ook deels een politieke keuze is.

De conclusie van ons onderzoek is dat het wel degelijk mogelijk is om een *acceptabel* en *beheerst* minimumniveau van toezicht vast te stellen. Dit rapport beschrijft hoe een bestaand raamwerk, ondersteund door open-source software⁶, gebruikt kan worden om dit op praktische wijze te realiseren. Dit raamwerk geeft antwoord op de vraag: Welke van alle door de Belastingdienst voorgestelde handavingsmixen levert de meeste waarde op voor de maatschappij? Onder de aanname dat het scala beschouwde mixen zoveel variatie bevat dat dit scala bij benadering alle mogelijke vormen van toezichtvormen representeert, kan de vorm die de meeste waarde oplevert beschouwd worden als mix waarin alle uitvoerings- en handavingsinstrumenten tezamen een acceptabel minimumniveau van toezicht opleveren. 'Acceptabel' wordt aldus geïnterpreteerd als 'volgens de maatschappij de beste van de beschouwde opties', of, negatiever geformuleerd als 'volgens de maatschappij de minst slechte van de beschouwde opties'. 'Beheerst' verwijst naar een analyse vooraf van de verwachte prestaties en risico's van de gekozen vorm toezicht.

Cruciaal in deze interpretatie is dat duidelijk is hoe de maatschappelijke waarde van enige vorm van toezicht bepaald kan worden. De bepaling van deze maatschappelijke waarde doet ook recht aan de opmerking van professor Speklé dat de Belastingdienst niet *intern* kan bepalen wat het minimumniveau van toezicht is. In de volgende paragraaf komen we terug op de vraag hoe de maatschappelijke waarde met hulp van een blik van binnen en buiten bepaald kan worden.

Een andere implicatie van onze interpretatie van de onderzoeksvraag is de behoefte aan een brede variatie van uitvoerings- en handavingsmixen die vergeleken kunnen worden in termen van de maatschappelijke waardering op de effecten van die mix. Immers, de beste van slechts twee beschouwde potentiële mixen kan moeilijk 'acceptabel' genoemd worden als er nog veel meer en wellicht betere potentiële mixen bestaan. De huidige uitvoerings- en handavingsmix van de Belastingdienst vormt hierbij één van de mixen die gewaardeerd en vergeleken zou moeten worden. Merk op dat we om een acceptabel minimumniveau van toezicht te bepalen, we dus breder moeten kijken dan toezicht alleen: We beschouwen niet alleen toezichtinstrumenten maar ook andere vormen van handhaving en de uitvoering. In de volgende paragraaf gaan we nader in op de vraag hoe verschillende uitvoerings- en handavingsmixen opgesteld kunnen worden.

Hoe een uitvoerings- en handavingsmix rendeert qua maatschappelijke waarde hangt uiteraard niet alleen af van die mix zelf. Tal van exogene onzekerheden spelen ook een rol, zoals de kans dat een aangifte ondanks de beste bedoelingen om deze compliant in te vullen, toch nog verkeerd wordt ingevuld. Hoewel een uitvoerings- en handavingsmix de impact van exogene factoren kan beïnvloeden, heeft de Belastingdienst deze exogene factoren niet zelf in de hand. In de volgende paragraaf bespreken we hoe scenario's kunnen worden opgesteld om met deze onzekerheden om te gaan.

Effecten van uitvoerings- en handavingsmixen zijn typisch gedefinieerd in verschillende meeteenheden. De monetaire kosten van een uitvoerings- en handavingsstrategie worden idealiter in euro's uitgedrukt, terwijl niet-gedetecteerde foutieve aangiften meestal worden weergegeven als een percentage of als een aantal. Teneinde te kunnen bepalen welke van de beschouwde uitvoerings- en handavingsmixen het grootste totale effect heeft, moeten deze verschillende eenheden wel geaggregeerd worden. In de volgende paragraaf gaan we dieper in op welke rol utiliteitscurves, die ook wel waarderingsfuncties worden genoemd, spelen om dit mogelijk te maken.

⁶ Open-source software is software die gratis gedownload en gebruikt kan worden. In dit geval is de software in de vorm van Python code en voorbeeldcasussen te downloaden van <https://github.com/responsible-business-decision-making/trbs>. Voor de volledigheid merken we op dat de Nederlandse overheid een "open-tenzij" beleid voert ten aanzien van software (zie [Rapport Opensourceswerken \(overheid.nl\)](#)) en dat het gebruik van open-source software niet impliceert dat de in- & output data van die software ook openbaar wordt.

3.2. Raamwerk ter bepaling acceptabel minimumniveau toezicht

In deze paragraaf bespreken we een hoe een bestaand raamwerk, dat uitgebreider besproken wordt in Roobeek (2023), toegepast kan worden om een acceptabel minimumniveau toezicht te definiëren door verschillende uitvoerings- en handavingsmixen met elkaar te vergelijken gegeven de uitvoering en andere vormen van handhaving. Binnen dit raamwerk worden de verwachte resultaten van verschillende uitvoerings- en handavingsmixen in termen van voor de maatschappij relevant geachte effecten als gelijkheid van behandeling, doelmatigheid, doeltreffendheid en doorlooptijd gesimuleerd en onderling vergelijkbaar gemaakt. Ook behandelt het raamwerk de afweging van het belang van diverse effecten en de onzekerheid daarin. De aldus hoogst gewaardeerde mix voldoet aan het maatschappelijk wens voor een acceptabel niveau van toezicht. Op deze manier vormt het raamwerk de identificatie van de hoogte van de determinanten welke van belang zijn bij het komen tot een acceptabel minimumniveau van toezicht. We beschrijven dit raamwerk, dat uit drie fasen bestaat, eerst op hoofdlijnen per fase en illustreren de werking ervan vervolgens aan de hand van een sterk versimpeld voorbeeld. In het volgende hoofdstuk volgt een gedetailleerdere beschrijving inclusief een projectie op de horecabranche.

3.2.1. Fase 1: Dialoog met belanghebbenden en experts uit verschillende disciplines

De eerste fase is exploratief van aard en vereist het betrekken van een zo breed mogelijk scala aan belanghebbenden en ervaringsdeskundigen met verschillende achtergrond. Deze fase bestaat uit vier stappen. De bepaling van de scope van toezicht vindt in de eerste stap plaats: voor welke doelgroep en – eventueel – belastingmiddelen wordt de vorm van toezicht bepaald? Uit onze interviews binnen de Belastingdienst blijkt dat – gegeven de heterogeniteit van belastingplichtigen dit een cruciale stap is die zal leiden tot het opknippen van het geheel tot een groot aantal verschillende scopes van toezicht, ieder voor een eigen doelgroep. Voor elk aldus bepaalde scope worden de drie fasen dus apart doorlopen. In de tweede stap worden de effecten van toezicht op de maatschappij en de Belastingdienst voor de gekozen doelgroep en belastingmiddelen gedefinieerd, geoperationaliseerd en extern afgestemd. Uit onze bureaustudie en de interviews blijkt dat de Belastingdienst (uiteraard) al vele prestatie-indicatoren heeft ingericht om deze effecten te meten. Deze stap is dus meer een kwestie van deze indicatoren inventariseren en reviewen op bruikbaarheid en volledigheid. De derde stap definieert - aan de hand van de vrijheidsgraden die de Belastingdienst heeft om toezicht in te richten langs verschillende toezichtinstrumenten - diverse mogelijke handavingsmixen. Ook hier blijkt uit onze bureaustudie en de interviews dat de Belastingdienst constant onderzoek doet hoe de toezichtsmix verbeterd kan worden door variaties en toevoegingen te overwegen en op basis van literatuur en historische data proberen te voorspellen wat de effecten daarvan zijn. Hierbij beschouwt men toezichtvarianten in samenhang met keuzes voor de uitvoering en andere vormen van handhaving dan toezicht. Het verschil in deze stap met wat er nu al gedaan wordt is dat enkele mogelijke variaties expliciet worden vastgesteld als mogelijke toezichtsmix en in de volgende fase ex ante gescoord worden. Uiteraard is één mogelijke uitvoerings- en handavingsmix al expliciet gedefinieerd, dat is de huidige mix. Door de exogene variabelen, die samen met de uitvoerings- en handavingsmix de effecten van toezicht bepalen, te laten variëren, worden in de vierde stap scenario's opgesteld. Deze scenario's geven mogelijke, coherente toekomstvoltrekkingen weer van factoren die de Belastingdienst niet zelf in de hand heeft maar wel mede bepalen wat de effecten van fiscaal toezicht zijn. Ook deze stap is niet geheel nieuw voor de Belastingdienst, want uit onze interviews blijkt dat er al veel nagedacht wordt over trends in de maatschappij (bijvoorbeeld de opkomst van influencers) en tot welke aanpassingen in de uitvoerings- en handavingsmix deze trends kunnen leiden.

3.2.2. Fase 2: Bureau- en veldonderzoek

De tweede fase is vrij technisch van aard en bestaat uit één stap waarin op basis van bureau- en veldonderzoek een simulatiemodel wordt opgesteld dat ex ante de effecten van elke mogelijke uitvoerings- en handavingsmix onder elk scenario doorrekent. Ook worden zogenaamde waarderingsfuncties bepaald. Deze functies maken het mogelijk om de effecten van toezicht in

termen van bijvoorbeeld gelijkheid van behandeling, doeltreffendheid, doelmatigheid en doorlooptijd, die verschillende meeteenheden kennen, onderling vergelijkbaar te maken. Uit de voor dit rapport aangereikte informatie en de interviews blijkt dat dat binnen de Belastingdienst zorgvuldige afweging van verschillende effecten al plaatsvindt. Dit vereist nu eenmaal enige vorm van het onderling vergelijkbaar maken van verschillende effecten. De expliciete bepaling van waarderingsfuncties in deze context lijkt dus wel nieuw te zijn voor de Belastingdienst, maar de impliciete bepaling ervan niet.

Uit Roobeek et al. (2023) blijkt ook dat bij belangrijke investeringsbeslissingen organisaties wel degelijk effecten met verschillende eenheden onderling vergelijken via een soort van waarderingsfuncties maar deze niet expliciteren of deze niet zo noemen. Een voorbeeld van zo'n impliciete waarderingsfunctie is vragen hoeveel extra netto contante waarde nodig is om te compenseren voor een één jaar langere terugverdientijd. Een voorbeeld van een waarderingsfunctie die niet zo genoemd wordt is de vermenigvuldiging van CO2 reductie met een zogenaamde internal carbon price in een business case voor de investering in energiebesparende apparatuur. Uit de casussen in Roobeek et al. (2023) komt overigens naar voren dat automatisch gegenereerde waarderingsfuncties in de praktijk vaak goed werken en daarmee ook enigszins impliciet blijven. Tegelijkertijd, om verantwoording af te kunnen leggen over waarom een gekozen vorm van toezicht van een acceptabel niveau is, verdient het aanbeveling om deze waarderingsfuncties wel onder ogen te zien en deze waar nodig te toetsen bij de de in de eerste fase benoemde belanghebbenden. In paragraaf 4.2.1 gaan we hier dieper op in.

De diepgang van deze stap 5 is heel schaalbaar en is afhankelijk van de beschikbare tijd en middelen, waarbij opgemerkt dient te worden dat elk simulatiemodel van deze aard hoogstens een beperkte weergave van de toekomst oplevert.

3.2.3. Fase 3: Afweging van strategische prioriteiten en bepaling risicobereidheid

De derde fase bestaat uit twee stappen. Eerst wordt gevraagd hoe belangrijk de verschillende effecten ten opzichte van elkaar geacht worden. Dit levert de strategische prioriteiten op en vereist een afweging van belangen, zoals bijvoorbeeld: hoe belangrijk is doorlooptijd ten opzichte van de kosten van uitvoering en de opbrengsten in de vorm van de verkleining van de zogenaamde tax gap? Omdat alle effecten dankzij de waarderingsfuncties al in dezelfde eenheid worden uitgedrukt, beperkt deze stap zich tot het ophalen van strategische prioriteiten in de vorm van relatieve gewichten per soort effect van toezicht. In de laatste, zevende stap, wordt gevraagd hoe zwaar elk scenario meegewogen moet worden in de bepaling van de meest robuuste uitvoerings- en handhavingmix. Net als de zesde stap beperkt deze stap zich tot het ophalen van relatieve gewichten, maar nu gaat het niet om een relatief gewicht per soort effect van toezicht ter bepaling van strategische prioriteiten, maar om een relatief gewicht per scenario ter vastlegging van de risicobereidheid.

In het volgende hoofdstuk worden deze stappen verder uitgewerkt voor de Belastingdienst, maar nu volgt eerst een illustratie van hoe dit raamwerk werkt in een sterk versimpelde situatie.

3.2.4. Illustratie aan de hand van een sterk versimpeld voorbeeld

Het sterk versimpelde voorbeeld op basis van fictieve data luidt als volgt. Stel dat de Belastingdienst zou kunnen volstaan met één "one-size-fits-all" vorm van toezicht voor alle doelgroepen en alle belastingmiddelen. Stel dat het resultaat van stap twee is dat er slechts drie effecten relevant geacht worden en dat per effect geldt dat de twee in tabel 1 vermelde indicatoren per effect volstaan om het effect te operationaliseren. We houden in dit voorbeeld de effecten aan zoals geformuleerd in Van Liere et al. (2023). In stap 3 stelt zij vier mogelijke vormen van toezicht op. De eerste kolom van tabel 1 bevat een beschrijving van elke van deze vormen. Het doel is om daaruit de best scorende te selecteren die als vorm van toezicht te beschouwen die voldoet aan een minimumniveau van toezicht. Voor het gemak nemen we aan dat de effecten van deze vorm van toezicht deterministisch bepaald kunnen worden. Er zijn dus geen exogene variabelen. In onderstaande tabel zijn de vier mogelijke vormen van toezicht gescoord op de zes in stap 2 bepaalde effecten (kolom 2 t/m 7).

Tabel 1. Uitwerking fictieve determinanten

Uitvoerings- en handhavingsmix	Doorlooptijd		Doelmatigheid		Rechtvaardigheid	
	Gemiddeld	Band-breedte	Kleine geld-stroom	Grote geld-stroom	Fout posi-tieven ⁷	Fout nega-tieven ⁸
Focus op belastingmoraal	2 maanden	6 maanden	€1,5 mln	€5 mln	70%	10%
Robuust aangifte proces	2 maanden	6 maanden	€1 mln	€8 mln	70%	10%
Intensief toezicht achteraf	4 maanden	8 maanden	€3 mln	€13 mln	40%	30%
Mix	3 maanden	7 maanden	€1,8 mln	€10 mln	50%	15%

Om deze scores, die verschillende meeteenheden hebben, onderling vergelijkbaar te maken, stellen we waarderingsfuncties op. De belangrijkste input hiervoor is aangeven of de resultaten van het type Larger-the-Better (bijvoorbeeld de grote geldstroom) of Smaller-the-Better (bijvoorbeeld de proportie foutpositief) is. Per effect kunnen ook de waarden voor minimale en maximale waardering worden ingesteld. Daarnaast zijn er verschillende vormen mogelijk: er kunnen zowel lineaire als ingewikkeldere waarderingsfuncties gebruikt worden. Zo kunnen waarderingsfuncties rekening houden met verzadigingspunten en met de wet van de afnemende meeropbrengst, of die zo ingesteld worden dat ze fungeren als knock-out criterium als een effect boven of onder een drempelwaarde komt. Zie Roobeek et al. (2023) voor meer details en verwijzingen naar open-source software die op basis van een paar parameters automatisch waarderingsfuncties bepaalt. Er kan ook gekozen worden om de automatisch bepaalde waarderingsfuncties in een strategische dialoog met de belanghebbenden verder te kalibreren of afhankelijk van elkaar te maken. Stel dat dit proces leidt tot de volgende tabel met waardering per vorm van toezicht per effect.

Tabel 2. Waardering van effecten op één schaal van 0 tot 100

Uitvoerings- en handhavingsmix	Doorlooptijd		Doelmatigheid		Rechtvaardigheid	
	Gemiddeld	Band-breedte	Kleine geld-stroom	Grote geld-stroom	Fout posi-tieven	Fout nega-tieven
Focus op belastingmoraal	70	60	60	55	40	80
Robuust aangifte proces	70	60	70	60	40	80

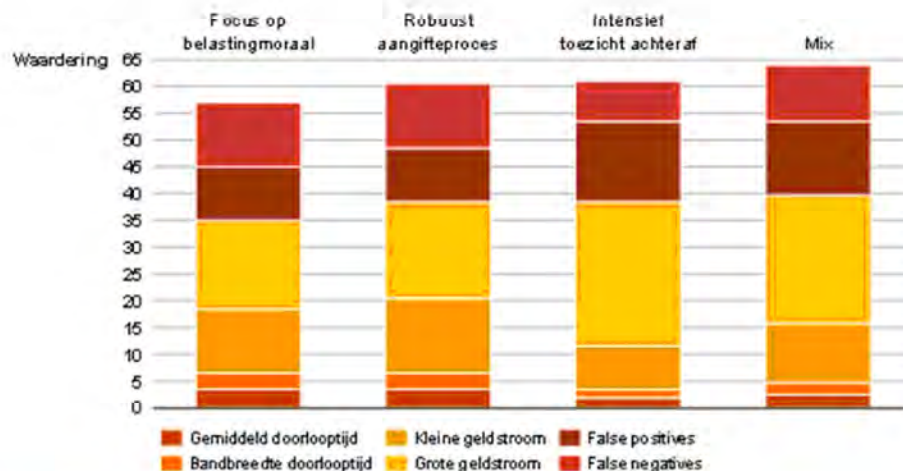
⁷ Dit betreft een nader onderzochte aangifte is die goed blijkt te zijn. Deze "fout positief" noemen is een enorme simplificatie, want in werkelijkheid kan er juist een preventieve werking uitgaan van zo'n onderzoek en kan dit gezien worden als een beloning voor correct aangifte doen. Er hoeft dus niet per se sprake te zijn van "fout".

⁸ Onterecht niet gecorrigeerd

Intensief toezicht achteraf	40	30	40	90	60	50
Mix	50	45	55	80	55	70

Nu de resultaten door waarderingsfuncties onderling vergelijkbaar zijn gemaakt, kunnen we ze consolideren in één totale waarde per uitvoerings- en handhavingsmix. Indien alle waarderungen van de resultaten eenvoudigweg opgeteld worden per mogelijke mix, dan worden alle effecten even zwaar meegewogen. Niet elke effect hoeft echter van even groot strategisch belang te zijn. Zo kan het belang van doelmatigheid hoger geacht worden dan dat van doorlooptijd of rechtvaardigheid. Wij pleiten voor een strategische dialoog om het relatieve belang van de waarden te bepalen bij het besluit over de meest geschikte mix en hierin afstemming te zoeken met belanghebbenden. Deze dialoog leidt tot een gewicht per soort effect dat aangeeft hoe belangrijk dat type effect wordt gevonden.

De waardering van effecten en de relatieve prioritering hiertussen maakt het mogelijk om de verschillende uitvoerings- en handhavingsmixen met elkaar te vergelijken. De resultaten uit het fictieve voorbeeld zijn weergegeven in figuur 6. In dit fictieve voorbeeld zou de optie "Mix" het hoogst gewaardeerd worden op alle resultaatdeterminanten (effecten) en daarmee het minimumniveau van toezicht omvatten. Merk op dat als men zou besluiten om veel meer belang te hechten aan fout negatieven, de vorm van toezicht die de nadruk legt op het verbeteren van de robuustheid, uiteindelijk de optie "Mix" zal verslaan. Immers, "Robuust aangifteproces" heeft samen met "Focus op belastingmoraal" de hoogste waardering op "Fout negatieven" en scoort op de andere effecten hoger dan "Focus op belastingmoraal".



Figuur 6. Waardering van vier mogelijke handavings- en uitvoeringsmixen

4. Toepassing afwegingskader

In dit hoofdstuk illustreren we per fase per stap de toepassing van het raamwerk in iets meer detail. We sluiten het hoofdstuk af met een verkenning van hoe het raamwerk zou kunnen werken in het fiscale toezicht op de horecabranche. We benadrukken dat dit conform de onderzoeksopdracht nog steeds slechts een verkenning betreft.

4.1. Fase 1: Dialoog met belanghebbenden en experts uit verschillende disciplines

4.1.1. Stap 1: Bepaling scope

De eerste stap van fase 1 bepaalt welke doelgroep en belastingmiddelen op een aantal fiscale kenmerken homogeen genoeg geacht worden om te adresseren met één vorm van toezicht. Uit de interviews blijkt dat dit uitdagend kan zijn binnen het MKB. Binnen een industrie-code kan bijvoorbeeld nog steeds veel variatie zitten. Om een groep bedrijven homogeen van omvang te maken moet een keuze gemaakt worden uit verschillende indicatoren. Binnen GO lijkt het moeilijk om überhaupt homogene groepen te vinden.

Voor de toepassing van het afwegingskader is het cruciaal om zoveel mogelijk uit te gaan van een homogene groep ondernemers waarvoor in zekere mate data beschikbaar is. Zo'n groep kan over de tijd door allerlei invloeden en nieuwe trends toch weer heterogener worden waardoor afbakening van zo'n groep een uitdagende taak is en het nodig kan zijn om deze over de tijd aan te passen.

4.1.2. Stap 2: Bepaling relevante effecten van toezicht

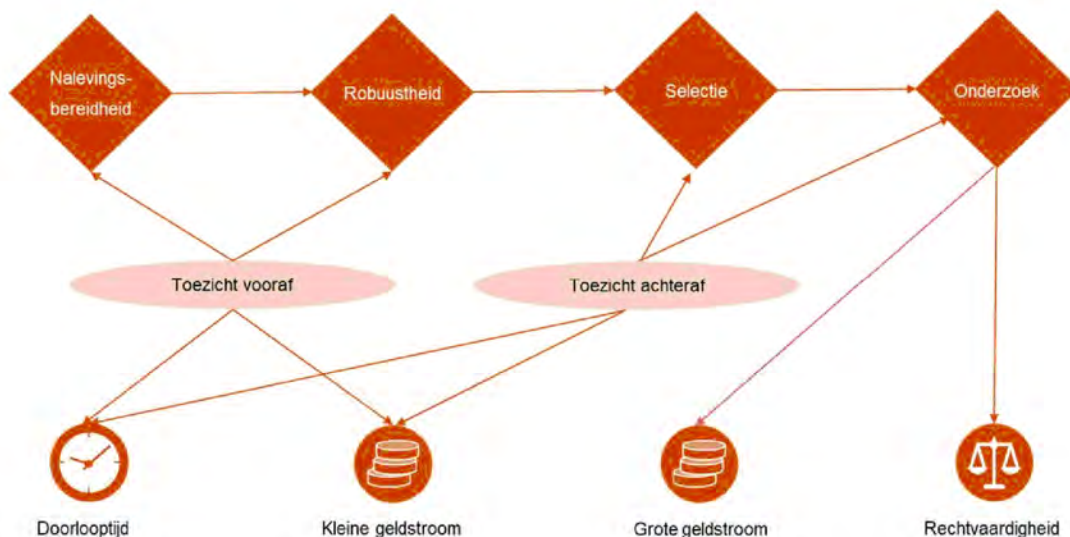
De tweede stap is het bepalen van de beoogde effecten van toezicht en de operationalisatie (meetbaar maken) van die effecten. Om inzichtelijk te maken wat een acceptabel minimumniveau van toezicht is voor de Belastingdienst is het van belang om breder te kijken dan enkel het restrisico op foutieve aangiften zonder correctie. Om een acceptabel minimumniveau van toezicht te bepalen is er een afweging nodig tussen meerdere effecten van toezicht. Wat deze effecten zijn dient afgestemd te worden met de maatschappij. In Van Liere et al. (2023) wordt geopperd dat de uitkomst van deze afstemming zou kunnen zijn dat de operationalisatie van effecten langs drie assen plaatsvindt: rechtvaardigheid, doorlooptijd en doelmatigheid:

- **Rechtvaardigheid** Het percentage of aantal "fout negatieven" (foutieve aangiften die niet als foutief worden aangemerkt en dus niet gecorrigeerd kunnen worden) en "fout positieven" (correcte aangiften die onterecht nader onderzocht worden en hinder of onrust zouden kunnen opleveren bij de aangever) kunnen gebruikt worden om rechtvaardigheid te operationaliseren. Zoals eerder opgemerkt is dit een sterk vereenvoudigde voorstelling van zaken en verdient het aanbeveling om "fout positieven" uit te splitsen in "fout positieven" die wel of geen substantiële hinder of onrust opleveren. Immers, zonder hinder kan een "fout positief" ook – mits de behandeling van de belastingplichtige als zodanig wordt ingericht - geïnterpreteerd worden als een waardevolle erkenning van een correcte aangifte. Ook kan in plaats van percentages voor aantallen gekozen worden.
- **Doorlooptijd:** De doorlooptijd betreft de tijd vanaf het moment van indienen van een aangifte tot de definitieve beschikking. Omdat de doorlooptijd per aangifte verschilt, ligt het voor de hand om zowel de gemiddelde doorlooptijd als de spreiding in doorlooptijd als effect te benoemen:
 - **Gemiddelde doorlooptijd:** De gemiddelde tijd tussen het indienen van de aangifte en de definitieve vaststelling
 - **Bandbreedte doorlooptijd:** Het verschil tussen de minimale en maximale doorlooptijd van aangiften (of de breedte van een 95% betrouwbaarheidsinterval)

- **Doelmatigheid:** De doelmatigheid omvat de kosten die de Belastingdienst maakt voor het uitvoeren van haar activiteiten (= kleine geldstroom) minus de opbrengsten van de activiteiten door het voorkomen van het mislopen van belastinginkomsten (grote geldstroom). Deze twee kosten worden vaak anders gewaardeerd door de maatschappij. Daarom bestaat doelmatigheid uit twee effecten
 - **Kleine geldstroom:** De kosten van het uitvoeren van toezicht
 - **Grote geldstroom:** De extra opgehaalde belastinggelden als gevolg van het toezicht.

4.1.3. Stap 3: Inventarisatie van mogelijke uitvoerings- en handhavingsmixen

Zoals eerder aangegeven interpreteren we een acceptabel minimumniveau van toezicht als een vorm van toezicht die de hoogste maatschappelijke waardering geniet van – bij benadering – alle mogelijke vormen van toezicht. Om recht te doen aan de bijzin “bij benadering” is het van belang om in deze stap een breed scala aan vormen van toezicht te verkennen en deze ook nog in samenhang met de uitvoering en andere handhavingsvormen dan toezicht te beschouwen. Dit verklaart waarom deze stap onderdeel uitmaakt van de exploratieve fase en gebruik maakt van experts uit diverse disciplines. Deze stap maakt onderdeel uit van een inductieve benadering: het gaat er niet om vooraf slechts een paar mogelijke vormen van toezicht te definiëren waarvan men denkt dat ze wel zullen werken. Het is juist de bedoeling om hier de breedte op te zoeken en pas in fase 2 op basis van een ex ante simulatiemodel dat gebaseerd is op bureau- en veldonderzoek de effecten uit te rekenen. Op hoofdlijnen laat Figuur 7 zien hoe toezicht voor- en achteraf niet alleen de beslismomenten (weergegeven met ruitjes) beïnvloeden, maar ook de doorlooptijd en de kleine geldstroom. Ook toont de figuur hoe de uitvoerings- en handhavingsmix indirect de grote geldstroom en de rechtvaardigheid beïnvloeden.



Figuur 7. Relatie tussen toezicht en effecten daarvan

Een uitvoerings- en handhavingsmix kan geparametriseerd worden aan de hand van zogenaamde endogene parameters, die door de Belastingdienst bepaald kunnen worden. Door deze endogene parameters van een waarde te voorzien ligt de uitvoerings- en handhavingsmix vast. Per beslismoment (eerste kolom) toont Tabel 3 slechts enkele voorbeelden van endogene parameters (derde kolom). We benadrukken dat dit rapport slechts een verkenning betreft van een manier om een minimumniveau van toezicht te bepalen. Bij het werkelijk toepassen van deze manier zullen deze parameters verder uitgewerkt en uitgebreid moeten worden. Merk op dat de

Belastingdienst als uitvoeringsorganisatie geen directe zeggenschap heeft over de fiscale wet- en regelgeving, maar wel bepaalt hoe duidelijk deze wordt uitgelegd in het aangifteproces.

Tabel 3. Mogelijke voorbeelden van endogene (stap 3) en exogene (stap 4) parameters

Beslismoment	Endogene parameters	Exogene parameters
Nalevingsbereidheid	Gemak om aan verplichtingen te voldoen	Belastingmoraal Sociale normen Gepercipieerde pakkans
Robuustheid	Duidelijkheid wet- en regelgeving Duidelijkheid procedures Ondersteuning bij het aangifteproces	Kennisniveau belastingplichtige Complexiteit activiteiten belastingplichtige Fiscale beheer belastingplichtige
Selectie	Personele capaciteit selectie Beschikbare doorlooptijd selectie Beschikbare gedraginsichten Beschikbaarheid selectiemodules en risicomodellen Mogelijkheid geautomatiseerde checks Omvang selectie	Heterogeniteit doelgroep Mate van verifieerbaarheid gegevens
Onderzoek	Personele capaciteit onderzoek Beschikbare toezicht producten Beschikbare doorlooptijd onderzoek Samenwerking (FIOD, uitvoeringsorganisaties en buitenlandse belastingdiensten) Informatievoorziening	Complexiteit van wet- en regelgeving

4.1.4. Stap 4: Inventarisatie van mogelijke scenario's

Naast de endogene parameters zijn er ook parameters die niet in handen zijn van de Belastingdienst maar die medebepalend zijn in wat de effecten van toezicht zijn. Kolom 3 van Tabel 1 toont voorbeelden van zulke parameters. Net zoals in stap 3 endogene parameters de ruimte van mogelijke uitvoerings- en handavingsmixen opspannen, dienen exogene parameters om alle mogelijke toekomstvoltrekkingen (scenario's) te parametriseren. Met andere woorden: door voor elke exogene parameter één inschatting te maken, ontstaat één mogelijk scenario. Door dit meerdere malen te doen ontstaan verschillende scenario's, die in de volgende fase ex ante doorgerekend kunnen worden om inzicht te krijgen in de effecten van uitvoerings- en handavingsmixen onder verschillende omstandigheden.

De sectorspecifieke exogene factoren binnen de horeca hebben hoofdzakelijk betrekking op de beslissende momenten aan de voorkant van het aangifteproces: de nalevingsbereidheid en

robuustheid. De endogene determinanten, oftewel de gehanteerde uitvoerings- en handhavingmix, zijn binnen de sector in een groeiende mate gericht op deze beslissende momenten via toezicht vooraf.

4.2. Fase 2: Bureau- en veldonderzoek

Fase 2 omvat één stap: die de ontwikkeling van een simulatiemodel betreft.

4.2.1. Stap 5: Ontwikkeling van simulatiemodel

Het simulatiemodel dient om ex ante de effecten van verschillende vormen van toezicht onder verschillende scenario's kwantitatief in kaart te brengen. Aldus worden de pijlen in Figuur 7 gemodelleerd, waarbij we opmerken dat deze relaties ook afhangen van exogene parameters. Hierbij kan geput worden uit onderzoek naar eerder onderzochte effecten van vormen toezicht, zoals bijvoorbeeld beschreven in Bijlage B. Ook kan gekeken worden naar interne data van de Belastingdienst waarin vastgelegd is wat de gevolgen waren van variaties in vormen van toezicht op de effecten van toezicht. Zo werkt de Belastingdienst al vaak met statistische steekproeven. Deze steekproeven zijn een vorm van waarheidsvinding die uitstekend gebruikt kunnen worden om ex ante het effect van bijvoorbeeld een nieuw risicoselectie met terugwerkende kracht te kunnen toetsen, zoals dit ook bij DUO heeft plaatsgevonden (PwC, 2024).

In gevallen dat noch extern onderzoek iets oplevert noch zulke interne data beschikbaar is, is er de mogelijkheid om gebruik te maken van expertmeningen, Monte-Carlo simulaties (het doorrekenen van vele trekkingen uit een statistische verdeling voor mogelijke waarden voor relaties). Een "last resort" is om inzicht te krijgen in effecten van toezicht door veldonderzoek: in een pilotstudie worden gedurende een beperkte periode gemeten wat mogelijke vormen van toezicht opleveren per doelgroep. Het moge duidelijk zijn dat er vele praktische nadelen aan zo'n pilotstudie kleven. Het is niet voor niet een last resort. In Roobeek e.a. (2023) staat beschreven hoe dit ook een iteratief proces is: Vaak blijkt in fase 3 dat slechts een paar relaties doorslaggevende invloed hebben in de keuze voor welke vorm van toezicht tot een acceptabel minimumniveau leidt. In de praktijk betekent dit dat relaties eerst vrij grof gemodelleerd worden en dat pas in een later stadium meer geïnvesteerd wordt om de relaties nauwkeuriger in kaart te brengen. Merk op dat de ex ante analyse in deze fase in het geval van risicoselectie op basis van algoritmen mooi aansluit bij de tendens om vooraf te willen analyseren wat de beperkingen van en biases in algoritmen zijn. Zo is het mogelijk om met een ex ante analyse te onderzoeken of de doelstelling om belastingplichtigen gelijk te behandelen behaald kan worden. Het recente onderzoek naar risicoselectie door Dienst Uitvoering Onderwijs (PwC, 2024) en ook de Responsible AI Act zijn hier voorbeelden van.

Als eenmaal de effecten van elke mogelijke vorm van toezicht onder elk scenario gekwantificeerd zijn ontstaat de vraag: hoe kunnen deze effecten, die elk een verschillend meeteenheid hebben, onderling vergeleken worden? Hoe vergelijken we bijvoorbeeld de waarde van de kleine en grote geldstroom, die in euro's gemeten worden, met de waarde van doorlooptijd, die in dagen gemeten wordt, met de percentages die weergeven hoeveel aangiften onterecht wel en niet gecontroleerd en/of gecorrigeerd worden? Hiertoe dienen we nog een extra stap te zetten opdat de waarden met elkaar vergeleken kunnen worden. We doen dit met behulp van zogenoemde waarderingsfuncties die de verschillende resultaten transformeren naar één meetschaal. Waarderingsfuncties worden in de economische literatuur ook wel utiliteitscurves genoemd. Er zijn veel geavanceerde manieren om utiliteitscurves te bepalen, bijvoorbeeld met conjoint analysis. Roobeek et al (2023) behandelt de meeste daarvan en onderzoekt de werking daarvan in de praktijk. In twee eerdergenoemde voorbeelden over de zoektocht naar de maatschappelijk meest aanvaardbare accommodatie van vluchtelingen en het best passende personeelsbeleid gegeven de strategie van de zorginstelling blijkt dat de volgende manier om waarderingsfuncties te bepalen leidt tot resultaten die in de ogen van de belanghebbenden al acceptabel werd gevonden: Van elke resultaatdeterminant (effect) wordt bepaald of deze "smaller-the-better" of "larger-the-better" is. De kleine geldstroom en het aantal fout positieven zijn voorbeelden van "smaller-the-better", de grote geldstroom is een voorbeeld van "larger-the-better". Vervolgens wordt bepaald of de waardering lineair toe- of afneemt bij een stijging van het resultaat of

dat de wet van de afnemende meeropbrengst van toepassing is. Voor accommodatiekosten werd bijvoorbeeld gekozen voor een lineaire waarderingsfunctie, maar voor de meeste componenten van kwaliteit van leven werd gekozen voor de wet van de afnemende meeropbrengst: boven een bepaald minimumniveau neemt de waardering niet meer lineair toe bij de stijging van de waarde van die component. Tot slot wordt bepaald welk bereik de waarden van de effecten heeft: wat zijn de hoogste en laagste waarde die het effect aanneemt? Afhankelijk van of er sprake is van "larger-the-better" of "smaller-the-better" wordt aan de hoogste en laagste waarde een waardering van 100 of 0 toegekend. De waardering op tussenliggende waarden wordt bepaald afhankelijk van of de waarderingsfunctie lineair is of niet. Deze simplistische manier van waarderingsfuncties kan eenvoudig verfijnd worden met open-source software⁹ door bijvoorbeeld verzadigingspunten (boven of onder een bepaalde waarde neemt de waardering niet meer toe) of show-stoppers (onder of boven een bepaalde waarde is de waardering zo negatief dat ongeacht de waardering op andere resultaatdeterminanten is de totale waardering te laag).

Nu de effecten onderling vergelijkbaar zijn gemaakt, hebben we een driedimensionale set van data: voor elk effect weten we de waardering voor elke beschouwde mogelijke vorm van toezicht onder elke scenario. Omdat het te complex is om op basis van een driedimensionale set te bepalen wat een vorm van toezicht is die het best presteert, wordt deze set in fase 3 geaggregeerd tot één dimensionale set.

4.3. Fase 3: Afweging van strategische prioriteiten en bepaling van risicobereidheid

Fase 3 bestaat uit twee stappen: het afwegen van strategische prioriteiten en de bepaling van de risicobereidheid. De uitleg van deze stappen luidt als volgt.

4.3.1. Stap 6: Afweging van strategische prioriteiten

De eerste stap bestaat uit het bepalen van strategische prioriteiten: in een strategische dialoog wordt bepaald hoe belangrijk de verschillende effecten van toezicht zijn door aan elke effect een relatief gewicht toe te kennen. Door de gewogen som van de waardering per effect te berekenen is de dimensie van effecten als het ware weggeaggregeerd. Dit resulteert in een tweedimensionale set van waarderungen: één waardering per vorm van toezicht per scenario.

4.3.2. Stap 7: Bepaling van risicobereidheid

In de tweede stap wordt de risicobereidheid bepaald: in een strategische dialoog wordt bepaald hoe zwaar elk scenario wordt meegewogen in de uiteindelijke beslissing over welke uitvoerings- en handhavingsmix de hoogste maatschappelijke waardering krijgt. Dit betekent dat de scenariodimensie weggeaggregeerd wordt door aan elk scenario een relatief gewicht toe te kennen. De aldus gewogen som over de waarderungen per scenario leiden tot één waardering per mogelijke vorm van toezicht. De vorm van toezicht met de hoogste waardering noemen we het toezicht met de grootste mogelijke maatschappelijke waardering onder de beschouwde mogelijke vormen van toezicht (gegeven de keuzes in de uitvoering en andere vormen van handhaving dan toezicht). Deze vorm toezicht is daarmee de uitvoerings- en handhavingsmix die voldoet aan het minimumniveau van toezicht. Hierbij merken wij op dat iedere vorm van uitvoerings en handhaving samen gaat met risico. Derhalve dient het risico meegewogen te worden in de bepaling van de uiteindelijke uitvoerings- en handhavingsmix.

Om de risicobereidheid zo vast te stellen dat er draagvlak voor is, dienen interne en externe belanghebbenden onderdeel uit te maken van het gesprek, waarbij eenieder eigenaarschap draagt om tot een gedragen niveau te komen. De Belastingdienst draagt als uitvoerder het eigenaarschap over het gehele proces en is dan ook verantwoordelijk voor de gesprekken met de belanghebbenden. De risico's die hierop betrekken hebben zijn zowel intern als extern te bepalen en verschillen per gekozen mix. Risico's kunnen bepaald worden per toezichtinstrument.

⁹ <https://github.com/responsible-business-decision-making/trbs>

Merk op dat stap 6 en 7 samen ook gezien kunnen worden als een selectie op basis van een SWOT analyse per vorm van toezicht: De Strengths zijn de effecten van toezicht waar een vorm van toezicht zich primair op richt om goed te scoren, de Weaknesses de effecten waarop wordt ingeleverd, de Opportunities de scenario's die gunstig uitpakken voor de vorm van toezicht, de Threats de scenario's die de vorm van toezicht juist kwetsbaar maken.

4.4. Illustratie van toepassing binnen de horeca

4.4.1. Stap 1: Bepaling scope

De eerste stap in fase 1 is het bepalen van de scope van het toezicht binnen de horecabranche. Hierbij richten we ons op alle belastingmiddelen die relevant zijn voor horecabedrijven. Door de specifieke kenmerken en uitdagingen van de horeca, zoals financiële problemen en arbeidstekorten, is het belangrijk een gerichte aanpak te ontwikkelen.

4.4.2. Stap 2: Bepaling relevante effecten van toezicht

Deze stap betreft de operationalisatie van effecten van toezicht. De eerder genoemde, algemene operationalisaties kunnen verbijzonderd worden naar wat er speelt in (het geselecteerde gedeelte van) de horecabranche. Zo heeft de horecabranche de afgelopen jaren te maken gehad met ernstige financiële problemen, verergerd door de coronacrisis. Veel horecaondernemers zijn hierdoor in de problemen gekomen en hebben de coronasteunmaatregelen als ontoereikend ervaren. De reserves van veel horecaondernemers zijn uitgeput en velen hebben moeite met het terugbetalen van leningen, waarbij de Belastingdienst en verhuurders de grootste schuldeisers zijn. Dit zou kunnen rechtvaardigen om het effect van toezicht op het inlopen van achterstanden te meten.

4.4.3. Stap 3: Inventarisatie mogelijke uitvoerings- en handhavingsmixen

In deze stap worden diverse mogelijke uitvoerings- en handhavingsmixen voor de horecabranche verkend. Hierbij wordt gebruik gemaakt van experts uit verschillende disciplines om een breed scala aan uitvoerings- en handhavingsmixen te ontwikkelen. Het doel is om verschillende vormen van uitvoering en handhaving inclusief toezicht te verkennen die later kwantitatief kunnen worden geanalyseerd. Die mogelijke vormen van uitvoering en handhaving worden geparametriseerd met endogene parameters (die door de Belastingdienst bepaald kunnen worden) en beïnvloed door exogene parameters (buiten de controle van de Belastingdienst). Een inventarisatie op basis van de inzet en resultaten van toezicht binnen horecabranche over een bepaalde periode kan hiervoor de basis vormen.

Tabel 4 toont een mogelijke invulling van de endogene parameters (determinanten) voor de horeca.

Tabel 4. Endogene determinanten horeca

Endogene determinanten	
Robuustheid	Duidelijkheid wet- en regelgeving en procedures: Een focuspunt van de handhaving in de horeca betreft een beoogde kwaliteitsimpuls van de administratie- en bewaarplicht. Hiertoe wordt er een zelfscan ontwikkeld waarmee ondernemers geïnformeerd worden over de kwaliteit van hun administratie en instrueert over verbeterpunten. Daarnaast is er een landelijke trechteraanpak voor starters ontwikkeld waarbij starters na aanmelding bij de KVK een brief krijgen met een verwijzing naar informerende sites, zodat ondernemers bekend reken met fiscale onderwerpen. Daarnaast wordt de mogelijkheid geboden om in contact te komen met de Belastingdienst, zowel fysiek als digitaal. Ondersteuning bij het aangifteproces: Ondernemers zonder fiscaal dienstverleners worden proactief benaderd via een telefonisch gesprek waarbij er wordt gekeken op welke wijze de ondernemer

	ondersteund kan worden, bijvoorbeeld via een startersbezoek waar ondersteuning of hulp geboden kan worden. Daarnaast wordt het mogelijk gemaakt voor lokale kantoren om een specifiek brancheonderdeel of gebied te benaderen vanuit de preventieve aanpak. Ook wordt erop ingezet om de kennis bij fiscaal dienstverleners te vergroten om op deze manier het aangifteproces bij een breed aantal ondernemers te ondersteunen.
Selectie	Beschikbare selectiemechanismes: De inzet van toezicht achteraf in de horeca volgt uit een vermoeden of een concreet signaal dat er sprake is van mogelijke nalevingstekorten. Daarnaast kan er een aanleiding voor bedrijfsbezoek gevonden worden op basis van niet-concrete signalen uit data-analyse. Bedrijfsbezoeken worden ingezet om de Belastingdienst te helpen bij het maken van gefundeerde vervolgkeuzes door het klantbeeld van de Belastingdienst de actualiseren. Afhankelijk van de bevindingen kan er worden opgeschaald naar een boekenonderzoek. De Belastingdienst kan ook kiezen voor het versturen van PM brieven, voorafgaand aan de selectie.
Onderzoek	Beschikbare toezicht producten: Als handhavingsinstrument worden bedrijfsbezoeken en boekenonderzoeken ingezet, maar enkel de boekenonderzoeken zijn meegenomen. De periode waarin een boekenonderzoek wordt afgerond is afhankelijk van de grootte van de onderneming. Samenwerking: In geval van vermoedelijke fiscale misstanden wordt actief de samenwerking gezocht met lokale ITE-teams en in ernstige gevallen met de FIOD.

4.4.4. Stap 4: Inventarisatie van mogelijke scenario's

Naast de endogene parameters moeten ook exogene parameters worden geïnventariseerd om de effecten van toezicht binnen de horecabranche te begrijpen. Deze exogene parameters helpen bij het creëren van verschillende scenario's die de diverse omstandigheden binnen de horeca weerspiegelen. Voor de horeca zijn nalevingsbereidheid en robuustheid van groot belang, vooral gezien de financiële druk en de personeelsproblemen die de sector ervaart. Daarnaast kampt de sector met tekorten aan arbeidskrachten. Veel personeel is tijdens de crisis vertrokken, deels door demografische veranderingen, een toename van het aantal zzp'ers en het feit dat personeel is overgestapt naar andere sectoren. Hierdoor hebben horecaondernemers moeite om voldoende personeel te vinden. Dit zou het risico op verschillende vormen van illegale arbeid kunnen vergroten.

Tabel 5. Exogene determinanten horeca

Exogene determinanten	
Nalevingsbereidheid	Belastingmoraal: Uit openbaar gemaakt onderzoek van de Belastingdienst blijkt dat er een hoge naleveringsbereidheid is binnen de Horeca. Financiële problemen en problematische schulden dragen echter op negatieve wijze bij aan de belastingmoraal en kunnen leiden tot verschillende vormen van illegale arbeid die hier apart benoemd kunnen worden als exogene determinanten. Aan ander

	voorbeeld van een exogene determinant is de impact van een bepaald type PM brieven op de nalevingsbereidheid.
Robuustheid	Kennisniveau belastingplichtige: Binnen de horecabranche starten maandelijks relatief veel ondernemers. Daarnaast is er beperkte ondersteuning vanuit fiscaal dienstverleners vanwege het gebrek aan middelen van de ondernemers, de beperkte beschikbaarheid van fiscaal dienstverleners en het gepercipieerde risico van het aannemen van nieuwe horecaklanten (mede vanwege strengere regelgeving zoals het UBO-register). Deze dynamieken hebben een negatieve impact op het kennisniveau van belastingplichtigen. Fiscale beheer belastingplichtige: Door de beperkte ondersteuning vanuit fiscaal dienstverleners zijn er administratieve achterstanden ontstaan binnen de sector en zijn er niet altijd jaarcijfers beschikbaar. Daarnaast is uit eerdere boekenonderzoeken gebleken dat de kasadministratie vaak zwak is binnen de horeca en dat er onvoldoende aandacht is voor administratie- en bewaarplicht. Complexiteit activiteiten belastingplichtige: De complexiteit van de activiteiten binnen de Horeca is relatief hoog. Dit komt onder andere door de inkomsten uit fooien, contact geldverkeer en wisselende bedrijfsconcepten.

4.4.5. Stap 5: Ontwikkeling van het simulatiemodel

Het simulatiemodel wordt ontwikkeld om de effecten van verschillende uitvoerings- en handavingsmixen op de horecabranche onder diverse scenario's kwantitatief in kaart te brengen. Hierbij wordt gebruik gemaakt van eerder onderzoek, interne data van de Belastingdienst en veldonderzoek binnen de horeca. Door waarderingsfuncties (utiliteitscurves) te gebruiken, worden de verschillende effecten onderling vergelijkbaar gemaakt. Dit model helpt te bepalen welke uitvoerings- en handavingsmix de grootste maatschappelijke waardering heeft binnen de horecabranche.

4.4.6. Stap 6: Afweging van strategische prioriteiten

Om de resultaten van de uitvoerings- en handavingsmixen te kunnen vergelijken is het van belang om de waardering van uitkomsten te waarderen. In de praktijk zal de waardering van resultaten op zichzelf en ten opzichte van elkaar geoperationaliseerd moeten worden middels een strategische dialoog.

In een strategische dialoog binnen de horecabranche wordt bepaald hoe belangrijk de verschillende effecten van toezicht zijn. Elk effect krijgt een relatief gewicht toegekend, wat resulteert in een gewogen som van de waarderings per effect. Dit helpt om de prioriteiten binnen de horecabranche duidelijk te maken.

4.4.7. Stap 7: Bepaling van risicobereidheid

Tot slot is het van belang om de verschillende scenario's ten opzichte van elkaar te wegen. Dit betreft een operationalisatie van de risicobereidheid. Bij een hoge risicobereidheid kan het optimistisch scenario, waarin het effect van PM brieven bijvoorbeeld hoog wordt ingeschat, relatief zwaar gewogen worden en bij een lage risicobereidheid kan het pessimistische scenario, waarin een moeilijk detecteerbare vorm van illegale arbeid veel voorkomt, relatief zwaar gewogen worden. Op basis van resultaten uit deze analyse kan geconcludeerd worden dat, gegeven de aangenomen waarderings van de resultaten, een uitvoerings- en handavingsmix die vooral gericht is op toezicht vooraf en niet veel capaciteit beschikbaar stelt voor onderzoek het hoogst gewaardeerd wordt.

5. Uitdagingen

Om het maatschappelijk acceptabel niveau van toezicht te identificeren, stelt dit rapport voor om een afwegingskader in de vorm van een raamwerk toe te passen. Dit raamwerk bestaat uit drie fasen, waarin in totaal zeven stappen moeten worden doorlopen. Iedere fase brengt specifieke uitdagingen en afhankelijkheden met zich mee die van invloed zijn op de haalbaarheid, efficiënte en effectiviteit van de toepassing van dit raamwerk. In dit hoofdstuk rapporteren we per fase welke uitdagingen en afhankelijkheden de Belastingdienst mogelijk kan ervaren bij het doorlopen van deze fasen.

5.1. Fase 1: Dialoog met belanghebbenden en experts uit verschillende disciplines

- **Sectorale diversiteit:** Het vormt een uitdaging dat het goede gesprek met belanghebbenden moet plaatsvinden om het abstracte niveau van toezicht per sector vast te stellen. De heterogeniteit van de sectoren vereist een scope en afbakening die in gezamenlijkheid wordt vastgesteld. De verschillende sectoren hebben unieke kenmerken en vereisten, die alle meegenomen moeten worden in de besluitvorming. Hierbij dient de kanttekening geplaatst te worden dat gelijke behandeling wel gewaarborgd moet blijven.
- **Maatschappelijke betrokkenheid:** Het verkrijgen van draagvlak binnen de samenleving is een complex proces dat aanzienlijke kennis, begrip, tijd en energie vergt. Een belangrijke uitdaging hierbij is het definiëren wie precies 'de maatschappij' vertegenwoordigt. Is binnen een parlementaire democratie het betrekken van Tweede Kamer voldoende om de stem van de burger te representeren? En zo ja, hebben parlementariërs voldoende kennis van fiscaal toezicht? Kortom, het betrekken van de samenleving bij het vaststellen van minimumtoezicht is moeilijk, mede door politieke invloeden. Dit proces vereist diepgaande kennis van de handavingspraktijken van de Belastingdienst en alle facetten van toezicht. Het is essentieel dat de samenleving goed geïnformeerd en actief betrokken wordt, zodat het toezicht dat bestempeld wordt als 'voorzien van maatschappelijk draagvlak' dat stempel ook verdient.
- **Vertrouwen en sturing:** De dialoog moet leiden tot de juiste sturing bij de Belastingdienst en een vertrouwenwekkende relatie tussen politiek en Belastingdienst, wat het vertrouwen van burgers vergroot. Dit vertrouwen is cruciaal voor de nalevingsbereidheid en vereist transparantie en consistentie in communicatie en handhaving. De juiste sturing op de in steekproefrapportages genoemde KPI's dient ook vanuit de politiek plaats te vinden. Hierbij dient wel gewaarborgd te worden dat invulling geven aan de toezichtsmix uiteindelijk binnen het eigenaarschap van de Belastingdienst valt.

5.2. Fase 2: Bureau- en veldonderzoek

- **Passend bij scope:** De relaties in het simulatiemodel moeten passen bij de in de eerste fase vastgestelde scope en dus duidelijk worden afgebakend. Het is belangrijk om nauwkeurig te definiëren en te begrijpen welke specifieke regels, normen, verbanden, historische data van toepassing zijn binnen de gekozen scope.
- **Evaluatie van invloeden:** Regelmatige bijstelling van het simulatiemodel is noodzakelijk vanwege nieuwe invloeden binnen sectoren en de complexiteit daarvan. Dit omvat het monitoren van technologische ontwikkelingen, veranderingen in wet- en regelgeving, en andere externe factoren die invloed kunnen hebben op de effectiviteit van toezicht.

- **Effectmeting:** Het inzichtelijk maken van het effect van toezicht binnen diverse sectoren en de perceptie van belastingplichtigen. Dit omvat ook het beoordelen van de effectiviteit van boekenonderzoeken en de zichtbaarheid van neveneffecten. Er moet een duidelijk systeem zijn om de resultaten van verschillende toezichtinstrumenten te meten en te analyseren. Hierbij dient een kanttekening geplaatst te worden dat het vergelijken van groepen ook kwetsbaar is. Daarnaast is het uitdagend om het effect van een los instrument ceteris paribus te bepalen.
- **Controlegroepen:** Het opstellen van controlegroepen is van groot belang om het effect van eigen handelen te meten, terwijl dit niet voor alle sectoren mogelijk is. Zulke controlegroepen helpen om causale relaties vast te stellen ter optimalisatie van de toezichtmix. De uitdaging bij het vaststellen van de controlegroep ziet op het definiëren van de groep voor het doen van een objectieve effectmeting.
- **Internationale samenwerking:** Uitwisseling van gegevens met buitenlandse belastingautoriteiten voorkomt niet alleen grensoverschrijdende belastingontduiking maar levert ook het op voor de toepassing van dit raamwerk. Internationale overeenkomsten, zoals de DAC7-richtlijn, ondersteunen deze samenwerking, maar het is essentieel om deze samenwerking nog te versterken.

5.3. Fase 3: Afweging van belangen en bepaling risicobereidheid

- **Politieke en economische overwegingen:** De politieke wil en bereidheid van beleidsmakers zijn cruciaal voor effectief toezicht. Daarnaast moeten de economische gevolgen van toezicht zorgvuldig worden overwogen. Dit omvat het balanceren van de noodzaak voor strikte handhaving met de potentiële impact op het bedrijfsleven en economische groei.
- **Verschillen in opdrachten tussen vakinhoudelijke experts en uitvoerders** kunnen leiden tot tegengestelde belangen en verminderde effectiviteit. Dat belangen verschillen heeft ook een logisch nut: Het mag schuren om te glanzen, waarbij glanzen gezien wordt als het behalen van de doelstellingen van de Belastingdienst als geheel. Echter, de relatie tussen verschillende opdrachten is vaak niet zichtbaar, wat begrip en samenwerking bemoeilijkt. Capaciteitsproblemen kunnen ook ontstaan in de uitvoering. De uitvoering van een kantoortoets of boekenonderzoek wisselen sterk qua doorlooptijd, wat effect heeft op inzetbare capaciteit.

5.4. Uitdagingen die niet fasespecifiek zijn

De Belastingdienst kan mogelijk tegen onderstaande uitdagingen aanlopen:

- **Transparantie en vertrouwen:** Het opbouwen van transparantie en vertrouwen tussen de Belastingdienst en belastingplichtigen is cruciaal. Gebeurtenissen van de afgelopen jaren hebben het vertrouwen verminderd, wat de nalevingsbereidheid negatief kan beïnvloeden. Het herstel van vertrouwen vraagt om transparantie communicatie, evenals maatregelen die laten zien dat de Belastingdienst handelt in het belang van de belastingplichtigen. De uitdaging voor de Belastingdienst ziet op het opbouwen van de vertrouwensrelatie met de belastingplichtige. Door externe factoren die betrekking hebben op een klein deel van de belastingplichtigen, kan het vertrouwen van de belastingplichtige in brede zin gedaald. Enerzijds kan de voorgestelde dialoog uit het raamwerk een startpunt zijn voor het verder opbouwen van de relatie met belastingplichtigen en een transparante informatiepositie van de Belastingdienst over

haar eigen handhaving en toezicht. Anderzijds kan het huidige vertrouwen in de Belastingdienst een obstakel zijn voor het opstarten van deze dialoog¹⁰.

- **Openbaarheid van informatie:** Duidelijke, eerlijke en transparante communicatie over wet- en regelgeving en toezichtresultaten schept duidelijkheid over verwachtingen van de burger en resultaten van het toezicht. Dit helpt belastingplichtigen om beter te begrijpen wat er van hen verwacht wordt en waarom bepaalde handhavingsmaatregelen worden genomen. De Belastingdienst is echter niet de enige belanghebbende in het opstellen van nieuwe wetgeving, maar is wel verantwoordelijk voor de uitvoering en handhaving hiervan. Bewustwording van de effecten van nieuwe wetgeving op de uitvoering kunnen bijdragen aan het behalen van betere resultaten in de compliance van belastingplichtigen, doordat duidelijke communicatie de verwachtingen duidelijk managet. De uitvoeringstoetsen vormen hierin een oplossing, waarin de Belastingdienst de effecten van compliance mee.
- **Flexibiliteit, aanpassingsvermogen:** Het vermogen van de Belastingdienst om snel te reageren op veranderingen in de omgeving, zoals nieuwe wet- en regelgeving, economische schommelingen, en technologische innovaties, is cruciaal. Dit vraagt om een adaptieve uitvoerings- en handhavingsmix en een cultuur van voortdurende verbetering. Om voor elke aanpassing eerste een afweging te maken inclusief een ex ante analyse zoals dit raamwerk voorstelt kan wringen qua doorlooptijd.
- **Terminologie:** Hoewel in hoofdstuk 6 een aanzet wordt gegeven hoe het raamwerk geïntegreerd kan worden met de Leidraad, is er afstand tussen werkwijze en terminologie in het raamwerk en de huidige praktijk binnen van de Belastingdienst. Samenvoegen van het raamwerk en de huidige praktijk, daar waar dat nu mogelijk is, kan een eerste stap vormen in de oplossing.
- **Abstractie en karakter van verkenning:** Het hier beschreven raamwerk kent – ondanks de illustratie voor de horecabranche - nog steeds een hoge mate van abstractie. Hoewel dat logisch is omdat het conform de onderzoeksvraag slechts om een verkenning gaat, is het lastig om het raamwerk echt op waarde te schatten zonder het eerst in werking te hebben gezien.
- **Capaciteit en expertise.** De interne capaciteit en expertise in het werken met dit raamwerk zijn nog beperkt. Het opbouwen en onderhouden van interne capaciteit en expertise binnen de Belastingdienst is überhaupt een voortdurende uitdaging. Dit omvat zowel het trainen van medewerkers, het implementeren van geavanceerde technologieën, en het ontwikkelen van methodologieën om toezicht effectief en efficiënt uit te voeren.

De bovenstaande uitdagingen onderschrijven de complexiteit van handhavingstaken die uitgevoerd worden door de Belastingdienst. Binnen een ingewikkeld stakeholderveld en vanuit de verantwoordelijkheid voor een grote groep belastingplichtigen is het bepalen van een mix die leidt tot een acceptabel minimaalniveau van toezicht ook met dit raamwerk nog steeds uitdagend.

¹⁰ Onderzoek effecten FSV Toeslagen (PwC 2021). Fiscale Monitor 2023. In gesprek over herstel, 2024.

6. Vervolgstappen

Uit veel van de interviews blijkt dat de in deze verkenning beschreven raamwerk kansrijk zou kunnen zijn, maar dat het raamwerk pas echt doorleefd kan worden door middel van praktische toepassing. Dit kan in eerste instantie binnen een kleine scope plaatsvinden, zodat eerst verder onderzocht kan worden welke van de in het vorige hoofdstuk beschreven uitdagingen in de praktijk het grootst zijn en in hoeverre het raamwerk inpasbaar is in een organisatie als de Belastingdienst. Pas als zo'n eerste toepassing tot succes leidt, is het verstandig om te gaan werken aan vergroting van het draagvlak voor het – eventueel aangescherpte – raamwerk. Ook kan in zo'n eerste toepassing in detail bekeken worden wat de overlap en de verschillen zijn met de huidige manier van het vaststellen van het uitvoerings- en handhavingsbeleid. Dit kan leiden tot een verdere integratie van de in dit rapport beschreven aanpak met de huidige aanpak. Aandacht voor veranderkunde is daarin cruciaal. In dit hoofdstuk bespreken we achtereenvolgens de toepassing binnen een kleine, nog te kiezen scope, de integratie met de huidige aanpak en de veranderkundige kant. We sluiten af met een samenvatting van de vervolgstappen.

6.1. Pilot

Op basis van de interviews ontstaat het beeld dat MKB de geschiktste uitvoeringsdirectie is om het nieuwe raamwerk kleinschalig uit te proberen. De reden hiervoor is dat binnen MKB geldt dat de heterogeniteit tussen doelgroepen groot is, maar één doelgroep homogeen genoeg – anders dan bijvoorbeeld bij GO – kan zijn om met één mix aan toezichtinstrumenten te kunnen werken. Tegelijkertijd is zo'n doelgroep ook groot genoeg om effecten van verschillende vormen van toezicht te kunnen voorspellen aan de hand van literatuuronderzoek en experimenten maar – anders dan bijvoorbeeld bij Particulieren – ook weer niet zo groot dat de term pilot niet meer gerechtvaardigd zou zijn. Welke MKB doelgroep het geschiktst is hangt ook af van het enthousiasme en de innovatiedrang van de direct betrokkenen, want zij fungeren als frontrunners. Als het raamwerk zich – na eventuele aanscherping en aanpassing – eenmaal heeft bewezen zijn minder enthousiasme en innovatiedrang nodig om het raamwerk verder uit te rollen.

6.2. Integratie van de twee raamwerken

De Leidraad Handhavingsregie ("Leidraad") en het Raamwerk ter bepaling van een acceptabel minimumniveau toezicht ("Raamwerk") bevatten overeenkomsten, maar ook aanvullende elementen. Zo richten beide benaderingen zich op het duidelijk definiëren van de scope en het identificeren van doelgroepen en belastingmiddelen waarop toezicht en handhaving mogelijk zijn. De Leidraad doet dit vanuit de vraag of sprake is van een nalevingsvraagstuk waarin getracht wordt antwoord te geven op het middel waarop gecontroleerd wordt en welke belastingplichtigen onderworpen zijn aan het middel. In dit kader komt de identificatiestap van de Leidraad overeen met de scopebepaling in de eerste stap van het Raamwerk, die leidt tot een gedetailleerde definitie van de scope voor toezicht. Daarnaast leggen zowel de Leidraad als het Raamwerk de nadruk op het analyseren van de effecten van toezicht, zoals gelijkheid van behandeling en doelmatigheid, om zo een beter begrip van de oorzaken van non-compliance te ontwikkelen en gerichte interventies mogelijk te maken.

Verder komen beide benaderingen overeen in hun focus op het gebruik van analyses en modellen om een mix van handhavinginstrumenten te definiëren en te valideren. In de Leidraad wordt bijvoorbeeld een analyse uitgevoerd om te bepalen welke instrumenten aansluiten bij de bereidheid en het vermogen van burgers en bedrijven om regels na te leven. Zowel de Leidraad als het Raamwerk onderstrepen het belang van een feedbackmechanisme waarbij strategische dialogen met belanghebbenden en evaluaties worden gebruikt om uitvoerings- en handhavingsmixen continu aan te passen en te verbeteren.

Toch biedt het Raamwerk een aantal aanvullingen op de Leidraad. Zo omvat het Raamwerk de ontwikkeling van een simulatiemodel dat de effecten van verschillende mixen van handhavingsinstrumenten onder diverse scenario's berekent en valideert. Ook voegt het Raamwerk een gestructureerde aanpak toe voor het ontwikkelen van scenario's die rekening houden met externe factoren die de effectiviteit van handhaving beïnvloeden. Dit maakt het mogelijk om te anticiperen op ontwikkelingen buiten de controle van de organisatie, een aspect dat in de Leidraad minder expliciet wordt behandeld. Daarnaast biedt het Raamwerk een manier om effecten van handhaving in de ogen van verschillende belanghebbenden onderling vergelijkbaar te maken en af te wegen. Dit maakt dat de impact van verschillende mixen van handhavingsinstrumenten in één oogopslag duidelijk wordt en dat de selectie van de uiteindelijke mix zowel op data als dialoog gebaseerd is.

Ook legt het Raamwerk nadruk op een veranderkundige benadering, door te erkennen dat de manier waarop toezicht- en handhavingsbeleid gepresenteerd wordt bijdraagt aan de acceptatie en implementatie ervan. Dit aspect is minder expliciet aanwezig in de Leidraad, maar kan van groot belang zijn bij het doorvoeren van beleidsveranderingen. Verder biedt het Raamwerk een meer flexibele en adaptieve benadering, waarin de mix van handhavingsinstrumenten continu wordt aangepast op basis van nieuwe inzichten en veranderende omstandigheden. Dit draagt bij aan een robuustere aanpak van toezicht en handhaving.

Hoewel de Leidraad en het Raamwerk elkaar overlappen in hun aanpak van het definiëren van toezichtdoelen en het belang van feedback, vormt het Raamwerk een verdiepende en meer adaptieve aanvulling.

6.3. Initiëren van de verandering

Na een eerste succesvolle toepassing van het raamwerk binnen een kleine scope, kan de aandacht zich richten op het initiëren van een veranderingsbeweging om het draagvlak voor het werken met het – eventueel aangepaste – raamwerk binnen de organisatie en daarbuiten te vergroten. Deze stappen worden van inherent ondersteund door het raamwerk gegeven de combinatie van data en dialoog daarin en dragen bij aan een geleidelijke en effectieve integratie van nieuwe uitvoerings- en handhavingsmixen.

Allereerst kan het opzetten van een veranderingsteam overwogen worden. Dit team zou bestaan uit medewerkers uit verschillende afdelingen, zoals senior management, operationele teams en specialisten, om een breed scala aan inzichten en expertise te waarborgen. Een veranderingsleider met voldoende autoriteit en invloed kan aangesteld worden om de verandering te sturen en te ondersteunen. Dit team zou fungeren als het kloppende hart van de veranderingsinspanning.

Het creëren van bewustzijn en draagvlak kan een belangrijke eerste stap zijn. Een communicatieplan kan ontwikkeld worden om de noodzaak en voordelen van de verandering duidelijk te maken aan alle medewerkers. Dit kan gebeuren door middel van nieuwsbrieven, het intranet en townhall-bijeenkomsten. Workshops en bijeenkomsten kunnen georganiseerd worden om medewerkers te informeren en hen de kans te geven hun zorgen en vragen te uiten. Open discussies zouden kunnen helpen om eventuele weerstand te verminderen en een gevoel van betrokkenheid te creëren.

Een grondige analyse van de huidige situatie kan noodzakelijk zijn om de verandering goed te kunnen sturen. Een uitgebreide nulmeting van de bestaande toezichtprocessen en prestaties zou kunnen worden uitgevoerd. Hierbij kan gebruik worden gemaakt van bestaande data en interviews met sleutelfiguren binnen de organisatie. Op basis van deze analyse zouden de belangrijkste knelpunten geïdentificeerd en gedocumenteerd kunnen worden, zodat een solide basis voor verdere actie ontstaat.

Met deze inzichten kunnen duidelijke, meetbare doelstellingen worden geformuleerd en indicatoren worden ingericht aan de hand waarvan de voortgang gemonitord kan worden. Denk daarbij aan: Op welk percentage van de opbrengsten per belastingmiddel vindt een via het raamwerk afgewogen

vorm van toezicht plaats? Een gedetailleerd stappenplan met tijdslijnen, mijlpalen en verantwoordelijkheden kan ontwikkeld worden. Dit plan zou flexibel genoeg moeten zijn om aanpassingen mogelijk te maken indien nodig. Door betrokkenheid van experts bij het opstellen van doelstellingen en plannen, zou worden gezorgd dat deze realistisch en haalbaar zijn.

Het raamwerk kan helpen bij het verbeteren van de veranderingsbeweging door een systematische en gestructureerde aanpak te bieden. Het kan ondersteunen bij het definiëren van duidelijke doelstellingen, het identificeren van knelpunten en het ontwikkelen van uitvoerings- en handavingsmixen om deze knelpunten aan te pakken. Door consistent gebruik te maken van het raamwerk, kan de organisatie een robuuste en adaptieve uitvoerings- en handavingsmix ontwikkelen die effectief, rechtvaardig en efficiënt is.

6.4. Samenvatting vervolgstappen

Het lijkt verstandig om te beginnen met het uitvoeren van een pilot met het raamwerk ter bepaling van een acceptabel minimumniveau toezicht binnen één of meerdere doelgroepen binnen MKB, omdat daarin relatief veel heterogeniteit is tussen doelgroepen, maar de doelgroepen nog steeds te groot zijn voor een individuele toezichtsaanpak. De landbouw- of horecabranche zouden zich bij uitstek kunnen lenen voor zo'n pilot, omdat daar al initiatieven lopen die zowel gericht zijn op de aanpassing van de huidige mix van toezichtinstrumenten en als de mogelijkheid bieden om het raamwerk te versterken en andersom. Zo'n pilot zal zorgen voor:

- meer praktische handvatten om met het raamwerk aan de slag te kunnen gaan binnen de Belastingdienst,
- verdere integratie van het raamwerk met de huidige aanpak binnen de Belastingdienst om het uitvoerings- en handavingsbeleid continu te verbeteren,
- eventuele aanpassing en aanscherping van het raamwerk voor toepassing binnen de Belastingdienst, en
- een startpunt van een veranderkundige aanpak om het draagvlak voor een – gedeeltelijk – nieuwe manier van werken te versterken, zowel bij de Belastingdienst als stakeholders.

7. Conclusies

De conclusie van het onderzoek naar het minimumniveau van toezicht bij de Belastingdienst benadrukt dat het vaststellen van een absoluut minimumniveau van toezicht niet haalbaar is. Dit komt doordat het bepalen van een minimumniveau van toezicht, zeker bij complexe organisaties zoals de Belastingdienst, niet alleen een technische of operationele kwestie is, maar vooral een intrinsiek politieke vraag. Het is een vraagstuk dat afhangt van normatieve keuzes over risicobereidheid en strategische prioriteiten. Bovendien hebben eerdere onderzoeken, waaronder die van de Algemene Rekenkamer en professor Roland Speklé, aangetoond dat het vaststellen van een minimumniveau toezicht problematisch is door de heterogeniteit van belastingmiddelen en belastingplichtigen, en de inherente politieke aard van het vaststellen van acceptabele restrisico's

In plaats van een absoluut minimumniveau, richt dit rapport zich op de mogelijkheid van een "acceptabel minimumniveau toezicht." Dit betekent dat er gezocht wordt naar een toezichtsniveau dat, gegeven de beschikbare middelen en strategische prioriteiten van de Belastingdienst, door de maatschappij als aanvaardbaar wordt beschouwd. Hierbij wordt uitgegaan van een raamwerk waarin verschillende uitvoerings- en handavingsmixen worden geëvalueerd op hun maatschappelijke waarde. Dit raamwerk bestaat uit drie fasen:

1. Dialoog met belanghebbenden en experts: Hierin worden de verschillende effecten van toezicht en handhaving in kaart gebracht, zoals rechtvaardigheid, doelmatigheid en doorlooptijd. De belastingplichtigen, toezichthouders en andere belanghebbenden leveren input voor het bepalen van de gewenste uitkomsten van uitvoering en handhaving.
2. Bureau- en veldonderzoek In deze fase worden de verschillende uitvoerings- en handavingsmixen gemodelleerd en gesimuleerd om te bepalen wat de effecten zijn van verschillende uitvoerings- en handavingsmixen op de nalevingsbereidheid, robuustheid en correctie van foutieve aangiften. Daarbij wordt gebruikgemaakt van data en onderzoek naar eerdere effecten van toezicht.
3. Afweging van strategische prioriteiten en risicobereidheid: Uiteindelijk wordt de keuze voor een uitvoerings- en handavingsmix gemaakt op basis van een gewogen afweging van de belangrijkste effecten en de risico's die men bereid is te accepteren. Hierbij worden scenario's geschetst die inzicht geven in de mogelijke uitkomsten van verschillende uitvoerings- en handavingsmixen onder uiteenlopende omstandigheden

De uitdagingen waarmee de Belastingdienst te maken heeft bij het vaststellen van een acceptabel minimumniveau toezicht, zijn talrijk. Deze variëren van het verkrijgen van voldoende kwalitatieve data, het omgaan met de inherente onzekerheden in de organisatie en haar omgeving, tot het vaststellen van duidelijke maatschappelijke doelstellingen die kunnen worden gebruikt voor het evalueren van verschillende uitvoerings- en handavingsmixen.

Concluderend biedt het rapport waardevolle inzichten en een methodologisch raamwerk waarmee de Belastingdienst een proportioneel en effectief toezichtbeleid kan ontwikkelen. Dit beleid moet niet alleen gebaseerd zijn op het minimaliseren van risico's, maar ook op het optimaliseren van maatschappelijke waarde. De Belastingdienst zal echter voortdurend keuzes moeten maken tussen doeltreffendheid, doelmatigheid en rechtvaardigheid, en deze afwegingen moeten worden afgestemd met de maatschappij en andere belanghebbenden

A. Geraadpleegde bronnen

A.1. Gevoerde gesprekken

- Corporate Dienst Vaktechniek/Handhaving op 18 maart 2024
- Directie Midden- en Kleinbedrijf op 2 april 2024
- Corporate Dienst Vaktechniek/Handhaving op 25 april 2024
- Innovatie & Strategie op 7 mei 2024
- Directie Particulieren op 19 juni 2024
- Corporate Dienst Vaktechniek/Handhaving op 3 juli 2024
- Directie Grote Ondernemingen op 5 juli 2024
- Directie Midden- en Kleinbedrijf op 8 juli 2024

A.2. Geraadpleegde documentatie

Literatuur

Braithwaite, V., & Braithwaite, J. (2001). An evolving compliance model for tax enforcement. In Crimes of privilege: Readings in white-collar crime. Oxford University Press.

L. van Liere MSc, Prof.dr. J. de Swart MBA en prof. dr. mr. E.C.J.M. van der Hel-van Dijk RA (2023). De daad bij het woord voegen: Fraudedetectiesystemen verankeren in de strategie van een organisatie. TaxTech 2023/6

Roobeek, A; Swart, J.J.B. de; Plas, M. van der (2023). Responsible Business Decision Making – Strategic Impact Through Data and Dialogue. KoganPage.

<https://www.koganpage.com/product/responsible-business-decision-making-9781398612280>

Kamerstukken/officialle overheidsdocumentatie

Tweede Kamer, Aanbiedingsbrief staatssecretaris Belastingdienst Jaarplan 2022, vergaderjaar 2021–2022, 31 066, nr. 958

Belastingdienst, Beleidskader handavingsplannen (uitvoerings- en handavingsstrategie Belastingdienst), september 2022

Belastingdienst, Format handhaving, 2022

Belastingdienst, Handreiking handavingsplannen, 2022

Belastingdienst, Leidraad Handavingsregie, 2022

Belastingdienst, Rapport analyse niveau toezicht achteraf, 2022

Belastingdienst, Toelichting op het Handavingsbeleid van de Belastingdienst 2022, 2022

Belastingdienst, Uitvoering, dienstverlening, toezicht en opsporing bij de Belastingdienst. 2022

Belastingdienst, Vaktechnische verdieping evalueren, 2022

Belastingdienst, Vaktechnische verdieping gedragsanalyse, 2022

Belastingdienst, Jaarplan 2023, 2023

Belastingdienst, Jaarplan 2024, 2024

PwC, Onderzoek Misbruik Uitwonendenbeurs

<https://www.rijksoverheid.nl/documenten/rapporten/2024/03/01/eindrapport-pwc-rapportage-onderzoek-misbruik-uitwonendenbeurs>, 2024

Europese Commissie, [EU AI Act: first regulation on artificial intelligence](#), 2024

B. Werkwijzen bij andere organisaties en eerder onderzoek

Deze bijlage bevat een selectie van kerninzichten uit eerdere onderzoeken naar en rondom uitvoerings- en handavingsmixen toegelicht en uiteengezet. Hierbij wordt allereerst verkend op welke wijze nationale en internationale toezichthoudende instanties omgaan met toezicht op basis van het rapport analyse niveau toezicht achteraf en rapportage van het OECD. Hierna worden belangrijke inzichten uit de literatuur opgehaald met betrekking tot determinanten die van belang zijn bij het inrichten van toezicht bij de Belastingdienst.

Rapport analyse niveau toezicht achteraf

Vergelijkbare organisaties Nederland

Het onderzoek naar het niveau van toezicht achteraf omvatte het uitvoeren van een enquête bij vergelijkbare organisaties en belastingdiensten, waaronder de Autoriteit Financiële Markten (AFM), de Nederlandsche Bank (DNB), de Dienst Uitvoering Onderwijs (DUO), de Inspectie SZW (ISZW), de Nederlandse Voedsel- en Warenautoriteit (NVWA), de Dienst Wegverkeer (RDW), en de Sociale Verzekeringsbank (SVB).

Binnen de uitvoerings- en handavingsmix leggen alle organisaties over het algemeen de nadruk op compliance van burgers en bedrijven, met een focus op toezicht vooraf. De samenwerking met andere en buitenlandse instanties wordt benadrukt, en risicoselectie wordt veel gebruikt om de inzet van toezicht te bepalen, waarbij impact breder wordt gezien dan alleen financieel.

Toezicht vooraf wordt gezien als de eerste stap om fouten te voorkomen, terwijl toezicht achteraf dient om eventuele overtredingen te detecteren die ondanks het voorafgaande toezicht zijn ontstaan. Voorlichting over rechten en plichten is de meest gebruikte vorm van toezicht vooraf, terwijl gedragsbeïnvloeding een belangrijk preventief middel is.

Het toezicht achteraf kan worden opgesplitst in kantoor- of veldcontroles, maar sommige grote uitvoeringsorganisaties hebben een daling van toezicht achteraf ervaren, voornamelijk tussen 2016 en 2018, als gevolg van een verschuiving naar toezicht vooraf.

Er is geen vastgesteld acceptabel niveau van toezicht, waarbij de meeste organisaties aangeven dat de financiering deels of volledig bepalend is. Aanbevelingen zijn gedaan om eerst een acceptabel risiconiveau vast te stellen en vervolgens de benodigde capaciteit vrij te maken voor toezicht.

Verschuivende enquêtes hebben de afgelopen jaren meer middelen gekregen voor toezicht. Nieuwe kostenkaders zijn overeengekomen met de AFM en DNB, waarbij gekeken is naar het risiconiveau, nieuwe wetgeving en ontwikkelingen, met de verplichting voor toezichthouders om deze kaders de komende jaren te volgen. Eventuele intensiveringen moeten door de toezichthouders zelf worden ingepast.

Buitenlandse belastingdiensten

Resultaten zijn verzameld van belastingdiensten in verschillende landen, waaronder Australië, België, Canada, Denemarken, Duitsland, Estland, Ierland, Letland, Litouwen, Noorwegen, Oostenrijk, Portugal, Spanje, het Verenigd Koninkrijk, de Verenigde Staten en Zweden. Uit deze resultaten komen gemeenschappelijke ontwikkelingen naar voren.

Een van deze gemeenschappelijke ontwikkelingen is dat aandacht wordt besteed aan het bevorderen van compliance vóór het indienen van de aangifte, zoals het invullen van gegevens vooraf (zoals in Portugal) of online dienstverlening (zoals in Zweden en Duitsland). Veel landen hanteren verschillende strategieën voor risicomanagement, afhankelijk van belastingbetalers en hun omvang.

Bijvoorbeeld een aparte benaderingen voor grote ondernemingen, het MKB, soms ook voor zeer vermogende particulieren (zoals in de VS en het VK).

Daarbij is bij meerdere landen het horizontaal toezicht geïmplementeerd. Bedrijven die onder horizontaal toezicht staan kunnen potentiële discussiepunten vooraf identificeren en oplossen in overleg met de desbetreffende belastingdienst. Een voorbeeld hiervan is het compliance assurance programme in de VS.

Tevens meten sommige landen, waaronder het VK, Portugal, Denemarken, Noorwegen, België en Spanje, de effectiviteit en efficiëntie van hun belastingstelsels door middel van geavanceerde analytische studies, specifieke kwantitatieve onderzoeken of strategische en resultaatsindicatoren. In enquêtes blijkt dat sommige belastingdiensten een afnemend niveau van toezicht achteraf ervaren, door veranderingen in het risicomanagement als reactie op verschillende inschattingen van voorzienbare ontwikkelingen.

OECD rapport

De OECD's Tax Administration Series (TAS 2023) benadrukt de omvang en complexiteit van belastingadministratie. Belastingdiensten verzamelen gezamenlijk netto-inkomsten van €13,4 biljoen en hebben ongeveer 1,7 miljoen medewerkers in dienst. Ze behandelen de belastingzaken van meer dan 900 miljoen belastingbetalers en hebben meer dan 450 miljoen contactmomenten per telefoon, in persoon, via e-mail of papier en meer dan 2,3 miljard contacten via onlinekanalen. Deze activiteiten worden uitgevoerd met een gecombineerd operationeel budget van ongeveer €95 miljard, wat hun efficiëntie weerspiegelt.

Het rapport biedt ook inzicht in initiatieven van belastingdiensten om compliance te bevorderen en de lasten voor belastingbetalers te verminderen, zoals 24/7 beschikbaarheid, digitalisering van identiteitsverificatie, decentralisatie van diensten en preventieve risicobeheersing. Internationale samenwerking blijft essentieel in deze tijd van veranderingen en digitalisering van de economie.

Braithwaite

"Taxing Democracy" van Valerie Braithwaite¹ is een boek dat onderzoekt hoe belastingssystemen de democratie kunnen bevorderen of ondermijnen. Braithwaite betoogt dat belastingssystemen een cruciaal aspect zijn van democratisch bestuur, omdat ze de perceptie van burgers over rechtvaardigheid, vertrouwen in de overheid en bereidheid om belastingwetten na te leven vormgeven.

Ze onderzoekt verschillende belastingbeleidsmaatregelen en -praktijken over de hele wereld, waarbij ze hun impact op democratische waarden zoals gelijkheid, participatie en verantwoording belicht. Door middel van casestudies en empirisch onderzoek laat Braithwaite zien hoe progressieve belastingheffing, transparantie en betrokkenheid van burgers de democratie kunnen versterken, terwijl regressieve belastingheffing, belastingontduiking en corruptie deze kunnen ondermijnen.

Het onderzoek biedt inzichten in de complexe relatie tussen belastingheffing en democratie en dringt er bij beleidsmakers op aan belastingssystemen te ontwerpen die sociale rechtvaardigheid en democratische idealen bevorderen.

Jongere individuen vertonen over het algemeen minder belastingnaleving dan oudere personen. Dit kan zich uiten in het niet volledig aangeven van inkomsten, het werken in de informele economie, het overdreven claimen van aftrekposten en het hebben van openstaande schulden bij de belastingdienst. Hoewel sommigen aanvankelijk dachten dat dit te wijten was aan een gebrek aan capaciteit of een minder ontwikkelde pro-belastinghouding, blijkt uit onderzoek dat de relatie complexer is.

Verschillende factoren spelen een rol bij de lagere belastingnaleving onder jongere mensen. Capaciteitsgerelateerde aspecten, zoals zelfvertrouwen, kennis van belastingverplichtingen en steun van sociale kringen, werden in eerste instantie gezien als verklaringen voor dit verschil op basis van

leeftijd. Echter, blijkt uit onderzoek dat deze factoren niet volledig de lagere naleving onder jongeren verklaren. In plaats daarvan lijken attitudes ten opzichte van belastingen een grotere rol te spelen. Een zwakkere morele norm met betrekking tot belastingbetaling, minder vertrouwen in eerlijke belastingadviseurs en een verminderde angst voor de gevolgen van niet-naleving zijn allemaal geassocieerd met lagere belastingnaleving. Bovendien zijn jongere individuen meer geneigd om de relevantie en prioriteit van belastingbetaling in twijfel te trekken, wat kan resulteren in een grotere afstand tot belastingautoriteiten.

Onderzoek wijst uit dat de invloed van deze attitudes op belastingnaleving groter is dan eerder gedacht. Hierbij blijken capaciteitsgerelateerde factoren, zoals kennis en vaardigheden, minder doorslaggevend te zijn bij het voorspellen van belastingontduiking. Deze bevindingen suggereren dat het aanpakken van morele en attitudele aspecten van belastingnaleving essentieel kan zijn om de naleving, met name onder jongere individuen, te verbeteren.



Figuur 8. Voorbeeld van compliance model (Braithwaite, 2003)

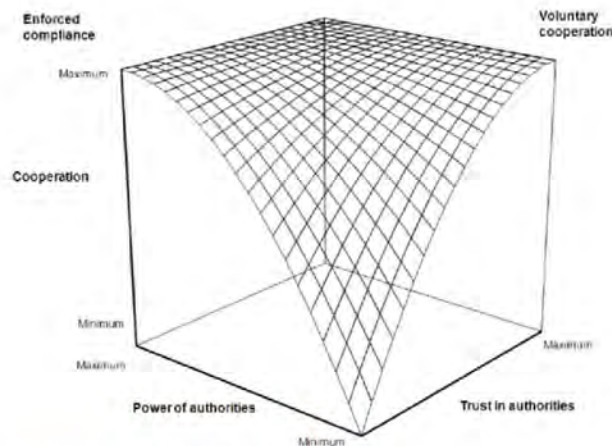
Kirchler

"The economic psychology of tax behavior" door Erich Kirchler² is een uitgebreide verkenning van de psychologische factoren die van invloed zijn op de naleving van belastingwetten door individuen. In het artikel duikt Kirchler in de complexe interactie tussen economische prikkels, sociale normen en psychologische processen bij het vormgeven van het gedrag van belastingbetalers.

Kirchler betoogt dat hoewel economische theorieën traditioneel de rol van financiële prikkels bij belastingnaleving benadrukken, psychologische factoren zoals morele overtuigingen, percepties van rechtvaardigheid en sociale normen ook een belangrijke rol spelen. Hij benadrukt onderzoeksresultaten die suggereren dat de bereidheid van mensen om belastingen te betalen wordt beïnvloed door hun perceptie van de eerlijkheid van het belastingstelsel, het vertrouwen in overheidsinstellingen en morele oordelen over de legitimiteit van belastingheffing.

Het artikel onderzoekt ook verschillende psychologische theorieën en modellen die helpen bij het verklaren van belastingnalevingsgedrag, waaronder de sociale identiteitstheorie, afschrikkingstheorie en het glijdende schaal-model. Kirchler benadrukt het belang van het overwegen van psychologische factoren bij het ontwerpen van belastingbeleid en uitvoerings- en handhavingsmixen, evenals bij het bevorderen van vrijwillige naleving onder belastingbetalers.

Concluderend biedt "The economic psychology of tax behavior" waardevolle inzichten in de multidimensionale aard van belastingnaleving en benadrukt de noodzaak van interdisciplinaire benaderingen die economische en psychologische perspectieven integreren om belastingontduiking effectief aan te pakken en belastingnaleving te bevorderen.



Figuur 9. Kirchler's framework (Kirchler et al., 2008)

Bomb crater effect

Belastingnaleving is een complex terrein waarbij zowel directe als langetermijneffecten een rol spelen. Het "bombcrater-effect"¹¹ is zo'n fenomeen dat zich voordoet na belastingcontroles, waarbij belastingbetalers direct na een controle hun naleving verminderen. Dit effect kan zowel op korte als op lange termijn van invloed zijn op het nalevingsgedrag. Op korte termijn kan de schok van een controle leiden tot een afname van naleving, vergelijkbaar met soldaten die schuilen in kraters na een explosie in oorlogstijd. Op lange termijn kan het effect van een controle echter variëren, waarbij sommige belastingbetalers hun gedrag aanpassen en anderen juist niet. Dit kan worden verklaard door verschillen in perceptie van kans en emotionele reacties op controles. Het begrip van het bombcrater-effect biedt waardevolle inzichten in de complexe factoren die van invloed zijn op belastingnaleving en kan helpen bij het ontwikkelen van effectievere strategieën om belastingontduiking tegen te gaan.

Recent onderzoek heeft de focus gelegd op het identificeren van waarschijnlijkheidsverwachtingen onder belastingbetalers en het effect van persoonlijke controle-ervaringen op nalevingsgedrag. Het onderzoek toont aan dat belastingbetalers die zelf gecontroleerd zijn, vaak een veranderd nalevingsgedrag vertonen, terwijl audits op andere belastingbetalers minder impact hebben op naleving. Dit is een vorm van een "bombcrater-effect", waarbij belastingbetalers, vooral die met een Bayesiaanse benadering, direct na hun eigen controle een sterke afname in naleving vertonen. De perceptie van een belastingcontrole als een persoonlijk risico blijkt de beslissingen van belastingbetalers aanzienlijk te beïnvloeden, terwijl controles op anderen minder impact hebben op het nalevingsgedrag.

Het onderzoek naar het bombcrater-effect heeft belangrijke implicaties voor beleidsmakers en de methodologie van laboratoriumonderzoeken op het gebied van belastingnaleving. Het benadrukt het belang van een zorgvuldig experimenteel ontwerp, inclusief controlegroepen van niet-gecontroleerde belastingbetalers, om het ware effect van controles op naleving te beoordelen. Deze bevindingen dragen bij aan een beter begrip van belastingbetalersgedrag en de rol van controles bij nalevingsbeslissingen, wat van cruciaal belang is voor het ontwikkelen van effectief beleid om belastingontduiking tegen te gaan.

De uitwerking in praktijk ziet voornamelijk op het gedrag van een belastingplichtige na controle. Dit gedrag is minder compliant dan het gedrag voor controle. Als beleidsmaker is het van belang om dit effect scherp te hebben en hierop de uitvoerings- en handavingsmix op aan te passen.

¹¹ Brandhorst, Sebastian & Kluge, Annette. (2018). The effect under the influence of audit feedback: Now you see me, now you don't. *Safety Science*. 110. 10.1016/j.ssci.2018.07.024.

Overeenkomsten

De samenhang tussen de onderzoeken van Braithwaite, Kirchler en het effect ligt in hun gemeenschappelijke focus op de factoren die belastingnaleving beïnvloeden, hoewel ze elk verschillende perspectieven en benaderingen bieden. De onderzoeken bevatten de volgende raakvlakken:

Morele en attitudinale invloeden: Zowel Braithwaite als Kirchler en het effect benadrukken de invloed van morele overtuigingen en attitudes op belastingnaleving.

Perceptie van rechtvaardigheid en vertrouwen: De perceptie van rechtvaardigheid en vertrouwen in het belastingstelsel en de overheid speelt een centrale rol in alle drie de onderzoeken. De perceptie van de belastingbetaling speelt een rol in hoe compliant een belastingbetaler zich opstelt in het proces voorafgaand aan de aangifte en na de aangifte.

Complexiteit van gedragsfactoren: Alle onderzoeken erkennen de multidimensionale en complexe aard van belastingnaleving, die verder gaat dan economische prikkels alleen.

Hieruit kan geconcludeerd worden dat effectieve belastingnaleving niet alleen afhankelijk is van strikte controles en economische prikkels, maar ook sterk beïnvloed wordt door morele en psychologische factoren. Beleidsmakers zouden daarom belastingstelsels moeten ontwerpen die sociale rechtvaardigheid bevorderen, transparantie waarborgen en het vertrouwen van burgers in de overheid versterken om vrijwillige naleving te bevorderen. Daarnaast is het essentieel om rekening te houden met de emotionele en perceptuele reacties van belastingbetalers op controles om ongewenste effecten zoals het effect te voorkomen.

Uit de bovenstaande theoretische kaders wordt in dit onderzoek een kader opgesteld die de bovenstaande onderdelen van een goed werkend belastingstelsel ondervangt, wat resulteert in een optimaal resultaat. Beleidsmakers kunnen met dit kader compliance bevorderen door acht te nemen van de morele en psychologische factoren die van invloed zijn op het gedrag van de belastingbetaler. Daarnaast omvat het model de wijze waarop controle op compliance efficiënter in gezet kan worden.