This is an Accepted Manuscript of an article

published by Taylor & Francis Group in International Public Management Journal on 12/03/2024,

Available online : https://doi.org/10.1080/10967494.2024.2315183

The accountability of collaborative innovations in tax administration: A Dutch-US comparison

Dennis de Widt

Cardiff Business School

Cardiff University

Cardiff, UK

dewidtd@cardiff.ac.uk

Emer Mulligan

J.E. Cairnes School of Business and Economics

University of Galway

Galway, Ireland

emer.mulligan@universityofgalway.ie

The accountability of collaborative innovations in tax administration: A Dutch-US comparison

Abstract

Analysing a collaborative public management innovation in tax administration in the US and the Netherlands, this article investigates how differences in accountability systems affect senior public servants' *felt* accountability, which itself appears critical to understanding why innovations evolve differently over time, even when sharing similar starting conditions. The article makes three key contributions. First, employing elite interviews, findings highlight how institutional factors critically shape public servants' felt accountability, hereby expanding existing scholarship and its concentration on microlevel psychological determinants. Second, against expectations, findings illuminate that ministerial-based senior public servants may display a greater degree of autonomy in innovation design and implementation compared to agency-based officials. Third, the article demonstrates that the configuration of accountability systems can change significantly over time, thereby critically affecting the evolution of innovations. This underlines the relevance of taking a longitudinal research approach to analysing senior public servants' attitude to collaborative innovations.

Keywords: collaborative public management, felt accountability, tax administration, the Netherlands, US

Introduction

Public administration scholarship has produced valuable insights as to how institutional features such as constitutional structures affect discretion held by senior public servants (Aberbach and Rockman 2006; Raadschelders and Van der Meer 2014). Few studies however analyse how institutional features influence senior public servants' *use* of their autonomy. Many studies appear to take it for granted that senior public servants willingly employ their autonomy even when this would produce risks for the official involved (Ansell and Gash 2008). One institutional factor that has so far been largely overlooked but can be expected to importantly affect public servants' use of autonomy is the configuration of accountability arrangements. Senior public servants operate in a web of multiple accountability relationships, with actors perceived differently in terms of legitimacy and expertise to hold one another to account. The aggregate of interactions within accountability relationships provides key impetus to the sense of accountability as felt by senior public servants.

To investigate how the institutional context influences senior' public servants felt accountability, this study employs a comparative case study analysis of the implementation of a collaborative public management innovation. Typically, the goal of collaborative public management innovations is to engage state and non-state actors in the implementation of new, creative solutions to address problems that cannot be easily solved by single organisations (O'Leary et al. 2006). Senior public servants have been identified as potentially conducting a key role in initiating and implementing collaborative innovations. However due to the risk of failure related to such innovations (Bommert 2010; Wegrich 2019), senior public servants' behaviour in relation to collaborative innovations is likely to be impacted by the accountability public servants perceive to be linked to such innovations, which itself will be influenced by country specific configurations of accountability systems.

Consequently, a country-comparative and longitudinal analysis provides an important avenue to enhance our understanding as to how country specific configurations of accountability systems influence public servants' perceptions of accountability. This paper undertakes such an analysis, thereby illuminating the dissimilar evolution of collaborative innovations across countries. This occurs even when they have similar beginnings.

The specific collaborative public management innovation examined in this article, called cooperative compliance, originates in the domain of tax administration in two countries, the US and the Netherlands. This collaborative initiative entails special arrangements for the interaction between tax authorities and large corporate taxpayers, in most cases multinational enterprises (MNEs). The interaction is between sophisticated actors with high levels of technical and practical experience who are working in a dynamic and challenging regulatory environment. Cooperative compliance initiatives have often emerged following close interaction between government officials and business representatives with the objective to realize a more cooperative relationship. This relationship is expected to enhance tax compliance by mechanisms other than what has traditionally been the tax administration's dominant approach of strict control-and-punishment (Bronzewska 2016). The innovation puts strong emphasis on collaboration between tax authorities and corporate taxpayers with the expectation this generates efficiencies for both sides. While cooperative compliance has rapidly spread internationally, helped by the OECD's endorsement (OECD 2013), little attention has been given to understanding why these innovations have evolved differently across country settings.

This study makes several contributions to the literature. First, the study connects accountability scholarship with collaborative public management research, fields that despite their obvious associations have developed rather separately. Second, the study sheds light on how accountability arrangements relate to senior public servants' felt accountability and its subsequent influence on the evolution of collaborative innovations. Third, despite their size and importance, no studies have addressed collaborative innovations in the context of tax authorities, and this study is the first to investigate a major collaborative initiative in two advanced tax authorities. Fourth, most studies focus on the origins, generation, and initial adoption of collaborative innovation solutions (e.g., Li and Chen 2022; Van Buuren and

Loorbach 2009), but few studies identify and explain what occurs after innovations have been initiated. By extending the analysis beyond a single point in time, our study examines the phenomena diachronically, enabling us to identify senior public servants' behaviour and accountability dynamics during the trajectory of collaborative initiatives.

The central question that we seek to answer is: *how does the relationship between accountability systems and felt accountability by senior public servants impact the implementation of collaborative public management innovations?* Following existing theoretical work, we distinguish between bureaucratic, professional, legal, and political accountability systems (Romzek and Dubnick 1987). It is hypothesised that accountability pressures deriving from these accountability systems will differ depending upon the institutional context, which subsequently gives rise to country specific public servant perceptions of accountability.

The rest of the article is structured as follows. First, the article begins by reviewing the theory on accountability and collaborative public management. The subsequent section discusses the study's methods, followed by the Dutch and US case studies. For each case, accountability dynamics are discussed during the design, rollout and consolidation phase of the innovation. The article subsequently includes a comparative analysis of the two cases whilst the last section concludes the article.

Theory

Accountability, collaborative public management, and felt accountability

The accountability literature has long been dominated by conceptual and normative oriented studies, which is reflected in principal-agent theory (PAT) being the field's dominant conceptual framework (McCubbins and Schwartz 1984; Moe 1984). As PAT has found mixed support in empirical work, more recent accountability studies have sought to develop alternative conceptual frameworks that improve understanding on how accountability dynamics work in the 'real world' (Brandsma and Schillemans 2013). Of key importance in this regard is Bovens (2007) who provides a wider definition of accountability, being "understood as a communicative interaction between an actor/accountee (person or organization) and a forum/account-holder". This wider approach to analyse accountability dynamics has been drawn upon in several empirical studies, and a conceptual consensus has emerged emphasising the relational dimension of accountability such as the expectation of being held accountable (Bovens *et al.* 2014).

Simultaneous with a rise of empirical focused accountability scholarship has been the expansion of collaborative innovation research. To improve public sector efficiency and responsiveness, governments around the world have faced pressures to become more innovative, which they have often sought to achieve by developing new ways of working with actors outside the public sector. This development has instigated various theoretical frameworks, including public entrepreneurialism, collaborative governance, and collaborative public management (Agranoff and McGuire 2004; Ansell and Gash 2008; O'Leary *et al.* 2006; Torfing 2016). The latter approach, which is the focus of this article, concentrates on the collaboration between government and third parties during policy implementation (O'Leary *et al.* 2006).

The literature identifies several drivers and barriers for collaborative innovation including political support, resource availability, levels of interdependencies amongst actors, and the perceived likelihood of gains from participation (Torfing 2016). Whilst the role of leaders of public sector organizations as 'catalyst' and 'convenors' of cross-boundary collaboration is widely emphasised (Wegrich 2019), few empirical analyses exist that do in fact analyse what makes senior public servants engage with collaborative innovations. Similarly, the collaborative literature has paid scant attention to the relationship between the implementation of collaborative innovations and accountability, despite many scholars having highlighted accountability concerns in relation to collaborative innovations (Bommert 2010; Han 2020; Sørensen 2012). A recent attempt seeking to address the accountability gap in collaborative innovation scholarship is the concept of ‘felt accountability’, which focuses on how engagement by senior public servants with collaborative innovations is influenced by the extent to which an accountability forum is perceived as legitimate and qualified, and public servants expect to be held accountable (Overman 2021; Schillemans *et al.* 2021). The assumption is that an increase in felt accountability will result in public officials reflecting more broadly and thoroughly on the consequences when they exercise discretion, leading to better-informed decisions (Hall *et al.* 2017; Wille 2010).

As the few existing studies investigating ‘felt accountability’ predominantly draw on psychological insights and apply a micro level approach (Schillemans *et al.* 2021), we have limited understanding of how the institutional context in which senior public servants operate influences ‘felt accountability’. By comparing the role of senior public servants in the design and implementation of a collaborative public management innovation in different institutional contexts, we aim to enhance insight into the relationship between institutional conditions and accountability pressures as perceived by senior public servants. In a second step, we use these insights and identify to what extent they help to explain the dissimilar evolution of collaborative innovations across countries .

Institutional dimensions of felt accountability – an accountability systems’ approach

In modern public organisations officials tend to operate in a web of formal and informal accountability arrangements that manage (and constrain) the behaviour and diverse expectations of actors within the organisation (Romzek 2000; Romzek and Dubnick 1987). To capture the multifaceted nature of accountability dynamics impacting upon public decision making, Romzek and Dubnick (1987) identified four accountability systems each based on variations alongside two dimensions: (1) whether the ability to define and control expectations is held by an entity inside or outside the organisation; and (2) the degree of control that entity is given over defining the organisation's actions. This results in two internally operating control systems – a bureaucratic accountability system, where accountability results from hierarchical relationships, and a professional accountability system, which defers control to staff with expertise within the organisation.

Next, two externally operating control systems are distinguished – a legal system, whereby accountability similarly follows from a hierarchical relationship but the controlling entity is based *outside* the organisation, and a political system, whereby accountability operates as responsiveness by public servants to those to whom they feel accountable, which may include as diverse constituents as the general public, agency clientele, and special interest groups. Whilst accountability systems operate in conjunction, inherent conflicts exist between them, such that in practice one or two accountability systems tend to be primary with the others 'in place but under-utilized' (Romzek 2000).

The implementation of collaborative innovations is expected to be critically affected by the extent to which accountability systems align with the nature of the innovation introduced, which will often necessitate shifting emphasis regarding the relative importance of different accountability systems. Cooperative tax compliance, the collaborative innovation central in this study, is a tax regulatory innovation that represents a shift in thinking for tax administrations. It moves away from a deterrence approach where taxpayers are coerced to comply with tax rules and threatened by audits and penalties, to a more responsive and collaborative way of working, aimed at generating mutual benefits for the tax administration and large corporate taxpayers (Larsen and Oats 2019). The collaborative element of the innovation concentrates on real-time sharing of information between both sides, beyond what is legally required.

Companies enrolled in the innovation voluntarily commit to providing a higher level of openness towards the tax authority on their fiscal affairs. The tax authority on its turn promises more predictability regarding its future actions and faster clarity with respect to the fiscal implications of decisions planned by companies part of the innovation, thereby enhancing corporate tax certainty.

In terms of accountability systems, arguably professional and political accountability systems will be key in introducing collaborative tax innovations since the emphasis in cooperative compliance is on tax administration staff strengthening their expertise of corporate taxpayers and enhancing their responsiveness to what are no longer referred to as taxpayers but service users. When comparing innovations implemented across countries with different configurations of accountability systems, we expect senior public servants to act differently when managing conflicting accountability expectations. If accountability pressures are weak, public servants will have more freedom to single out accountability systems they prefer responding to, whilst if accountability pressures are strong public servants have less discretion to favour particular accountability systems as they have to balance pressures deriving from multiple accountability systems.

The configuration of accountability systems underlies public servants' perceptions of accountability, with felt accountability subsequently conditioning how public servants operate in relation to collaborative innovations. At the opposite ends of the spectrum, it could be hypothesised that senior public servants who have greater ability to define and control expectations in relation to collaborative innovations and control their organisation's actions (Romzek and Dubnick 1987), will put greater emphasis on seeking to meet accountability pressures deriving from stakeholders at which the innovation is aimed, i.e., the innovation's target audience. However, if public servants are more restrained in their ability to control expectations and organisational actions, they are likely to put greater effort in seeking to balance between multiple interests. This will include accountability pressures deriving from the innovation's target stakeholder group but equally those deriving from actors not direct part of the collaborative innovation relationship, such as business representative bodies, professional associations (e.g., tax adviser associations), taxpayer groups other than large corporate taxpayers, and other societal actors taking an interest in tax compliance of large companies such as media and NGOs.

Methods

The empirical analysis focuses on cooperative compliance, which consists of a series of collaboration-oriented innovations in monitoring of large corporate taxpayers by tax administrations. Given the aim to identify the impact of felt accountability by senior public servants for the evolution of collaborative public management innovations, a research design is needed that maximizes variation in terms of accountability systems as these are expected to be critical determinants of felt accountability. This consideration resulted in the selection of cooperative compliance programmes in the US and the Netherlands. In contrast to many countries which do not have cooperative compliance, or have such arrangements at an embryonic stage, the two countries selected are widely considered to possess advanced cooperative compliance regimes, with their tax administrations leading innovators in reconfiguring supervision in the field of corporate taxation (Bronzewska 2016). In addition, with cooperative compliance having emerged in both countries around 2005, the programmes' currently mature stage facilitates drawing longitudinal comparisons across the two systems.

- *insert table 1 about here* -

The empirical analysis presented in the next section is informed by a series of semi-structured interviews, conducted between 2015 and 2018 and primarily face to face, with 43 actors from the tax authorities, the tax advisory profession and in-house tax directors working within large businesses. Representatives from each group were interviewed across the two countries, with the Dutch interviews totalling 27 and the US interviews 16. The interviews lasted on average 60-90 minutes and were mainly conducted in the premises of the interviewee. All interviewees have had experience of working directly with cooperative compliance or acting in an advisory capacity in relation to these programmes, and most interviewees were able to reflect upon the evolution of the innovations since their introduction in the early 2000s. The interviews focused on the factors giving rise to cooperative compliance innovations in both countries, their initial design, subsequent rollout, and evolution overtime, with a particular emphasis on the role of involved senior public servants. Interviewees included public servants part of the top leadership team of both countries' tax administrations, who were personally involved in the design and rollout of the innovations, as well as frontline tax administrators with direct responsibility for interacting with large corporates. In addition, interviews were undertaken with senior tax specialists working for large corporates, and specialist tax advisors from Big-4 firms who provide tax advice to companies and frequently interact with tax administrations on their clients' behalf. This multiplicity of stakeholder perspectives strengthened the depth and comprehensiveness of interviewees' perspectives.

The interviews were digitally recorded and then professionally transcribed and analysed using NVivo analysis software, which also served to manage the data. For the analysis of the data, a three-stage process developed by Corbin and Strauss (2007) was followed: open, axial and selective coding. Open coding was used for identifying, breaking down, conceptualizing and categorizing the data. Examples of concepts and categories identified are enhanced tax management control arrangements adopted by corporate taxpayers part of the collaborative innovation, fiscal risk profiling by tax administrations, and anticipated benefits of cooperative compliance as perceived by companies versus tax administrations. Subsequently, axial coding was used in order to relate categories to their subcategories. A key phenomenon emerging at this stage were stakeholder concerns regarding the technical ability and resource capacity of tax administrations to adequately assess companies' fiscal risk profiles and collaborate with companies who had joined cooperative compliance but needed strengthening of their internal fiscal control systems in light of their experienced reduction of tax administration audits.

Finally, selective coding was used to integrate and refine major categories in order to create a larger theoretical scheme linked to Romzek and Dubnick's accountability systems (Corbin and Strauss 2007). This process demonstrated, for example, that in the Dutch case tax administrators perceived a hierarchical pressure to persuade companies to join cooperative compliance even when they held the view these companies were not (yet) in a position to depart from the tax administration's traditional monitoring. Similarly, but pointing in the opposite direction, bureaucratic accountability dynamics in the US case were reflected in front level tax administrators being highly cognisant of the Internal Revenue Service' (IRS) leadership disfavouring a rapid expansion of cooperative compliance. During our selective coding, we similarly connected with bureaucratic accountability systems, hierarchical pressures, as experienced by interviewed tax administrators, to adopt a shift in mental attitude when auditing corporate taxpayers (*'from distrust to trust'*). Another category connected with bureaucratic accountability systems during the selective coding is the nature of political-administrative relations, as interactions between senior tax officials and government ministers appeared key to explain the frequency and confidence with which senior officials utilized hierarchical arrangements when implementing the innovation.

A similar three-stage analysis was conducted to identify key concepts that could be linked to each of the other accountability systems, with the coding schemes providing the starting point for the subsequent thematic analysis, which focused on identifying similarities and variations in findings across the two countries. This abductive analysis comprised multiple readings of the transcripts to identify recurring themes in light of the theory and extant literature. Whilst many similarities were traced between the two countries concerning the drivers and design aspects of their respective cooperative compliance regimes (see below), the focus was on understanding features of the collaborative innovation and the accountability context in which they developed that helped to explain why cooperative compliance played out differently in the two settings.

Using this constant comparative method (Corbin and Strauss 2007), highlighted multiple factors that help to explain the dissimilar rollout of the innovation across the two countries. This includes greater staff pressures faced by the US IRS when compared to the Netherlands Tax and Customs Administration (NTCA) (see Figure 1), and a more formalised design of cooperative compliance in the US when compared to the Netherlands, the latter resulting in the innovation being more time consuming to manage for the IRS compared to the NTCA. However most interviewees in both countries highlighted what they perceived as the critical role fulfilled by senior public servants during the decision-making on the design and methods of evaluation of cooperative compliance, with these decisions greatly influencing the pace of the innovation's subsequent expansion. Hence, whilst the fieldwork identified many similar causes for the implementation of the collaborative innovation, it is through the constant comparative method that considerations and actions of senior public servants emerged as key to illuminate *differences* in the implementation trajectories of the collaborative innovation across the two countries.

Next to the comparative method, validity of this research was sought to be enhanced through comprehensive data treatment. This encompassed the continuous analysis of interview results and extensive discussion between the two researchers, both of whom were involved in the fieldwork. This indicated that concerning the key lines of inquiry interviewees articulated highly similar views, regardless as to whether they were based inside or outside the tax administration, or the level of seniority they occupied, thereby suggesting external transferability. Interview findings were also constantly compared with archival material and extant research, in particular policy documents such as tax administration annual reports and reviews of cooperative compliance as conducted by oversight bodies. Documentary analysis was conducted prior to the interview work, thereby informing the interview questions, but also following the interviews as a method to triangulate interview data which strengthened the validity of the research findings.

Research context: Cooperative compliance in the US and the Netherlands

The two countries offer telling institutional contrasts to enable comparison in terms of their tax systems and accountability cultures, the latter resembling the country specific configuration of accountability systems. The accountability culture in US government services has been characterised as pragmatic but also blame and rule oriented with an emphasis on short-term results over long-term outcomes (Romzek 2005). The literature on the US IRS is mixed, with some observers claiming the organisation is ‘out of control’ (Tanfani *et al.* 2013). Others however suggest the IRS is responsive to elected political principals due to the presence of highly developed output control mechanisms, which are frequently employed by the US legislative branch for monitoring of government agencies including the IRS (Scholz and Wood 1998). The US has long been recognized as a country with adversarial regulatory relationships, with interactions between the IRS and taxpayers portrayed as formal, distant, rigid and suspicious (Thorndike 2001). The Dutch administrative culture is more consensus based, with a regulatory preference for informal oversight mechanisms. This regulatory approach is reflected in the Dutch fiscal culture, which has been characterized by relatively high openness and easy access by large companies to the Dutch tax authority and Finance Ministry (De Widt 2017). A further institutional difference between the two countries is that unlike the IRS, the Dutch tax administration is not an agency but integral part of the Dutch Finance Ministry.

In relation to the tax administration of large companies, the two countries also share similarities that incentivized developing collaboration-based innovations. Most importantly, the tax authorities in both countries experienced growing backlogs and frequent disputes with corporate taxpayers in the 1990s and 2000s, which resulted in considerable uncertainty for companies regarding their tax liabilities. As corporate audits frequently took multiple years to complete, they consumed large administrative resources for companies and tax authorities which was perceived as increasingly problematic by both sides (Bronzewska 2016). In addition, the occurrence of corporate scandals in the late 1990s and early 2000s (e.g., Enron and Ahold) incentivized strengthening of corporate governance codes, which enhanced companies' internal control systems. Both the US and Dutch tax administrations were eager to take advantage of the improvements in companies' governance systems when modifying their supervision.

In terms of institutional features, the Dutch and US cooperative compliance programmes share many similarities. In both countries, cooperative compliance is primarily the preserve of dedicated units within the tax authority tasked with managing the relationship with large business taxpayers, with both tax authorities having invested in upskilling tax administrators on the complexities and context of large corporate taxpayers. In comparison to traditional monitoring, tax administrators use more sophisticated procedures as part of the programmes to assess the fiscal risk profile of companies. The shared aim of the Dutch and US programme to provide companies enrolled in the innovation with early fiscal certainty has enabled both tax administrations to finalise companies' tax liabilities much faster, subsequently resulting in a much-welcomed reduction in companies' fiscal workload (Bronzewska 2016; De Widt 2017; De Widt et al. 2019).

Next to their similarities, several differences exist between the Dutch and US programme. Whilst cooperative compliance has been introduced in both countries using pilots, the length of the pilots differs significantly, ranging from around a year in the Netherlands to six years in the US. Further, whilst in both countries companies can join the programmes on a voluntary basis, the number of companies having joined differs significantly – being substantial in the Netherlands but having remained limited in the US (see Table 1). Another key difference relates to the level of formalization of access to the programmes – whilst in the Netherlands a rather loose trust focused covenant is signed (NTCA 2013), in the US a more formalized approach is taken, with companies having to go through a rigorous application process before they can access the innovation (De Widt et al. 2019).

The next section investigates the background of these differences and traces a key role for senior public servants perceived accountability in relation to the innovation. Reflecting our longitudinal approach, we first consider the initial design and rollout of the innovation, followed by an analysis of the consolidation phase.

The Netherlands

Accountability dynamics during HM's programme design and rollout

In the Dutch case, the initiative for developing cooperative compliance, known as horizontal monitoring (HM), originated from senior public servants based in the Ministry of Finance. The term HM refers to the more equal (horizontal) relationship between the tax administration and taxpayers as opposed to traditional top-down regulatory oversight. The need for an alternative supervision approach became particularly visible in 2004 when senior tax officials initiated a meeting with the Dutch business community. At the meeting, officials faced unexpected strong criticism from business representatives stating that the NTCA “was adopting a ‘them and us’ mentality” and emphasized the need for a more collaborative regulatory approach (interview senior official). These criticisms, combined with backlogs and fiscal disputes which were putting significant pressure on NTCA resources, encouraged senior officials to develop a supervision model that would rely more strongly on collaboration with businesses.

The effort was led by the Management Team of the NTCA, and in particular one of its members, Theo Poolen, who also occupied the position of NTCA’s deputy Director-General. The NTCA’s Management Team is part of the Finance Ministry, and is, alongside the tax authority itself, part of the Ministry’s Directorate-General for the Tax & Customs Administration. The Management Team constitutes the executive board of the NTCA but is chaired by the Director-General Tax Administration of the Finance Ministry. Even though part of the Finance Ministry, the Directorate-General for the Tax & Customs Administration, and the Management Team as part of it, has a tradition of operating in relative independence from the rest of the Finance Ministry. Due to the NTCA’s highly layered management structure, the Management Team is also relatively far removed from the tax authority, for which it operates as the executive governing body (Borstlap and Joustra 2017).

Interviewees indicate that the Management Team developed the new supervision model in a highly intuitive fashion, hereby reflecting an informal governance style they associate with the Management Team. One interviewee indicates that there was one concise document including ‘a couple of points’ which served as the model’s starting point, whilst another senior public servant, who had been a member of the Management Team himself, comments:

> We have done it spontaneously, more or less ourselves. I have not started to look in the literature or so. But at one point there is also intuition. Yes, and intuition I always say,

> that is solidified expertise. You see everything, you hear everything and then you combine things. At a certain point this rolled out.

Although public servants were leading the reform with little involvement from politicians, public servants did ensure the new supervision model received political backing. This applied particularly to the State Secretary of Finance, – who in the Dutch system is responsible for tax administration – with one interviewed senior official stating: ‘we managed to convince the State Secretary of Finance that this was the right way of working’. Obtaining political backing was eased by the hybrid institutional position of the NTCA’s Management Team which provides senior tax officials with easy access to government ministers and positively influenced trust levels between government ministers and officials.

In addition, as Dutch executive-legislative relationships have grown increasingly monistic over the past decades (Andeweg and Irwin 2014, 174ff), the political environment in which senior officials operate contributed to ensuring solid ministerial and parliamentary support for policies signed off by the cabinet. This sympathetic legislative environment explains the limited role attributed by interviewees to legal accountability mechanisms during the initial implementation of HM. The supporting political environment also emboldened willingness amongst senior officials to take risks by implementing HM at a relatively large scale and fast pace. Following the model’s introduction in 2005 and 2006 with two pilots including forty companies, the innovation expanded rapidly, with companies enrolled in HM encompassing around 40% of the NTCA’s 9,600 sized Large Business Division by 2016 (De Widt 2017). HM further expanded to medium-size and small businesses, even though acquiring a different shape.

Companies were initially hesitant to join HM due to the greater degree of fiscal transparency they need to provide as part of the collaborative way of working, however corporate anxiety disappeared quickly once companies detected the benefits of HM. Companies also felt encouraged to enrol in the innovation when they observed the dedication of senior tax officials in seeking to ensure HM would deliver on its promised benefits for companies. Relevant too were personal visits made by members of the Management Team to company boards aimed at encouraging companies to join HM. However not all interviewed senior public servants thought this was necessarily a role Management Team member 'should get involved with'.

The decision to expand HM was taken following the two pilots which had merely been evaluated through a user experience survey, conducted amongst the participating companies and tax administrators involved. A more advanced evaluation, such as one including a null measurement, was not chosen as considered 'too cumbersome', as stated by multiple interviewed senior officials. During the initial years, tax inspectors were encouraged by senior officials to enrol as many companies as possible as quantity was perceived vital in order 'to see how HM would perform in practice' (senior tax official quoted in Stevens et al. 2012, 212).

Within the tax administration however, senior officials met significant resistance against the rollout of the regulatory innovation, with one interviewed senior official stating:

> There was a lot of resistance. [-] If you've worked a certain way for 25 years and you think you're doing a good job, and most of them did a good job, then you're not going to give up easily when some madman suddenly comes along from The Hague [i.e., the seat of the Dutch government] and says, 'this is it'. I wouldn't trust that either.

Senior management however was determined in 'giving the model a go', and, strongly relying on bureaucratic accountability mechanisms, significant resources were dedicated to seeking to ensure NTCA front level staff adopted the new working methods, such as compulsory training, whilst those 'who did not buy into the new arrangements' (interviewee NTCA) were frequently transferred to functions in the tax administration where they no longer had direct client facing interaction. Those tax officials who felt reluctant about HM were particularly concerned about the impact of the innovation on companies' tax compliance levels, and, as interviewees indicate, often felt there was insufficient attempt in the innovation's design to seek to capture its compliance effects.

Another reason for senior officials seeking a speedy rollout of the innovation relates to the emergence of the term compliance in Dutch tax debates during the early 2000s, which derived from the international (tax) regulatory discourse. It was increasingly felt across the Dutch tax community that tax administration performance indicators which had been commonly used until then, such as number of tax audits conducted, were of little relevance to assess taxpayers' compliance levels. It is in the search for more appropriate tax compliance indicators that, upon the suggestion of senior tax officials, the number of enrolled HM companies was introduced in the Government Budget (which includes a section on tax administration) as an indicator of corporate tax compliance. Based on the assumption that an increase in the number of HM enrolled companies would signify an improvement of cooperative compliance, the Ministry of Finance started to use the indicator in its drafting of performance contracts with the NTCA. When the NTCA on subsequent occasions failed to reach these targets, Members of the Lower House started to pressure the NTCA to accelerate HM's expansion. Hence although legal accountability mechanisms were insignificant in relation to HM at first, actions by senior tax officials activated pressure from the legislature, something which one interviewed senior official referred to as a 'self-inflicted problem'.

Accountability dynamics during HM's consolidation phase

With HM having acquired a prominent and preferred status within the NTCA, MPs but other stakeholders too, such as the Dutch accountancy profession, started to raise questions as to how the NTCA decided on a company's admittance to the programme. For this purpose, the NTCA had initially relied on a loosely formulated covenant agreement between the NTCA and the joining company, which focused on establishing a relationship of trust between both sides, with the company indicating *willingness* to gain tax control and be transparent on its tax affairs towards the NTCA. However, the emergence of accountability pressures following the programme's rapid expansion incentivised the NTCA to shift away from its original trust-based focus towards a model seeking to assure stakeholders more robustly that the NTCA did not act out of blind faith when enrolling companies in the programme. Consequently, a key role was assigned to the Tax Control Framework (TCF), which refers to a company's internal systems that support its tax risk management and should ensure a company is fit to join the programme in terms of its tax compliance levels. On this transition, one interviewed tax official comments:

> HM started very much as a gentleman's agreement – we agreed 'this is how we will interact with each other'. [-] However later on, an enormous number of guidelines were introduced including how to conduct a compliance review [of companies applying to HM]. We [the NTCA] now said 'there must be justifiable trust'. This was all made up in retrospect and not at the time HM was introduced.

Although senior officials state they felt the urgency to implement changes to HM when facing intensifying accountability pressures, the rapid rollout of HM meant that the implementation of agreed upon changes and assurances caused major resourcing challenges to the NTCA as they made the innovation's administration significantly more time consuming. The programme's rapid expansion including to companies with a limited level of fiscal control now avenged itself, as these companies required significant tax administration support to bring their level of tax compliance to an acceptable level. The NTCA however was unable to provide this due to capacity constraints, as stated by our interviewees, and found in evidence elsewhere (Stevens et al. 2012).

In turn, the modifications brought to HM and the NTCA's subsequent resourcing challenges generated new accountability pressures from businesses enrolled in HM. The resourcing problems rendered the NTCA incapable of intensifying its supervision of companies deemed to be at high risk of being noncompliant, something to which the NTCA had committed itself during the innovation's rollout. Hence, corporate interviewees indicate they felt that the NTCA did not 'keep its side of the bargain'.. As the prospect of intensified supervision constituted an incentive in itself for businesses to join HM, enrolled companies now started to question 'whether it was worth making the effort' of voluntarily providing greater transparency towards the tax administration if businesses not part of HM enjoyed a similarly low level of scrutiny in practice. Further, multiple interviewees criticised HM for enhancing the risk of corporate tax avoidance as it allowed companies not officially enrolled in HM to enjoy the innovation's attraction of receiving light-touch supervision, even though they had not been subjected to NTCA reviews aimed at establishing they were in fact low-risk companies (cf. Awater and Kuijpers 2017).

Doubts about the real benefits of the regulatory innovation have been further aggravated by the perceived reduction in speed by which the NTCA provides clarity on fiscal queries raised by companies, something which itself has been caused by increased scrutiny by the European Commission of Dutch tax administration practices (e.g., *FT* 2019). Senior officials interviewed displayed awareness of this being an important corporate concern and stated that they act upon this felt accountability. In particular, they sought to chase decision making within the NTCA if a fiscal query concerned a company enrolled in HM to ensure the model 'delivers' for opted-in companies. Consequently, intensifying legal accountability pressures deriving from the European level have made that senior officials have started to feel more conflicted about how to ensure HM companies remain satisfied. Another difficulty faced by senior officials promoting the innovation is that the role of national level NTCA working groups, which are aimed at addressing complex fiscal queries arisen in the interaction with corporate taxpayers. These working groups principally consider tax queries from a tax specialist' perspective, consequently feel less accountable to opted-in companies than senior public servants and frontline tax officials to maintain a good relationship with HM companies.

The rapid expansion and growing concerns, within and outside the NTCA, about frictions occurring in HM's implementation, encouraged the State Secretary of Finance to initiate a first extended evaluation of HM in 2011. The review was conducted by the Stevens Committee, named after its chair Leo Stevens, a widely respected Dutch tax professor, with other members including two professors, both with links to the NTCA, a public servant from the Interior Ministry, and a former tax director of Royal Dutch Shell. In its published report, the Committee stated it had struggled to obtain 'self-evident' (Stevens et al. 2012, 4) policy information and concluded that the NTCA had devoted 'insufficient energy and effort to a management approach on the basis of adequate (management) information'(ibid), making it impossible to draw conclusions on the innovation's impact.

In response to the Committee's observations, the NTCA decided in 2012 to conduct an empirical evaluation of HM, results of which were published five years later (Belastingdienst 2017). Findings from the follow-up investigation show that HM companies did not outperform or underperform companies under traditional supervision in terms of number of corrections made by the tax administration. Interviewees more supportive of HM emphasised these results showed that the innovation is not negatively related to companies' fiscal compliance. However, those critical of HM found confirmation of their view that HM's light touch supervision is unwarranted given that HM companies failed to demonstrate a higher compliance level, which is opposite to what the model predicts.

Although interviewed officials indicated being aware of concerns in terms of HM's impact for NTCA staffing and tax revenues collected, these issues were not considered in the NTCA's evaluation, according to NTCA interviewees involved with the study, due to practical and research limitations. Other interviewees however emphasised the NCTA's leadership had displayed little willingness to identify the innovation's administrative and revenue impact, with one interviewed tax official stating it was widely accepted that HM was an expensive monitoring model, a view reflected by most interviewees. An interviewed tax adviser suggested this reflected a political choice, and a price the Dutch government was willing to pay in order to preserve international tax competitiveness.

Over time, institutional defence for HM within the NTCA has diminished as senior officials who introduced the innovation moved to other positions with their successors taking limited ownership of the innovation or criticizing it in internal communication, some of which subsequently leaked to Dutch media (e.g., Kuijpers 2017). In addition, (former) tax inspectors have publicly criticised HM (Awater and Kuijpers 2017), whilst increasingly critique has also been articulated by stakeholders such as the legal and accounting profession, although generally in a subdued manner (e.g., Dutch Association for Tax Science 2015). Interviewed senior officials expressed awareness of these criticisms and perceived a need 'to further develop the model', which triggered some changes gradually implemented to HM during 2020-2022.[1] Whilst the impact of these changes remains to be seen, the NTCA has indicated it remains committed to the principles of HM, despite the enduring knowledge gaps regarding its administrative and revenue impact.

[1] This includes the termination of HM covenants with the one hundred largest companies operating in the Netherlands and replacing those covenants with an 'individual tax monitoring plan'. Medium-sized companies will continue to be able to conclude a covenant but under stricter requirements, whilst other large companies can only participate in HM through their tax adviser (Olsthoorn 2022).

The US

Accountability dynamics during CAP’s design and rollout

Prior to the introduction of the Compliance Assurance Process (CAP) programme, the IRS had experimented with various mechanisms to solve disputes between the IRS and large corporate taxpayers including a fast settlement programme in 2001. CAP can be seen as a continuation of these initiatives and was introduced in 2005 as a pilot programme with 17 voluntary participants. A major driving force behind CAP was the Commissioner of the IRS at the time, Mark Everson, who encouraged IRS staff to develop innovative solutions for speeding up the process between filing corporate tax returns and finalizing tax liabilities.

Interviewed senior IRS officials emphasize that Everson acted largely in response to pressures from the corporate sector and Congress to make the IRS more service oriented and increase its efficiency. The IRS had continuously faced such scrutiny, but political pressure intensified during the second half of the 1990s when, with a Democratic President in place, Republicans took over Congress. Congress’ dissatisfaction with IRS performance and divided government, led Congress to take on the initiative and draw up plans to reform the IRS, including the transformation of the IRS’s regional structure into four divisions organized around categories of taxpayers, including a Large Business division in which CAP originated (Rossotti 2005; Thorndike 2001).

CAP was first introduced to a selected number of business representatives during a one-day event in the Summer of 2004 where, with use of highly detailed flowcharts, businesses were presented with what one IRS interviewee referred to as a ‘conceptually thought through’ supervision model. For example, and opposite to the Dutch model at that time, CAP foresaw an appeal function. IRS interviewees emphasize that CAP was very much the result of a collective process involving administrators from across the IRS Large Business Division, not limited to its senior leadership team. In addition, the Tax Executives Institute (TEI), which is the leading US association of corporate/in-house tax professionals, worked with the IRS in the development of CAP. The TEI Executive Director at the time described it as a 'collaborate effort to resolve issues at the lowest level' (McCormally 2005).

As CAP was significantly different from traditional auditing, concerns over capacity issues CAP could generate for the IRS were present from the outset. One interviewed senior IRS official involved in CAP's introduction comments: 'We were sure in the IRS we could do it, but we were not sure at what scale'. This nervousness was echoed by an IRS spokesperson who said in 2005, according to minutes published by TEI, 'The agency is committed to making the initiative work, but the quality of examinations [audits] cannot be compromised' (TEI 2005).

The pilot, which occurred in close collaboration with companies, was evaluated through a survey undertaken by an independent research firm and showed that most companies expressed a desire to continue with CAP. The collaborative approach in CAP is reflected in companies working with an IRS team to identify and resolve potential tax issues before their annual tax return is filed, which subsequently leads to a shorter post-filing examination period as fewer contentious items will need to be dealt with. According to former IRS Commissioner Nolan, it is due to CAP that 'for the first time ever, corporate taxpayers are working cooperatively and transparently with LMSB agents to resolve their tax compliance issues' (Bronzewska 2016, 209).

Despite significant corporate interest for the innovation, concerns amongst IRS senior management on the resource implications meant that CAP was rolled out reluctantly, leading to a pilot period of six years. Reluctance by IRS leadership in expanding CAP amongst corporate taxpayers and in promoting it amongst IRS staff, translated into a lack of enthusiasm amongst tax administrators when being approached by companies who wanted to join the programme. One corporate interviewee stated that when their company expressed interest in joining CAP ‘they [IRS tax administrators] were sort of almost discouraging, like “well…”’. Due to this, CAP expanded slowly, also after the IRS’ decision in 2011 to make the programme permanent, resulting in only 181 CAP companies by 2016 (‘a drop in the bucket’, according to one interviewee). Interviewees emphasize that there has been no lack of interest in CAP from companies, amongst whom CAP proved popular almost instantaneously (cf. KPMG 2018).

The wider political constitutional context in which the IRS operates is relevant to understanding senior officials’ decision-making regarding CAP’s rollout. With divided government characterizing parts of the Clinton and Obama administrations, appropriations to the IRS have been heavily politicized. In addition to opposition by Congressional tax adversaries, IRS’ relationships with Congress further soured around 2013 when it was revealed that the IRS had selected political groups for scrutiny, including the Tea Party, which triggered several investigations and led to the IRS being accused of political bias (Lederman 2016). In addition to political pressures, the IRS has faced a significant increase in workload due to the enactment of new legislation, such as the Affordable Care Act. The unpopularity of this legislation with many Republicans in particular means that Congress has frequently applied reductions on Presidential budget requests for the IRS, especially after Republicans took over control of the House of Representatives in 2011. For example, the recommended budget by the IRS Oversight Board for FY2014 was $13.074billion, and whilst President Obama requested $12.861 billion, Congress appropriated only $11.290 billion (IRS Oversight Board 2014, 8). Although Congress has not had a particular interest in CAP, the huge budgetary pressures faced by the IRS have strongly affected IRS decision-making about CAP. An interviewed senior IRS official comments:

> Because the IRS [-] is a relatively underfunded tax administration compared to our peers, we need to be very clear with our employees what we want them to do. Because the effects of CAP are unclear, and sometimes seem to be the opposite with what we aim for, it would be difficult for me at this moment to defend relocating staff to CAP.

Budgetary pressure means that the relationships with both the executive and legislative branch are crucial for IRS performance. This need is enhanced by the fact that the IRS receives less natural support from the Treasury when compared to support the NTCA receives from 'its' Finance Ministry. Although the IRS is not as independent as other US agencies (Gore 2013), due to the IRS' position as a federal government executive agency and a bureau of the Treasury, the Treasury has historically tried to distance itself from IRS operations to alleviate historical concerns that the IRS would be used for political purposes, combined with the Treasury's natural tendency to focus on policymaking rather than implementation (Thorndike 2001). The IRS Commissioner, together with the IRS Chief Counsel, are the only two political appointees in the IRS, appointed by the President with Senate consent. To remove partisanship from the IRS, both appointees have five-year terms that overlap the four-year presidential election cycles (Gore 2013). In addition to being under-resourced (Johnson 2016), the IRS is subjected to frequent judicial review (Hickman 2017) and faces a large, according to some observers excessive number of regulatory overseers (Forman and Mann 2015). As the next section shows, this has given rise to strong accountability awareness amongst IRS senior public servants, which is reflected in CAP's rollout.

Accountability dynamics during CAP's consolidation phase

Constant political and budgetary pressure on the IRS reduced goal ambiguity during the implementation of CAP and resulted in a strong focus on the programme's efficiency implications, primarily for the IRS. Interview findings and documentary review demonstrate that pilot CAP participants were generally satisfied with the programme, in particular the benefits of having old issues resolved (Boocock et al. 2007). From inception, the TEI introduced bi-monthly conference calls amongst corporate tax directors engaging with CAP. In 2012, the IRS Commissioner is recorded as inviting TEI members to contribute to improving the accountability document that taxpayers and the CAP audit team use (TEI 2012). In the second year, an IRS spokesman noted the need for caution in measuring success of CAP. In particular, the performance measures traditionally used for IRS tax officials such as yield, assessments and issues raised, do not fit well with CAP and emphasis was put on ensuring that tax officials are evaluated properly (TEI 2006).

By the third year, the metrics used to evaluate CAP remained the same as non-CAP cases with the addition of qualitative 'diagnostic tests such as cycle time and employee satisfaction' to compare CAP and non-CAP. In addition, since CAP's introduction, all CAP cases have been subjected to independent quality review. In response to a query from the TEI about quality assurance, 'Ms Nolan [the IRS Commissioner of the Large Business Division] replied that a number of checks and balances are incorporated in the programme because [the Large Business Division] knew it would have to explain and justify the changed approach', and further, 'the worst that can happen is that LMSB might be accused of "giving away the store"' (TEI 2007). As CAP entered its fourth year, the IRS announced that in-house tax experts would be consulted, and it would reach out to corporate taxpayers to help develop diagnostic tools. Further, IRS interviewees emphasise IRS senior officials actively encouraged staff to investigate cooperative compliance experiences in other countries, something starkly contrasting with the Dutch tax administration's approach at the time, which focused on promoting the Dutch model internationally through international forums such as the OECD. The continuing lack of standard metrics made allocation of resources for CAP challenging during the IRS budget process, and interviewees and TEI correspondence indicate that from 2013 concern is growing amongst companies that tax administrators applying CAP were moving back to a more traditional audit approach, thereby undermining the programme's purpose (e.g., TEI 2013).

IRS' cautiousness to expand CAP has been further instigated by the multiple bodies overseeing IRS' performance. This includes the Government Accountability Office (GAO) who stated that the IRS lacks 'a credible base of information for understanding the current effectiveness of CAP' (GAO 2013, 12). Further, a review was conducted in 2013 by the Treasury Inspector General Tax Administration (TIGTA), which is based in the US Treasury but conducts independent oversight of IRS activities. TIGTA's review focused on the resourcing and revenue implications of CAP for the IRS. Although the method is not without shortcomings, TIGTA has sought to compare the impact on the IRS of CAP versus traditional supervision, concluding that the former consumes substantially more IRS staff hours, with the hourly revenue rate collected under CAP claimed to be around a third of that collected under traditional audits (TIGTA 2013). The TIGTA review further concluded that whilst there was some favourable feedback on CAP, especially from enrolled companies, additional analysis of cost and benefits was needed before considering expanding the programme (TIGTA 2013). Noteworthy too is the Taxpayer Advocate Service (TAS), which is an independent organization based in the IRS and acts as taxpayers' ombudsman. Whilst TAS has not specifically considered CAP, its unceasing pressure on IRS leadership to provide an equal level of tax administration service across all categories of US taxpayers makes IRS officials cautious of the potential detrimental impact of allocating more tax administration resources to a specific category of taxpayers, such as large corporates (Cords 2020).

The IRS decided in 2016 to put CAP on hold, stating a reassessment of CAP was required 'given today's challenging environment of limited resources and budget constraints'[2]. In addition to the difficulty of measuring CAP's revenue impact, interviewed senior IRS officials also referred to the role of the CAP's external reviews as contributing to the IRS' decision to put CAP on hold. Subsequently, a two-year review period was conducted, generating corporate concerns that the programme would be abolished altogether. In 2018, the IRS announced a CAP 'recalibration', which has resulted in the first significant changes to CAP since its original introduction. The changes, devised in consultation with external stakeholders including the TEI, concentrate on more explicit rules as to how CAP companies are expected to engage with the IRS, with the IRS providing more explicit examples of the type of behaviour that would preclude acceptance of a company into the programme (De Widt et al. 2019). Further, a more stringent CAP application process applies with companies required to provide additional information, which should better enable the IRS to evaluate whether a taxpayer's circumstances can be best addressed within CAP or through traditional filing processes (KPMG 2018, 2022). Whilst CAP was reopened for new applicants in 2019, the programme has only seen a limited expansion hitherto, mostly due to ongoing difficulties in developing CAP metrics.

Comparative Analysis

This section compares the accountability dynamics in relation to the collaborative innovation investigated in the two countries. The expectation amongst senior officials of being held accountable differs substantially between the two cases, and to a significant extent is a product of the country specific constellation of accountability systems as perceived by senior public servants. First, we review internally operating accountability systems (i.e., bureaucratic and

[2] https://www.irs.gov/businesses/corporations/irs-continues-comprehensive-assessment-of-the-cap-program.

professional), which is followed by an analysis of external accountability systems (i.e., legal and political).

Professional accountability pressures

Key for the success of innovations is the presence of a match between the innovation introduced and the accountability systems present. In the Dutch case, the introduction of HM enlarged discretionary space of frontline tax officials as enabling them to resolve tax queries more quickly and assess risk profiles of corporate taxpayers. By putting increased weight on the role of frontline staff, the innovation increased the role of professional accountability systems. Over time however new accountability pressures deriving from the EU level increasingly complicated HM's reliance on professional accountability, and instead growing use has been made of bureaucratic accountability systems aimed at enhancing uniformity in tax administration decision making. A misalliance between the innovation and accountability systems has not occurred to a similar extent in the US, both owned to the fact that in the US the innovation did not change substantially over time, nor did major shifts occur in accountability pressures, as happened in the Dutch case ensuing intensifying scrutiny by the European Commission.

Whilst in both systems performance measurement has proven challenging, in the US objectives of the innovation have been relatively clearly defined and remained largely unchanged over the lifetime of the innovation, unlike the Netherlands. The greater budgetary pressures experienced by the IRS enhances accountability felt by senior IRS officials in response to legal accountability pressures. In the Netherlands in contrast the lower level of accountability pressures accounts for lower efforts by senior public servants to narrow down the innovation's objectives and explains both why different and at times contradictory objectives were pursued, and why critical elements of the innovation were added after it had been rolled out amongst many companies.

A similarity between both cases is that companies can exit the innovation, however this appears to have generated more accountability pressures for the NTCA compared to the IRS, given the NTCA's preference for large companies to join and remain part of the innovation. This is reflected by what is in practice a continuing participation of companies in HM, once they have been enrolled in the programme, whilst in the US case, companies must annually reapply to remain eligible for CAP.

Bureaucratic accountability pressures

In case of bureaucratic accountability systems, accountability results from hierarchical relationships. Bureaucratic accountability pressures appear to have significantly impacted the Dutch innovation, given the high degree of control by the NTCA Management Team over the tax administration organisation. The hierarchical way of operating of the Management Team vis-à-vis the tax administration executive organisation, which some interviewees suggested is at odds with common structures in Dutch government, has been enhanced by the NTCA's Management Team being located within the Finance Ministry but operating relatively independently from the rest of the Finance Ministry. The resulting proximity between Dutch senior tax officials and government ministers made it easier for Dutch senior public servants, as compared to their agency-based US counterparts, to manage expectations held by government ministers. The assurance of political backing appears to have increased confidence of the NTCA's leadership to operate hierarchical coordination mechanisms when rolling out HM across the tax administration. In the US, in contrast, the presence of greater external scrutiny made IRS senior officials' more wary to ensure they incorporated expertise from across the IRS and to slow down the innovation's rollout given the difficulty of generating of performance metrics.

Legal accountability pressures

IRS senior officials demonstrated great wariness that any significant changes they bring to IRS tax administration processes are likely to be scrutinized by organizations external to the IRS. This involves scrutiny by political actors, especially Congress, whose scrutiny intensifies in situations of divided government. This was reflected by IRS interviewees emphasizing that executive branch support is insufficient, making it equally critical for the IRS to carefully maintain political accountability relationships with the legislative branch. Further, and unlike the Dutch case, IRS senior officials displayed high awareness of non-political, governmental bodies taking an interest in IRS performance. References were made to TIGTA and its review of CAP, with TIGTA perceived by senior officials as an authoritative monitoring body, aided by the organisation's impartiality and its exclusive focus on tax administration which has enabled TIGTA to build up strong tax administration expertise.

In contrast to the IRS, legal accountability pressures are perceived as less important by Dutch senior officials. This difference can be partly attributed to the more natural supportive attitude Dutch public servants receive from government ministers. In their turn, the constitutional functioning of the Dutch system enables government ministers to manage, to a large extent, legislative scrutiny taking place of the tax administration's undertakings. Legal oversight of the NTCA appears noticeably less intense compared to the IRS – there are fewer oversight bodies whilst those that do exist have taken limited interest in HM. This has undermined the extent by which monitoring organizations are perceived by senior public servants as knowledgeable to scrutinize the complexity of tax administration processes including HM. With an equivalent to TIGTA absent in the Dutch system and other oversight bodies such as the Dutch Court of Audit demonstrating limited interest in HM, it was the Dutch Government itself who initiated a review of the innovation. Despite the critical observations produced by this review, the fact that the Stevens Committee members were appointed by the State Secretary of Finance, who like the Finance Ministry was known to be a strong proponent of HM, reduced the accountability pressures arising from the first official review. This is reflected by the NTCA not providing an official response to the Stevens Committee' recommendations, and might similarly explain why the Committee provided an endorsement for HM despite the near unavailability of any effect measurements the Committee had identified in relation to HM.

Political accountability pressures

Political accountability is defined by Romzek and Dubnick (1987) as responsiveness by public servants to those to whom they feel accountable, and are not formally controlled by. This includes diverse constituents in both the US and the Netherlands, which similarly instigated different behaviours of senior officials in relation to the collaborative innovation. In the Dutch case, the main political accountability pressure is in relation to the innovations' target stakeholders, i.e., corporates enrolled in HM. There has historically been a closer interaction in the Netherlands between large companies and the Finance Ministry as compared to the equivalent relationship in the US. The closer interaction in the Dutch case partly reflects the greater efforts small countries need to display to retain large companies, however it is also a result of the rapid rollout of the collaborative innovation and the personal commitment by Dutch senior officials in seeking to encourage companies to join and remain in HM. Their personal efforts in seeking to expand the innovation and building 'trust-based relationships' with companies increased Dutch senior officials' felt accountability towards companies. This is unlike the US where there has not been a similar level of commitment and investment by senior officials in building trust-based relationships. The more restrained rollout of CAP and the innovation's more formalized design, makes that IRS officials were able to maintain a greater level of independence towards corporate taxpayers, reducing officials responsiveness in relation to political accountability pressures deriving from the corporate stakeholder group. Nonetheless, accountability pressures deriving from corporates are relevant in case of the US innovation too, reflected by IRS officials referring to the difficulty of entirely abolishing CAP given the expectations it had raised amongst large corporates and the 'stakeholder group we [the IRS] have created'. However the comparatively much smaller size of CAP has made it easier for officials to withstand accountability pressures deriving from corporates.

The large scale and speedy pace of HM's rollout increased external scrutiny from actors including journalists and tax advisers, however interview findings demonstrate that this did not substantially enhance senior public servants' responsiveness towards these criticisms. With regard to media, critical reporting about HM remained restricted to a small number of Dutch media outlets (especially *Trouw*, a small Dutch newspaper), reducing the impetus for senior officials to give account. Further, tax advisers who criticised HM have mostly argued for a further formalization of HM's processes and structures, akin to US CAP. Most interviewed Dutch tax officials suspected these criticisms were driven by tax advisers' commercial considerations as moving towards a more formalized HM model would better enable tax advisers 'to sell tax advisory products' to companies wanting to join HM. This presumed commercial motivation clearly reduced the legitimacy, as perceived by tax officials, of the political accountability pressures exerted by tax advisers.

Conclusions

This article identifies various factors that explain the different evolution of cooperative compliance in the US and the Netherlands, including growing scrutiny of national tax administration practices by supranational actors (i.e., the EU, in case of the Netherlands), and country specific government budget dynamics putting pressure on tax administration resources (the US). Whilst the qualitative nature of our study prevents us from drawing definite conclusions, a consistent finding arising from the fieldwork in both country cases is the critical role occupied by senior public servants. Their perceptions of accountability in relation to the innovation, including to whom they should be responsive, appear key to understand the divergent trajectories of the innovations, even when sharing similar starting conditions. The pressures perceived to derive from the multiple accountability systems, which we summarize in table 2, apply to both the processes applied for implementing the innovation and the corporate-tax administration relationships emerging form it. Senior officials' perceived accountability is critically influenced by the configuration of accountability systems, hereby illuminating the importance of institutional factors and the need to go beyond microlevel psychological approaches when analysing accountability perceptions and their impact on public servants' behaviour.

The expectation of being held accountable is lower amongst Dutch senior public servants, instigating a more moderately experienced sense of accountability and explaining an accountability responsiveness by Dutch senior civil servants concentrating on external stakeholders directly benefiting from the collaborative innovation, i.e., companies enrolled in the innovation. The fact that Dutch senior public servants' felt accountability was predominantly driven by political accountability pressures originating from corporates, amongst whom the innovation proved popular, illuminates why the innovation was rolled out swiftly with little consideration for the wider tax administration and revenue implications. The similarly collaborative nature of the cooperative compliance innovation introduced by the IRS instigated a felt accountability amongst IRS senior officials to seek to ensure the programme would deliver for companies who had voluntarily joined the innovation. However the more forceful presence of *additional* accountability systems clearly moderated the sense of accountability senior IRS officials feel towards companies as the innovation's core stakeholder group. It also instigated a more comprehensive awareness of the costs and benefits officials believed needed to be considered, and more diligent attempts to seek to evaluate these prior to decisions on the innovation's expansion.

The comparison also demonstrates that accountability pressures may substantially change over time, highlighting that in order to enhance the long-term robustness of innovations policymakers would benefit from considering accountability pressures from systems that may be weak or absent at the outset, but can grow stronger over time. It also underlines the research relevance of undertaking a longitudinal analysis when analysing accountability pressures experienced by senior public servants as they may critically explain an innovation's evolution.

The comparative analysis highlights that if innovations are more strongly linked to commitment by individual public servants and have been subjected to limited accountability pressures at the outset, as in the Dutch case, the innovation is more fragile over time to changes occurring in the configuration of accountability systems. This contrasts to a policymaking environment where accountability pressures are more substantial and diverse, as in the US, which impels senior public servants to be much clearer on the objectives sought to be achieved by introducing an innovation. These differences in policymaking environments are even more relevant in case of collaborative innovations where there is a greater risk of primarily seeking to satisfy actors directly involved in the collaboration and ignoring effects for wider stakeholders.

The Dutch case indicates that public servants may feel accountable and initiate performance reviews on their own accord, even when accountability pressures from politicians are limited. Initiating such reviews by the organisation who has introduced the innovation may have tactical advantages as enabling public servants to influence the scope of the review, however thereby reducing the prospect to satisfy accountability pressures deriving from a diverse array of stakeholders. Further, some scholars (e.g., Egeberg and Trondal 2009) claim that compared to senior officials based in a ministerial department, agency officials are likely to assign greater weight to user and client concerns as they are relatively insulated from political processes. Our comparison however demonstrates that it was in fact Dutch ministerial-based officials who demonstrated the greatest responsiveness to tax administration clients, whilst agency-based IRS officials displayed more alertness to political signals.

This article has identified some preliminary evidence about the relationship between senior public servants' felt accountability and the implementation of collaborative innovations in tax administration. Whilst the study identifies different accountability dynamics across the country cases and time periods investigated, further research of a larger range of collaborative innovations, including across different country contexts and policy domains, promises to provide further vital insights into the extent to which senior public servants' felt accountability and the subsequent actions these perceptions give rise to, are related to different institutional environments.

Notes on contributors

Dr Dennis De Widt is a Reader (Associate Professor) in Accounting and Finance at Cardiff Business School, Cardiff University, U.K.. His main research interests include tax administration, local government and intergovernmental relations, and public sector financial management.

Professor Emer Mulligan is a Professor of Taxation and Finance in the J.E. Cairnes School of Business and Economics at University of Galway, Ireland. Her research investigates the administrative, social, political and economic dimensions of taxation.

References

Aberbach, Joel, and Bert Rockman. 2006. "The Past and Future of Political-Administrative Relations: Research from *Bureaucrats and Politicians to In the Web of Politics*—and Beyond." *International Journal of Public Administration 29*(12): 977–95.

Agranoff, Robert, and Michael McGuire. 2004. *Collaborative Public Management*. Washington: Georgetown University Press.

Andeweg, Rudy and Galen Irwin. 2014. *Governance and Politics of the Netherlands*. Basington.

Ansell, Chris, and Alison Gash. 2008. "Collaborative Governance in Theory and Practice." *Journal of Public Administration Research and Theory 18*(4): 543-71.

Awater, Femke and Karlijn Kuijpers. 2017. "Inspecteurs zonder inspectiebevoegdheid." *De Groene Amsterdammer*, 8 February.

Belastingdienst. 2017. *Rapport Onderzoek Grote Ondernemingen.* Utrecht: Belastingdienst/Centrum voor Kennis en Communicatie.

Bommert, Ben. 2010. "Collaborative Innovation in the Public Sector." *International Public Management Review 11*(1): 15–33.

Boocock, Stephen W., John T. Ernest, and Melody L. Twigg. 2007. "Dealing with Fin 48: Techniques to Manage Tax Uncertainty through IRS Practices and Procedures." *Tax Executive 59*(2): 170–79.

Borstlap, Hans, and Tjibbe Joustra. 2017. *Onderzoek naar de Besluitvormingsprocedures Binnen de Belastingdienst.* The Hague.

Bovens, Mark. 2007. "Analysing and Assessing Accountability: A Conceptual Framework." *European Law Journal 13*(4): 447–68.

Bovens, Mark, Thomas Schillemans, and Robert Goodin. 2014. "Public Accountability", in Mark Bovens, Robert Goodin and Thomas Schillemans (eds.), *The Oxford Handbook of Public Accountability*. Oxford: Oxford University Press, pp. 1–22.

Brandsma, Gijs Jan, and Thomas Schillemans. 2013. "The Accountability Cube: Measuring Accountability." *Journal of Public Administration Research and Theory 23*(4): 953–75.

Braun, Virginia, and Victoria Clarke. 2006. "Using Thematic Analysis in Psychology." *Qualitative Research in Psychology 3*(2): 77-101.

Bronzewska, Katarzyna. 2016. *Cooperative Compliance. A New Approach to Managing Taxpayer Relations*. Amsterdam: IBFD.

Burga, Ruben, and Davar Rezania. 2015. "A Scoping Review of Accountability in Social Entrepreneurship." *SAGE Open 5*(4): 2158244015614606.

Corbin, Juliet, and Anselm Strauss. 2007. *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory*. Sage.

Cords, Danshera W. 2020. "Tribute to National Taxpayer Advocate Nina Olson." *Pittsburgh Tax Review 18*(1): 1-7.

De Widt, Dennis, Emer Mulligan, and Lynne Oats. 2019. "The US Compliance Assurance Process: A Relational Signalling Approach." *Journal of Tax Administration 5*(1): 145–61.

De Widt, Dennis. 2017. "Dutch Horizontal Monitoring: The Handicap of a Head Start." *FairTax Working Paper Series No. 13.*

Dutch Association for Tax Science. 2015. *Commissie Hoge Raad: 100 jaar Belastingrechtspraak.* Amsterdam: Dutch Association for Tax Science.

Forman, Jonathan Barry, and Roberta F. Mann. 2015. "Making the Internal Revenue Service Work." *Florida Tax Review 17*(10): 725–815.

FT. 2019. "EU Court Annuls Brussels' €30m Tax Ruling Against Starbucks." *Financial Times*, 24 September.

GAO. 2013. *Corporate Tax Compliance: IRS Should Determine Whether Its Streamlined Corporate Audit Process Is Meeting Its Goals.* Washington, DC: Government Accountability Office.

Gore, D. 2013. 'IRS Not So 'Independent', *The Annenberg Public Policy Center*, https://www.factcheck.org/2013/05/irs-not-so-independent/

Hall, Angela T., Dwight D. Frink, and M. Ronald Buckley. 2017. "An Accountability Account: A Review and Synthesis of the Theoretical and Empirical Research on Felt Accountability." *Journal of Organizational Behavior 38*(2): 204–24.

Han, Y. 2020. "The Impact of Accountability Deficit on Agency Performance: Performance-Accountability Regime." *Public Management Review 22*(6): 927–48.

Hickman, Kristin E. 2017. "Administrative Law's Growing Influence on US Tax Administration." *Journal of Tax Administration 3*(1): 82–94.

IRS Oversight Board. 2014. *FY2015 IRS Budget Recommendation Special Report*. Washington, DC.

Johnson, Steve R. 2016. "The Future of American Tax Administration: Conceptual Alternatives and Political Realities." *Columbia Journal of Tax Law 7*(5): 5–35.

KPMG. 2022. "IRS Opens Application Period for 2023 Compliance Assurance Process (CAP) program." *TaxNewsFlash United States*, no. 2022-262, September 15. Washington, DC: KPMG.

KPMG. 2018. "IRS Extends CAP Program, Modifying Some Rules and Signalling More Significant Changes May Lie Ahead." *What's News in Tax*. Washington, DC: KPMG.

Kuijpers, Karlijn. 2017. "Aangifte mag van de fiscus minder soepel worden." *Trouw*, 27 May.

Lederman, Leandra. 2016. "IRS Reform: Politics as Usual." *Columbia Journal of Tax Law 7*(1): 36-79.

Li, Yanwei, and Cheng Chen. "What Makes Networks Innovative? A Configurational Analysis of 18 Collaborative Innovation Cases." *International Public Management Journal* (2022): 1-22.

McCormally, T.J. 2005. "TEI Testimony Before the IRS Oversight Board." *Tax Executive*, 57.

McCubbins, Matthew D., and Thomas Schwartz. 1984. "Congressional Oversight Overlooked: Police Patrols Versus Fire Alarms." *American Journal of Political Science 28*(1): 165–179.

Moe, Terry M. 1984. "The New Economics of Organization." *American Journal of Political Science 28*(4): 739–77.

OECD. 2013. *Co-operative Compliance: A Framework*. Paris: OECD.

O'Leary, Rosemary, Catherine Gerard, and Lisa Blomgren Bingham. 2006. "Introduction to the Symposium on Collaborative Public Management." *Public Administration Review 66*(S1): 6–9.

Olsthoorn, Peter 2022. "Wordt horizontaal toezicht 2.0 wel een groot succes?" *Accountant*, 31 May.

Overman, Sjors. 2021. "Aligning Accountability Arrangements for Ambiguous Goals: The Case of Museums." *Public Management Review 23*(8): 1139–159.

Raadschelders, Jos, and Frits Van der Meer. 2014. "Administrative Elites in the Netherlands from 1980 to 2011: Making the Invisible Visible." *International Review of Administrative Sciences 80*(4): 726–45.

Romzek, Barbara S. 2000. "Dynamics of Public Sector Accountability in an Era of Reform." *International Review of Administrative Sciences 66*(1): 21–44.

Romzek, Barbara S., and Melvin J. Dubnick. 1987. "Accountability in the Public Sector: Lessons from the Challenger Tragedy." *Public Administration Review 47*(3): 227–38.

Rossotti, Charles. 2005. *Many Unhappy Returns*. Boston, MA: Harvard Business School Press.

Schillemans, Thomas, Sjors Overman, Paul Fawcett, Matthew Flinders, Magnus Fredriksson, Per Laegreid, and Martino Maggetti. 2021. "Understanding Felt Accountability: The Institutional Antecedents of the Felt Accountability of Agency-CEO's to Central Government." *Governance 34*(3): 893–916.

Scholz, John T., and B. Dan Wood. 1998. "Controlling the IRS: Principals, principles, and public administration." *American Journal of Political Science 42*(1): 141-62.

Sørensen, Eva. 2012. 'Measuring the Accountability of Collaborative Innovation." *The Innovation Journal 17*(1): 2–18.

Stevens, L, M. Pheijffer, H. van den Broek, T. Keijzer, and L. van der Hel-van Dijk. 2012. *Tax Supervision – Made to Measure*. The Hague: Committee Horizontal Monitoring Tax and Customs Administration.

Tanfani, Joseph, Matea Gold and Melanie Mason. 2013. "How the IRS spun out of control." *Los Angeles Times*, 18 May. Available via: https://www.latimes.com/nation/la-xpm-2013-may-18-la-na-irs-conservatives-20130519-story.html

TEI. 2013. "TEI's statement to the IRS Oversight Board." *Tax Executive 65*(2): 127–32.

TEI. 2012. "Minutes of a meeting between TEI delegates and the Commissioner of Internal Revenue." *Tax Executive 64*(3): 247–49.

TEI. 2007. "Minutes of the Tax Executives Institute meeting with LMSB 2007." *Tax Executive 59*(2): 199–203.

TEI. 2005. "Minutes of the Tax Executives Institute meeting with LMSB 2005." *Tax Executive 57*(2): 175–79.

Thorndike, Joseph J. 2001. "Reforming the Internal Revenue Service: A Comparative History." *Administrative Law Review 53*(2): 717–80.

TIGTA. 2013. *The Compliance Assurance Process Has Received Favorable Feedback, but Additional Analysis of Its Costs and Benefits Is Needed*. Washington, DC: Treasury Inspector General Tax Administration.

Torfing, Jacob. 2016. *Collaborative Innovation in the Public Sector*. Washington, DC: Georgetown University Press.

Van Buuren, Arwin, and Derk Loorbach. 2009. "Policy Innovation in Isolation? Conditions for Policy Renewal by Transition Arenas and Pilot Projects." *Public Management Review 11*(3): 375-92.

Wille, Anchrit. 2010. "Political–Bureaucratic Accountability in the EU Commission: Modernising the Executive." *West European Politics 33*(5): 1093–116.

Yesilkagit, Kutsal and Sandra Van Thiel. 2008. “Political Influence and Bureaucratic Autonomy.” *Public Organization Review 8*(2): 137-53.

Table 1. Cooperative compliance innovation in the US and Dutch tax administration.

	US – Internal Revenue Service (IRS)	**The Netherlands – Netherlands Tax and Customs Administration (NTCA)**
Name programme	Compliance Assurance Process (CAP)	Horizontal Monitoring (HM)
Introduction pilot	April 2005	March 2005
Pilot length	Approx. 1 year	Approx. 6 years
Enrolment on voluntary basis	✓	✓
Number of enrolled companies (2016)	181	3,840 (approx. 40% of 9,600 eligible companies)
Main cooperative compliance focused evaluations	Continuing internal reviews by the IRS of the ongoing pilot, 2005-2011	Limited internal review in 2006 by the NTCA of the pilot, focused on user experience survey amongst participating companies and NTCA's processing teams assigned to businesses
	2013 review by US Government Accountability Office (GAO), requested by US Senate Committee on Finance. Focused on impact CAP on IRS resources and collected tax revenues.	2011 review Committee Horizontal Monitoring Tax and Customs Administration (Stevens Committee), requested by State Secretary of Finance. Limited attention for impact HM on NTCA's resources and collected tax revenues.
	2013 review by Treasury Inspector General for Tax Administration (TIGTA), TIGTA's own imitative. Focused on impact CAP on IRS resources and collected tax revenues.	2017 review Investigation Large Enterprises, conducted by NTCA. Limited attention for impact HM on NTCA's resources and collected tax revenues.

Table 2. Summary of accountability dynamics in relation to cooperative compliance collaborative tax innovations.

	Dutch HM	US CAP
Definition of stakeholders	**Narrow** – focus on target group of innovation (i.e., large corporates)	**Broad** – target group of innovation considered alongside other direct and indirect stakeholders
Definition of objective(s) of the innovation	Loosely defined & changes over time	Stable & continuous focus on tax compliance and tax administration resource impact
Professional accountability system	**Weak** – mainly developed by NTCA senior leadership	**Medium** – initiated by IRS senior leadership but significant input from lower levels
Bureaucratic accountability system	**Strong** – use of hierarchical pressure by NTCA senior leadership to accelerate implementation	**Weak** – cautious approach by IRS senior leadership averts hierarchical pressure to accelerate implementation
Legal accountability system	**Weak** – limited perceived legislative pressure as NTCA senior leadership is backed by government ministers	**Strong** – substantial legislative pressure due to frequent scrutiny and financial reliance of IRS on Congress, & TIGTA oversight and IRS' agency status resulting in limited natural government backing
Political accountability system	**Strong** – pressure from opted-in companies to keep HM attractive. Limited perceived pressure from other stakeholders (e.g., media, tax profession)	**Medium** – limited perceived pressure to keep CAP attractive but some pressure experienced to not abolish CAP

Figure 1 Tax administration staff evolution IRS and NTCA, 2004-2018