How previous policy experiences affect the frontline

Understanding implementation success and failure through a general policy alienation lens

Nadine van Engen

How Previous Policy Experiences Affect the Frontline:

Understanding implementation success and failure through a general policy alienation lens

Nadine van Engen

How Previous Policy Experiences Affect the Frontline:
Understanding implementation success and failure
through a general policy alienation lens

***

Hoe eerdere beleidservaringen de frontlinie beïnvloeden:
Beleidsimplementatie- en mislukking vanuit een algemeen
beleidsvervreemdingsperspectief

Thesis

To obtain the degree of Doctor from the
Erasmus University Rotterdam
by command of the
rector magnificus
Prof.dr. R.C.M.E. Engels

and in accordance with the decision of the Doctorate Board
The public defense shall be held on

January 10, 2019

by

Nadine Alida Maria van Engen
Born in Utrecht, the Netherlands

Erasmus University Rotterdam

Promotores

Prof.dr. A.J. Steijn

Prof.dr. V.J.J.M. Bekkers

Co-promotor

Prof.dr. L.G. Tummers

Doctoral committee

Prof.dr. E.H. Klijn

Prof.dr. S.M. Groeneveld

Prof.dr. P. Mascini

Voor Ton van Engen
Ik wens iedereen zo'n vader toe

TABLE OF CONTENTS

chapter 1

Introduction

1.1 INTRODUCTION

Teachers, healthcare workers, and police officers, as well as other public employees working on the frontline of public service delivery, are often confronted with new policy programs that, usually, lead to new rules and regulations that have to be implemented. We broadly define policies as "purposive courses of actions followed by a government in dealing with a problem or a matter of concern" (Anderson, 1975, p. 3). The fact that these 'frontline workers' are often confronted with new policies is, of course, in itself, not problematic: democratically elected governments have the mandate to do so (Dunsire, 1978; Barrett, 2004). However, this can influence the way in which frontline workers perform their tasks, as grown practices may be challenged – over and over again. For the successful implementation of policies, policymakers are dependent on the willingness of frontline workers – sometimes also termed 'street-level bureaucrats', 'frontline officials' or 'public (service) employees' – to co-operate (Lipsky, 1980; Sabatier & Mazmanian, 1980; Meyers *et al.*, 1998; Hill & Hupe, 2009; Tummers *et al.*, 2009; Brodkin, 2012; Gofen, 2014; Van Engen *et al.*, 2016). The latter need to tailor the formers' policies to their clients' needs (Sommer Harrits & Ostergaard Moller, 2014), deal with conflicting demands from different policies (Tummers *et al.*, 2015), and have discretion in doing so (Lipsky, 1980). However, research has shown that frontline workers' actual behavior during policy implementation does not necessarily align with policymakers' ambitions (Brehm & Gates 1999; Maynard-Moody & Musheno, 2003; May & Winter, 2009). For instance, frontline workers may make accessing the public service more difficult for clients when work pressure is high (Tummers *et al.*, 2015), such as when a police officer telling a client who wants to report a crime: "The office is very busy today, return tomorrow if you wish" (cf. Triandafyllidou, 2003) – even though a responsive and citizen-oriented police culture may be a top priority for the government. In doing so, frontline workers can create major difficulties for (new) governments, democratically mandated to change policy and to implement new rules and regulations.

Indeed, it can be seen that frontline workers not only *can*, but also *do* create difficulties for governments, as well as societies. This can be illustrated by the following three examples. First, in 2007, 550.000 students in Israel were not receiv-

ing education, as their teachers had gone on strike to protest against a large-scale education reform (Berkovich, 2011). Second, in 2016, treatments for thousands of patients in England were disrupted when hospital doctors staged their first strike, thereby escalating political tensions over a publicly funded health care system (Castle, 2016). Third, in 2017, more than six times the average daily rate of homicides in Brazil was reported as a result of a police crisis (The Guardian, 2017).

This is problematic, because such actions may ultimately result in diminished legitimacy of the government (Bekkers *et al.*, 2007). It can cause tension and conflicts (Nutting *et al.*, 2011), and result in suboptimal circumstances for society at large, as public funds are invested in the formulation and implementation of government policies that are, apparently, not supported by frontline workers. Ultimately, this impedes the improvement of public service provision, as this is only likely to be achieved if actors operating at different levels of the system collaborate (Bryson *et al.,* 2015; Page *et al.*, 2015).

Scholars have held contradictory views on policy implementation. For a long time, policy implementation was treated as a rather mechanistic activity. Van Meter and Van Horn (1975, p. 450) even described policy implementation as "a series of mundane decisions and interactions unworthy of the attention of scholars". Top-down perspectives tend to treat deviations from the policy-on-paper as a control problem where room for interpretation makes it increasingly likely that policy means and ends will be mismatched (Howlett, 2004). Research usually has served to support a normative approach that prescribed clear policy goals and the operational steps needed to achieve them (Brodkin, 2012). However, insights changed as it became clear that implementation may lead to a reformulation of policies, to other outcomes than expected, or even to outright failure (e.g., Elmore, 1979; Lipsky, 1980; Sabatier & Mazmanian, 1980). These critiques showed that the earlier models of policymaking and public management were "not effective in practice, nor convincing in theory" (Parsons 1995, p. 468). As a result, new models were developed that emphasized the complexities of policy implementation and the prominent role fulfilled by frontline workers as a consequence of their discretion. From this bottom-up perspective, frontline workers are seen as de facto policymakers and problem solvers who adjust policies to the specific context and needs of their clients (Elmore, 1979; Brodkin, 2011; Tummers & Bekkers, 2014). In this view, "if local implementers are not given the freedom

to adapt the program to local conditions it is likely to fail" (Matland, 1995, p. 148). This underscores the importance of alignment between policymakers and policy implementers for successful policy implementation and, particularly, the relevance of evaluating how frontline workers perceive and enact government policies.

Surprisingly, the experiences of frontline workers with new policies have been often studied in isolation (e.g., Handley & Howell-Moroney, 2010; Sager *et al.*, 2014). This ignored the fact that these policies were and are not developed in a vacuum (Hogwood & Peters, 1982). For instance, studies might consider how teachers perceive a new inclusive education policy, how healthcare workers appreciate new procedures to finance healthcare, or how police officers evaluate new guidelines for criminal investigations and how this relates to their willingness to implement them. Our point is not that this focus is not meaningful or relevant for academics or practitioners – because it is. However, as we will show throughout this thesis, if we want to fully understand implementation success or failure of specific policies, we should take into account frontline workers' experiences with previous government policies. Very often, their experiences with current policies have 'a history' - because they build upon earlier experiences with other (related) policies. This process can be described as policy accumulation (In ´t Veld, 1989) where there is continuous aggregation of policies that follow upon each other. What this accumulation notion suggests is that frontline workers have a certain predisposition, with varying degrees of positivity, towards policies in general. Insights from change management studies - where terms such as 'change fatigue' and 'change cynicism' are used - show that employees' previous experiences of change affect their openness and willingness to change at a later stage (Bordia *et al.*, 2011). This implies that frontline workers could become indifferent to whatever new policy (change) is introduced, viewing new policies as just the new 'political flavor of the month' (cf. Herold *et al.*, 2007).

To systematically and coherently analyze frontline workers' experiences with specific government policies, Tummers, Bekkers and Steijn (2009) developed the policy alienation framework. Policy alienation is defined as a "cognitive state of psychological disconnection from the policy..." (2009, p. 268) and consists of two main dimensions: policy powerlessness and meaninglessness.

In the realm of policy formulation and implementation, policy powerlessness relates to the degree of influence frontline workers have (or rather lack) over shaping the policy as introduced by the government. This power may be exercised on the strategic, tactical, or operational level (Tummers *et al.*, 2009). Strategic powerlessness refers to the perceived influence of professionals on decisions concerning the content of a policy, as captured in rules and regulations at the government level. Tactical powerlessness refers to professionals' perceived influence over decisions concerning the way a policy is executed within their own organization. Operational powerlessness relates to the degree of discretion professionals have during actual policy implementation. Examples include the degree to which doctors and nurses have the impression that they can influence healthcare reforms at the national level, or the degree to which teachers feel that they have discretion during the implementation of inclusive education.

Policy meaninglessness refers to the perception of the contribution a policy makes to a greater purpose. Two types of meaninglessness are distinguished: societal and client meaninglessness. The former refers to the perception of professionals concerning the added value of policies to socially relevant goals. The latter reflects the perception of professionals regarding the value added for their own clients. Examples include the degree to which police officers believe that the instalment of a national DNA database results in an increased number of solved crimes, or the degree to which youth care workers believe that the decentralization of youth care is beneficial for their young clients.

1.1.1 Main research question

However, as we discussed above, it is relevant to investigate not only how frontline workers experience specific policies, but also how they identify with *government policies in general*. We argue that not taking the latter into account might result in a failure to understand why the implementation of new government policies is, or is not, supported by frontline workers. We can illustrate this with an example. Elizabeth and Jack are both secondary school teachers and both confronted with a new government policy that aims to tackle the growing educational inequality in their country. When asked, both Elizabeth and Jack indicate that they support the new policy's goals. They believe that the funding, as well as the training opportunities, are sufficient and they have considerable discretion when imple-

menting the policy, and feel self-confident in doing so. However, it turns out that Jack acts in line with the policy and contributes to make its implementation a success, whereas Elizabeth does not. Why? Not because they differ in terms of their support for the specific policy, but because Jack identifies with government education policies in general, whereas Elizabeth does not. Overall, Jack supports government education policies. He has the impression that they address relevant problems, that they allow school leaders and teachers to have a say in their set-up, and that they leave room for tailored implementation at the school level. Elizabeth, on the other hand, is much more critical of government education policies. She feels that policies do not tackle urgent problems – and if they do, they do so often in an ineffective way. She feels school leaders and teachers are not listened to during policy formulation, and that there is little discretion for school leaders and teachers during implementation. This, their 'policy predisposition', influences Jack and Elizabeth's willingness to implement a new policy – regardless of their evaluation of the newly introduced policy in itself.

The policy alienation framework (Tummers, 2012) does not take the effect of the accumulation of previous experiences into account and does not allow for the evaluation of general government experiences. Therefore, in this thesis, we investigate how we can further develop the framework to take this effect into account and how this can be helpful to analyze the effect of previous policy experiences on frontline workers. In doing so, we introduce the term 'general policy alienation', based on distinctions made in the literature between general and specific trust (Kenning, 2008) and general and specific self-efficacy (Schwoerer *et al.*, 2005). Whereas we use the term policy alienation to refer to experiences with specific policies, we use the term general policy alienation to refer to frontline workers' experience with overall government policies. Besides further developing the policy alienation framework, we first investigated which factors might influence general policy alienation. In this way, this thesis aims to provide more insight in the role factors, such as policy consistency, policy discretion, and policy accumulation, may play in policy support. Second, we investigated the influence of general policy alienation on implementation willingness. Based hereon, this thesis aims to provide more insight regarding the extent to which general policy perceptions, in interaction with evaluations of specific policy characteristics, might influence how willing frontline workers are to implement new policies. As

such, our study connects to broader debates on policy implementation, policy legitimacy and alignment, and discretion at the frontline (e.g., Wallner, 2008; Hupe & Hill, 2009; Brodkin, 2012, Gofen, 2014; Tummers *et al.*, 2015).

Summarizing, the main goal of this study was to analyze whether and to what extent frontline workers experienced general policy alienation, but also why this was the case and what the implications are for frontline workers' implementation willingness. Therefore, the central research question of this thesis is:

How can the general policy alienation of frontline workers be conceptualized and measured, what are its causes and what is its influence on implementation willingness?

1.1.2 Subsidiary research questions

To answer the main research question, a number of subsidiary research aims were formulated.

First, we aimed to conceptualize and measure general policy alienation. As outlined above, we developed this concept in order to allow for the conceptual distinction between frontline workers experiences with specific and overall government policies. First, we investigated how we could conceptualize general policy alienation, and how we could clearly distinguish general from specific policy alienation. Providing a straightforward conceptualization is especially relevant in light of our second research aim: measuring general policy alienation. Hinkin (1998) stated that if you do not provide a clear conceptualization, you would end up with a scale that is not valid: it does not capture the phenomenon you aim to measure. We were interested in developing a measure of general policy alienation. This would allow us to quantitatively test its relationship with other variables, including, as we discuss below, policy consistency and implementation willingness. In doing so, we adopted a systematic approach to develop valid and reliable scales for general policy alienation, because "The point is not that adequate measurement is 'nice'. It is necessary, crucial... Without it we have nothing" (Korman, 1974, p. 194).

Second, we aimed to further investigate the effects of policy accumulation on how frontline workers perceive and implement policies. Hogwood and Peters (1982) noted that in the study of policymaking and policy analysis, scholars often

speak of creation, birth, and innovation, as though policies came new into the world. In reality, they argued, new policies are rarely written on a tabula rasa, but rather on a well-occupied or even crowded tablet of existing laws, organizations and clients. Pollitt and Bouckaert (2011, p. 8) even stated, "the detail of public sector reforms turns out to be more like geological sedimentation, where new layers overlie but do not replace or completely wash away the previous layer". However, we do not argue that the accumulation of policies is, in itself, problematic for frontline workers. Rather, we believe that particularly the degree to which policies are consistent – both over time and in relation to each other – influences how frontline workers perceive policies. When frontline workers have to decide whether or not to put effort into implementing a new policy, their government's past performance in maintaining their policies becomes an important consideration (cf. White *et al.*, 2013). Furthermore, frontline workers can suffer from a status quo bias, i.e. they have a preference for policies as they currently are (Arnold & Fleischman, 2013). This argues in favor of a positive effect of policy consistency on frontline workers' policy perceptions, including how meaningful and legitimate policies are. So far, this has not been tested empirically.

Third, we aimed to evaluate how general policy alienation influences frontline workers' implementation willingness. The topic of discretion continues to be debated in policy design and policy implementation (Tummers & Bekkers, 2014; Veronesi & Keasey, 2015; Cooper, 2017). However, little attention has been paid to the implicit link assumed between frontline workers' discretion and the motivation to implement government policies. To explore the motivational effects of discretion, we drew on the logic of the Thomas theorem: "If men define situations as real, they are real in their consequences" (Thomas, 1928). Hence, we focused on the perceived degree of discretion, and investigated to what extent frontline workers feel powerful (i.e., the opposite of powerlessness). Although psychological and change management literature suggests a positive link between powerfulness and motivation (Lines, 2004; Gagné & Deci, 2015), scholars have not found a strong, consistent symmetric relation between policy powerfulness and implementation willingness (Tummers, 2011; Thomann, 2015). Therefore, it might be that an asymmetric explanation is more suitable, which we tested empirically.

Summarizing, the three subsidiary research questions are:

1. How can we conceptualize and measure frontline workers' general policy alienation?
2. What are the effects of policy consistency on frontline workers' evaluations of meaningfulness and legitimacy?
3. How does (general) policy alienation influence frontline workers' implementation willingness?

1.2 CASE STUDY: THE DUTCH SECONDARY EDUCATION SECTOR

Having introduced our main research questions, we now discuss our case study: the education sector. Research has indicated that frontline workers in this sector experience many problems with public policies. More specifically, school leaders and teachers experience many problems with national education policies. This is particularly relevant because they play a crucial role in delivering services. In 2016, for instance, over 12.000 teachers in Poland demanded not only a pay rise and retention of early retirement privileges, but also the dismissal of the Minister of Education (NOS, 2016). In 2015, stress levels among teachers in England were soaring: a survey among 3.500 members of a teaching union showed that more than two-thirds of respondents considered quitting the profession with their top-concerns being: work load, pay, inspection, and curriculum reform (Precey, 2015). In 2015, over 5.000 teachers in Seattle in the US started a strike. This strike was motivated by more than just by a wage dispute, as the following teacher's quote shows: "In spite of your portrayal of this being all about salary, it is about much, much more – it's about properly funding schools, respecting educators, giving kids the recess time that research shows they need, reducing severely over-crowded classrooms, dealing with racial inequity in our schools, making up for years without adequate cost-of-living adjustments in the state's most expensive city, and so much more" (Young, 2015).

The specific case we studied is the Dutch secondary education sector. This case is relevant for three reasons. First, the sector has experienced many problems in recent decades as a result of the reshuffling of authority and responsibili-

ties between the ministerial and the school levels (Pijl & Frissen, 2009). Second, the sector is characterized by numerous policy changes (Bronneman-Helmers, 2008). Third, research has shown that many school leaders and teachers are critical of government-initiated reform (Tweede Kamer, 2008). This makes it a suitable case to investigate policy implementation challenges, the consequences of policy accumulation and antecedents and effects of general policy alienation in-depth. This is illustrated by the following three quotes:

> "*Annoying are the continuous changes and additional tasks. A perfect example is the introduction of an obligatory social internship for all high school students. We embraced this policy, invested many of our funds in it, and really saw its added value. Then, the obligation was withdrawn, as well as the government funding. This, in my opinion, rewards schools that act negligent. As a result of this, when new policies are introduced by the government and you do want to loyally implement them – you eventually start thinking: Why would we?*"
>
> - School leader, interviewed for this thesis (2013)

> "*The maths test* [a new, obligatory maths test introduced by the Dutch government] *once again shows that Dutch education is unmanageable. Politicians play angry bosses who shout 'SIT' to their dogs. And whisper immediately afterwards: 'Okay, keep on lying then.' Schools know this, nod yes and do nothing. Consequence? The results of policies are always the opposite of the goals.*"
>
> - Ton van Haperen, Dutch education blogger (2015)

> "*What is wrong? The ink of a policy letter isn't dry yet, or the next one is on its way. Changes do not get the chance to get 'crystallized.' The teacher is the professional… but…*"
>
> - Teacher, surveyed for this thesis (2016)

The Dutch Ministry of Education, Culture and Science is responsible for the education system, its financing, and education quality. It formulates policies, takes measures and specifies certain standards, such as the content and quality

of final exams. Within this context, schools have to take responsibility themselves (Onderwijsraad, 2012). Schools and the Ministry are intrinsically connected and strongly dependent on each other. Both are intertwined by politics. However, over the last decennia, it seems a 'gap' has grown between these actors. Schools, in particular, have the impression that the Ministry and politics (i.e., the government) do not understand them. In the 2008 Dutch parliamentary research on education reform, the research commission (Commission Dijsselbloem), for instance, concluded that "political support seemed to be more important than the support of schools" (Tweede Kamer, 2008, p. 177) and that "the support of representatives of professional interest groups was equated with support of schools, while teachers, parents and students were hardly listened to during the policy process" (p. 188).

The Ministry of Education, Culture and Science has the political responsibility for the educational system and is bound by national legislation. The Ministry is responsible to a large extent for the financing of the education system. It also defines the general education policy and specifies the admission requirements, structure and objectives of the education system on general lines (EP-Nuffic, 2015). The Dutch education system consists of eight years of primary education, four, five or six years of secondary education (depending on student capacities) and two to six years of higher education (depending on the type of education and the specialization).

Unique to the Dutch system is its duality and the freedom of education. This freedom of education is a concept included in the Dutch constitution, article 23. As a consequence of 'article 23', the Dutch government provides the same financial support to public and private schools, as long as the schools meet certain basic quality and financial standards (EP-Nuffic, 2015). Article 23 also specifies the relationship between the government and school organizations. On the one hand, the government has the task to take active care of the education system as a whole. On the other hand, the government has to give discretion to schools, as all Dutch schools have the freedom to be organized according to their own convictions and ambitions. In the school year 2015-2016, approximately 960.000 Dutch students between the age of 12 and 18 followed secondary education. This is, depending on student capacities, either a preparatory secondary vocational education (four years), senior general secondary education (five years), or university

preparatory education (six years). They follow their education at 700 different schools at approximately 1.400 school locations (DUO, 2017). In total, there are almost 75.000 teachers working in Dutch secondary education and almost 3.100 school principals.

When we conducted our study, between 2013 and 2017, the cabinet of the Netherlands was the 'Rutte-Asscher cabinet', formed by the liberal People's Party for Freedom and Democracy ('VVD') and the social democratic Labour Party ('PvdA'). The Minister of Education was a PvdA member, and the Secretary of State (i.e., Junior Minister) for Education was a VVD member. The Secretary of State was responsible for secondary education, and introduced multiple policies, including a teacher development agenda, anti-bullying policy, and policies to stimulate excellence in secondary education. There is mixed empirical evidence on how positive secondary school teachers and principals were about these politicians and their policies. Our own study results, for instance, indicate that our respondents had relatively low trust in these politicians and were quite critical towards their policies.

1.3 RELEVANCE OF THE THESIS

Having introduced our main research interests, research questions and our case study, we now discuss why and how this thesis is relevant from both an academic and practitioner perspective.

1.3.1 Academic relevance

Our aim was for this study to contribute theoretically and methodologically to the public administration literature.

Theoretically, we aimed to contribute to knowledge on policy implementation and street-level bureaucracy by introducing the concept of general policy alienation and highlighting the importance of policy history. Although the literature recognizes the important role of frontline workers for policy implementation (Lipsky, 1980; Freidson, 2001; Bekkers *et al.*, 2007), public administration and management research still tends to marginalize the perspectives and experiences of those who enact the policy in practice (O'Toole, 2000; DeLeon & DeLeon,

2002; Barrett, 2004; Saetren, 2005; Werts and Brewer, 2015) and, particularly, the micro-level (psychological) underpinnings of this (cf. Grimmelikhuijsen *et al.*, 2017) - however, see Andersen and Jakobsen (2017) and Raaphorst (2018a) for recent, notable exceptions. This is peculiar, since policy implementation is often complex, contradictory, and still one of the main challenges for civil servants worldwide (O'Toole, 2004). Therefore, this matter deserves the ongoing attention of public administration and management scholars.

Our first contribution was to introduce the concept of general policy alienation and, thereby, acknowledge that frontline workers bring with them a history of government policy (changes), and, hence, general ideas about their effectiveness, legitimacy, and meaningfulness. In doing so, we emphasized that frontline workers were not neutral implementers. By studying policy experiences in relation to their historical context, we extend the theoretical work on policy accumulation (In 't Veld, 1989) and related concepts, such as policy succession (Hogwood & Peters, 1982), and institutional layering (Thelen, 2004).

Our second contribution was to investigate how frontline workers' policy perceptions might be improved. It has been noted "the cataloguing of failures when putting policies in place has been the hallmark of implementation studies since the 1970s" (May, 2015, p. 277). We, on the other hand, propose and show that policy consistency may contribute to improved policy perceptions of frontline workers by relying on work emphasizing the benefits of a rational policymaking perspective (e.g., Dunsire, 1979), as well as mostly political research on the status quo bias of civil servants (e.g., Fleming *et al.*, 2010).

Methodologically, we contribute by adopting relatively new and methodological approaches. First, we conducted quantitative street-level bureaucracy research. Traditionally, this type of research has been quite qualitative (Pressman & Wildavsky, 1973; Lipsky, 1980; Sandfort, 2000, Maynard-Moody & Musheno 2003; Hill & Hupe, 2009). In this regard, it is not surprising that it has been noted "making the study of street-level bureaucracy both generalizable and comparative is an issue in its own right" (Hupe *et al.*, 2015, p. 326). Our quantitative approach allows for the large-scale testing of relevant theories and assumptions and, thereby, complements previous qualitative research (Van Engen, 2019). For instance, we adopt an experimental approach to investigate the effects of policy consistency on frontline workers' policy perceptions. Although experiments

often manipulate situations (i.e. situations are not 'real', which limits ecological validity), they also allow us to isolate and explore causal effects of interest in ways that other methods cannot (Grimmelikhuijsen *et al.*, 2017; James *et al.*, 2017). By doing so, we contribute to the emerging tradition of a ´behavioral public administration´ (Grimmelikhuijsen *et al.*, 2017). Second, we developed two valid and reliable measures: a measure of general policy alienation (26 items) and a short measure of general policy alienation (5 items). It has been noted that the field of public administration lags behind as compared to other social sciences in the development of measurement scales (Perry, 2016). We proposed and used systematic procedures that, we hope, are helpful to researchers who aim to develop scales themselves. A greater emphasis on measurement, that we have contributed to, can help street-level bureaucracy and implementation research in making inferences that are also comparable across studies and contexts (Grimmelikhuijsen *et al.*, 2017).

1.3.2 Practical relevance

The discussion regarding the gap between research and practice in public administration and policy implementation research has never lost its salience (O'Toole, 2004). As noted by recent scholars, creating lasting and dynamic evidence-based policymaking systems requires a long-term commitment by both researchers and policymakers (VanLandingham & Silloway, 2016). Particularly, progress towards evidence-informed policymaking requires both improving the supply of research that is reliable, timely, and relevant to the policy process. In this light, as stated above, it is surprising that public administration and management research still tends to marginalize the perspectives and experiences of those who enact the policy in practice (O'Toole, 2000; DeLeon & DeLeon, 2002; Barrett, 2004; Saetren, 2005; Werts & Brewer, 2015).

Our study focused on policy implementation and contributes to the public administration practice in three ways. Firstly, our research may help national and local policymakers – basically, all (government) actors involved in policy implementation – to understand better why the implementation of their policies succeeds or fails. This applies, in particular, to sectors where governments rely heavily on frontline workers to achieve their intended policy changes. This includes: the healthcare sector (where governments rely on medical doctors and

nurses); the safety sector (where government rely on police and military); and, the education sector (where governments rely on school board governors, school leaders and teachers). Secondly, we developed measurement scales that can be used by policymakers or applied policy researchers to comprehensively (long, 26-item measure) or efficiently (short, 5-item measure) analyze how frontline workers experience government policies, also over time. If changes occur, or frontline workers indicate they feel extremely alienated, this may call for the introduction of appropriate interventions. In this way, this monitoring might help to improve the policy implementation process. Taking the experiences of frontline workers seriously may be a helpful tool to improve the relationship between policymakers and policy implementers. Thirdly, the practical recommendations we postulate – based on our research results – provide quite straightforward suggestions for politicians, public managers and civil servants to strengthen their policy implementation.

1.4 OUTLINE OF THE THESIS

The three subsidiary research questions introduced in this chapter are addressed in four empirical chapters as summarized in Table 1.1 (on the next page). The columns refer to the chapters in this thesis and the specific research question to which they relate, as well as the data sources, and the methods applied. These aspects are explained in more detail below. Please note that the empirical chapters of this thesis were originally written as independent journal articles and can be read separately. As a consequence, there is some overlap of ideas between the chapters.

1.4.1 Empirical chapters
The following provides a brief abstract of all the empirical chapters.

Chapter 2. Taking previous policy experiences into account: Conceptualizing and measuring general policy alienation
To explicitly take account of frontline workers' previous experiences with government policies, we introduced the concept of general policy alienation. This

is defined as an overall cognitive disconnectedness from government policies, and conceptualized with two main dimensions: policy powerlessness and policy meaninglessness. Building on the policy alienation framework of Tummers (2012), we developed a valid and reliable 26-item measure of general policy alienation. This measure consists of five dimensions: strategic, tactical and operational powerlessness (six items each), and societal and client meaningfulness (four items each). In line with our assumptions, we found a relationship between general policy alienation, specific policy alienation (i.e., towards specific policy programs), policy consistency, transformational school leadership and, finally, implementation willingness.

Table 1.1 Outline of the empirical chapters

RQ		Chapter		Data source	Method	Published
1	How can we conceptualize and measure frontline workers' general policy alienation?	2	Taking previous policy experiences into account: Conceptualizing and measuring general policy alienation	Teachers (*N=1.096*)	Exploratory and confirmatory factor analyses, structural equation modeling	*Public Management Review* (2016)
		3	Developing a short measure of general policy alienation	School leaders and teachers (*N=1.183; N=354; N=933*)	Exploratory and confirmatory factor analyses, structural equation modeling	*Public Administration* (2017)
2	What are the effects of policy consistency on frontline workers' evaluations of meaningfulness and legitimacy?	4	Determining whether consistent government policies lead to greater meaningfulness and legitimacy on the frontline	Teachers (*N=779*)	ANOVA, ANCOVA, t-tests, regression analyses	*Public Administration* (2018)
3	How does (general) policy alienation influence frontline workers' implementation willingness?	5	Evaluating how powerfulness and meaningfulness influence implementation willingness	Teachers and healthcare workers (*N=1.087; N=1.004*)	Large-N set-theoretic configurational analysis	*Journal of Public Administration Research and Theory* (2018)

Chapter 3. Developing a short measure of general policy alienation

Given the limited questions that can be included in a typical survey, the 26-item length of the measure could have limited use for research. Therefore, one main goal of this study was to develop a valid and reliable short measure of general policy alienation. First, this frees up survey time researchers can then use to measure additional variables (Liden *et al.*, 2015). Second, an overload of items can introduce fatigue, or even boredom, among respondents, which may negatively influence the quality of the responses obtained (Crede *et al.*, 2012). Third, a short measure is more likely to be useful for other fields of public administration where general policy alienation is not the core subject matter, but could form a relevant contextual or explanatory factor. However, short measures may compromise validity if not developed using rigorous methods. This problem can be managed by applying the stringent 10-step approach we developed based on guidelines by, among others, DeVellis (2012) and Smith *et al.* (2000).

Chapter 4. Determining whether consistent government policies lead to greater meaningfulness and legitimacy on the frontline

Research has shown that frontline workers actual behavior during policy implementation does not necessarily align with policymakers' ambitions (Maynard-Moody & Musheno, 2003; May & Winter, 2009). This can be problematic for (new) governments democratically mandated to introduce new policies. Therefore, it is important to understand better how frontline workers, as well as other stakeholders, perceive and experience their policies over time. In this study, we focused on the effects of policy consistency. In other words, we studied how the continuity, certainty and predictability of policies over time influenced frontline workers. Specifically, we investigated with a survey experiment how policy (in)consistency affects perceived policy meaningfulness and government legitimacy. We also took into account the fact that this relationship may depend on policy content. Furthermore, given the apparent importance of discretion for frontline workers (Lipsky, 1980; Maynard-Moody & Musheno, 2003), we also investigated if, and how, this effect is moderated by the degree of discretion policies allow for.

Chapter 5. Evaluating how powerfulness and meaningfulness influence implementation willingness

It has been shown repeatedly, that frontline workers have an important role in the successful implementation of policies as, inevitably, they retain some degree of discretion (Davis, 1969). However, little attention has been paid to the implicitly assumed link between frontline workers' discretion and the motivation to implement government policies. This is surprising, given that "research performed in ignorance of the understanding that implementing actors have about their circumstances is likely to miss important parts of the explanation" (O'Toole, 2000, p. 269). Therefore, in this study, we aimed to further disentangle the relationship between policy powerfulness, meaningfulness, and implementation willingness. Contrary to the other empirical chapters, we relied on an asymmetric explanation of policy implementers' motivation and, accordingly, investigated whether we could find evidence for an asymmetric relation between powerfulness and implementation willingness.

1.4.2 Data sources

Multiple data sources were used to conduct this study. Below, we briefly describe our data. More information is provided in the relevant chapters.

First, we collected large-scale survey data in June 2013 and June 2016. The respondents were members of a larger voluntary panel of public employees ('Flitspanel'), funded by the Dutch Ministry of the Interior and Kingdom Relations. In total, the panel consisted of 35.000 Dutch public sector employees. To ensure the representativeness of the panel, the members were selected through the records of the pension fund for all Dutch government employees ('ABP') – in which all public employees are legally obliged to participate (for more information on the panel see http://www.internetspiegel.nl; other recent studies making use of this panel are Van Loon *et al.*, 2016 and Van der Voet & Vermeeren, 2017). The 2013 sample consists of 1.183 school leaders and teachers. The 2016 sample consists of 993 school leaders and teachers. In some studies school leaders were excluded from the sample based on the research aims of the respective empirical chapters.

Second, to conduct the studies reported in chapter 2 and 4, we partially used secondary data. Firstly, 'Regioplan', a Dutch independent research organization, collected one dataset used for chapter 2. Survey data were collected in 2015 as

part of a policy evaluation study, conducted at the request of the Dutch Ministry of Education, Culture and Science. This sample consists of 354 school leaders and teachers. Secondly, Tummers (2012) collected one dataset used for chapter 5. Survey data were collected in 2010 to analyze, among others, frontline workers experiences with a new policy program. This sample consists of 1.317 healthcare professionals.

Open access

Transparency and reproducibility are key to good science. Two ingredients are essential for reproducibility in any field in science, namely: full disclosure of the methods used to obtain and analyze data; and, availability of the data that went into and came out of the analysis (Open Science Collaboration, 2015). To adhere with transparency and reproducibility guidelines, we will make the two datasets we collected for this research publicly available via dataverse. Please note that *all data we provide is fully anonymized and cannot be traced down to individual respondents*.

We believe that making our data publicly available will allow both researchers and practitioners to (re)use our data for academic, practical, as well as educational purposes. In addition, researchers can use our data to investigate other relationships than those reported in this thesis - not all data we collected were used. Academic teachers can use our data in statistical courses, so that students can use 'real life' data to learn how to conduct descriptive statistics or regression analyses. Also, survey respondents can experience survey fatigue due to overexposure to surveys. This is considered a main cause of increasing nonresponse (Steeh, 1981). By allowing other researchers to make use of our data, we hope to contribute to limiting questionnaires Dutch school leaders and teachers are confronted with.

1.4.3 Methods

We applied multiple methods and used multiple statistical programs to conduct our study.

First, we developed a 26-item measure of general policy alienation. Here, we applied exploratory and confirmatory factor analyses to evaluate the factorial structure of the proposed measure. Then, we tested the construct and conver-

gent validity of the scale with structural equation modeling. To evaluate the convergent validity of our measure, we investigated the relationships between the five general policy alienation dimensions, the perception of a specific policy program, policy consistency, transformational leadership, and implementation willingness.

Second, we developed a short, 5-item measure of general policy alienation using a systematic 10-step procedure we developed based on guidelines by, among other, DeVellis (2012) and Smith *et al.* (2000). First, we evaluated face validity and reviewed our item pool with experts. Second, we evaluated the internal consistency reliability. Third, we applied exploratory and confirmatory factor analyses to evaluate the factorial structure of the proposed measure. Furthermore, we conducted multi-group confirmatory factor analyses and tested whether our proposed short measure has measurement invariance, also known as measurement equivalence, across groups (Byrne, 2008). Finally, we evaluated convergent and discriminant validity. We evaluated the former by relating our short measure to the perception of a specific policy program, policy consistency, implementation willingness, and trust in government, and the latter by relating our short measure to the number of students at a school and whether a school was publicly or privately owned.

Third, we conducted a survey experiment to investigate the effect of policy consistency versus policy inconsistency on frontline workers perceptions of meaningfulness and legitimacy. Here, we applied ANOVA and ANCOVA tests to evaluate whether – in line with our hypotheses – respondents randomly assigned to the consistency manipulation score higher on meaningfulness and legitimacy than respondents randomly assigned to the inconsistency manipulation. We conducted regression analyses to further understand the effect of policy consistency. Here, we not only investigated the direct effect of policy consistency on meaningfulness and legitimacy, but also if, and how, this effect is moderated by discretion and by policy content.

Fourth, we used large-N set-theoretic configurational analysis, formal theory evaluation and comprehensive robustness tests (Ragin, 1987, 2000; Schneider & Wageman, 2012) to investigate the relationship between powerfulness and meaningfulness and implementation willingness – which, in chapter 5, we as-

sumed was asymmetric and non-linear. Set-theoretic configurational methods are designed to access such relations in term of necessity and sufficiency.

As stated above, we used multiple statistical programs for the analyses. We used version 6 of the statistical program *Mplus* (Muthen & Muthen, 2012). We used the statistical program *R*, specifically the *R*-packages 'lavaan' (Rosseel, 2012), 'psych' (Revelle, 2015), and 'semTools' (semTools Contributors, 2016) and *R*-packages 'QCA' (Dusa, 2007) and 'SetMethods' (Medzihorsky *et al.*, 2017). For the majority of the descriptive statistics, we used version 21-24 of the statistical program *SPSS.*

chapter 2

Taking previous policy experiences into account: Conceptualizing and measuring general policy alienation

ABSTRACT

Research mainly looked at problems frontline workers[1] have with specific policy programs. However, policies are not developed in a vacuum. Frontline workers are often confronted with (a series of) policy changes, intended to refine, replace or complement other policies. This policy accumulation results in frontline workers having a certain predisposition towards policies in general. To conceptualize this predisposition, we introduce the term general policy alienation. We investigate whether the earlier developed policy alienation scale can be adapted to measure general policy alienation. Our analyses show that the scale performs satisfactorily. Theoretical relevance, as well as directions for practical applications are discussed.

---

1   Please note that in the original article (Van Engen *et al.*, 2016), we apply the term 'public professionals' instead of 'frontline workers'. To increase readability, we apply the term 'frontline workers' throughout this whole thesis, including this chapter.

2.1 INTRODUCTION

"*This strike is about much more than the compulsory seven and a half hours teachers should spend daily at school.*"

- President of the Norwegian Teachers Union

(Education International, 2014)

This quote illustrates that frontline workers who regularly work on the frontline of public administration (such as teachers), where they interact directly with citizens, are confronted with government policies that they do not always support. Moreover, they have an important role in the success of these policies given their discretion during implementation (Lipsky, 1980; Maynard-Moody & Musheno, 2003; Hupe & Hill, 2007; Gofen, 2014). As such, their support influences the effectiveness and legitimacy of government policies (Freidson, 2001; Bekkers *et al.*, 2007).

Surprisingly, the experiences of frontline workers with new policies are often studied in isolation (e.g., Handley & Howell-Moroney, 2010; Sager *et al.*, 2014), ignoring the fact that these policies are not developed in a vacuum (Hogwood & Peters, 1982). Very often, these experiences have a history because they build upon earlier experiences with other related policies. This process can be described as policy accumulation (In ´t Veld, 1989): the continuous aggregation of policies that follow each other. What this accumulation notion suggests is that frontline workers have a certain predisposition, with varying degrees of positivity, towards policies in general. Insights from change management studies - where terms such as 'change fatigue' and 'change cynicism' are used - show that employees' previous experiences of change affect their openness and willingness to change at a later stage (Bordia *et al.*, 2011). The same mechanism may also apply to frontline workers and their receptivity of new policies, and this will also influence the effectiveness and legitimacy of these policies. If we want to increase our understanding of the influence of this policy predisposition, we first have to conceptualize it and, second, have to operationalize and measure it.

Tummers, Bekkers, and Steijn (2009) proposed a policy alienation framework to systematically analyze whether public actors identify with a specific policy. However, the framework does not take the accumulation of previous experiences

into account. Here, we introduce the term *general policy alienation* based on distinctions made in the literature between general and specific trust (Kenning, 2008) and self-efficacy (Schwoerer *et al.*, 2005). To demonstrate that this is a phenomenon that frontline workers genuinely experience, we return to the quote that introduced this article: 'This strike is about much more than the compulsory seven and a half hours...'. This strike by Norwegian secondary school teachers *did* start as a reaction to the introduction of a new controversial government proposal, but the strike was about more than that. Months before the strike started, the Norwegian teachers had already voted against another government proposal because they perceived it as a threat to their professional autonomy and their ability to deliver high quality education. The later attitude of these Norwegian teachers was therefore in line with our conceptualization of general policy alienation as a state of mind reflecting accumulated past policy experiences. Alongside conceptualizing general policy alienation, we also investigate whether an adapted version of the previous policy alienation scale (Tummers, 2012) can be used to assess frontline workers' general perceptions of government policy, thereby helping in the analysis of the effect of frontline workers' past policy experiences. By taking history into account, this would contribute to a more realistic and context-sensitive approach when studying policy implementation.

This article is structured as follows. In the first part, we discuss the existing theory on policy accumulation and policy alienation. The second part presents the empirical component of this study based on data from a survey among 1.096 Dutch secondary school teachers. Here we report the steps taken in the development of a reliable and valid measurement scale, including exploratory and confirmatory factor analysis, and internal and convergent validity tests. After discussing the results, we conclude by evaluating our contribution to the policy implementation literature. Finally, we discuss how our results can benefit public administration scholars and practitioners in their continuous quest to improve public service delivery.

2.2 THEORETICAL FRAMEWORK

2.2.1 Policy accumulation

When studying policies, history matters (Pierson, 2000). A policy's past should therefore not be ignored. Hogwood and Peters (1982) noted that scholars often speak of creation, birth, and innovation as though policies come new into the world. In reality, they argued, new policies are rarely written on a clean slate, but rather on a well-occupied or even crowded tablet of existing laws, organizations, and clients. Policies fit within a certain tradition of policies and policy changes. Attention has also been paid to this notion of history in studies of institutional change. Here, Thelen (2004) introduced the concept of institutional layering to explain transformation as a process in which new elements are attached to existing institutions, thereby gradually changing their status and structure. The institution is not replaced, but new layers, such as policies, policy processes, actors, or rules, are added to it. Pollitt and Bouckaert (2011, p. 8) commented that "the detail of public sector reforms turns out to be more like geological sedimentation, where new layers overlie but do not replace or completely wash away the previous layer". The introduction of a new policy is thus shaped by interactions with the pre-existing policies it is intended to either specify, replace, or complement as it adapts to unanticipated implementation circumstances and evolving political needs (Van Gunsteren, 1976; Wildavsky, 1979). The term 'policy accumulation' is used to refer to these processes (In 't Veld, 1989). Due to this accumulation process, frontline workers will have a certain predisposition towards policies in general, and this will affect their receptivity towards new policies.

2.2.2 General policy alienation

Tummers, Bekkers, and Steijn (2009) conceptualized policy alienation in order to systematically and coherently analyze why frontline workers do, or do not, identify with government policies. Policy alienation is defined as "a cognitive state of psychological disconnection from the policy program being implemented by a public professional who regularly interacts directly with clients" (Tummers *et al.*, 2009, p. 688). They distinguished two main dimensions of policy alienation: policy powerlessness and policy meaninglessness. In this chapter, we make a conceptual distinction between frontline workers' specific policy alienation

(disconnection from a specific policy program) and general policy alienation (an overall disconnect from government policies).

We first need to define the terms 'profession' and 'frontline workers'. However, distinguishing professions from non-professions has proven difficult. Several authors have argued that professionals must have specific knowledge and do certain things to be professional (content), and they must be part of a professional association (control) to acquire content and be regarded as professionals with special privileges (Elliot, 1972; Abbott, 1988; Freidson, 2011). Others, such as Etzioni (1969), proposed a distinction between professions and semi-professions. The latter referring to professions with limited discretion and decision-making responsibility. In light of our research topic, we use a fairly broad definition of professions offered by Gabe, Bury, and Elston (2004, p. 163): "to describe an occupation as a profession may be simply to identify it as a particular kind of occupation, typically one with high status and high rewards, requiring long formal training and delivering a personal service". In line with this, a semi-profession is then an occupation without high status and high rewards. We subsequently define frontline workers as employees working in professions (such as medical doctors) and semi-professions (such as teachers or social workers) in the public sector. With this definition, we want to emphasize that our research is relevant for understanding both professionals' and semi-professionals' experiences with national policies. For our study, two distinguishing characteristics of frontline workers are relevant. The first is that these frontline workers are responsible for implementing and thereby defending the policies of the government (Lipsky, 1980; Maynard-Moody & Musheno, 2003; Hupe & Hill, 2007). The second is that, in doing this, they have a certain degree of discretion in their regular interactions with citizens (Sandfort, 2000; Brodkin, 2011).

In the realm of policy formulation and implementation, policy powerlessness relates to the degree of influence frontline workers have (or rather lack) over shaping a policy program. This power may be exercised on the strategic, tactical, or operational levels (Tummers *et al.*, 2009) where it influences, respectively, the national level, the organizational level, and the actual policy implementation. The second dimension of policy alienation is meaninglessness. In the context of policymaking and implementation, meaninglessness refers to frontline workers' perceptions of the contribution a policy makes (or fails to make) to some greater

purpose. Societal meaninglessness refers to the perception of frontline workers concerning the value that policies add to socially relevant goals (Tummers *et al.*, 2009). For instance, frontline workers may perceive a policy program as not actually providing desirable public services or outcomes, such as improved educational quality. Client meaninglessness reflects frontline workers' perception of the value added for their own clients. If frontline workers perceive that they are not helping their clients by implementing certain policies, this amounts to a high level of client meaninglessness. The latter should logically be most pertinent to public servants such as teachers who have direct working relationships with citizens (as clients) and we use the term 'frontline' to refer to those in such a relationship.

The policy alienation framework has primarily been used to analyze frontline workers' experiences with single policies. In this study, however, we focus on general policy alienation. Do frontline workers have the impression that they can, in general, influence the shaping of government policies? Further, do they have the impression that government policies are, in general, meaningful and add value for society as a whole and for their own clients? As with specific policy alienation, general policy alienation can be conceptualized using five dimensions. We conclude this section by summarizing and defining these dimensions in Table 2.1 (on the next page). This table also shows, for each dimension, the definition of specific policy alienation in order to clarify the distinction between the two concepts. Further, an example is provided of each dimension.

Here, we should emphasize that we are not claiming that the way frontline workers respond to new policies is dependent only on their alienation towards a specific policy or their general policy alienation: other factors are also relevant. These include the influence of professional culture and organizational socialization (Oberfield, 2010; Hatmaker *et al.*, 2011). Furthermore, personality characteristics can play a role, such as psychological reactance and self-efficacy (Bandura, 1977; Brehm & Brehm, 2013). This is fully acknowledged, and will be discussed more extensively in the concluding section. However, since the main goal of our article is to capture, using the new concept of general policy alienation, how past policy events influence later responses of frontline workers, we do not focus explicitly on such aspects.

Table 2.1 Definition of general policy alienation: Five dimensions

Dimension	Policy alienation*	General policy alienation	Example high general policy alienation
Strategic powerlessness	The perceived influence of frontline workers on decisions concerning the content of policy X as captured in rules and regulations.	The influence that frontline workers usually perceive themselves as having on decisions concerning the content of government policies as captured in rules and regulations.	A teacher feeling that the government drafts education policies without involving teachers.
Tactical powerlessness	Frontline workers' perceived influence on decisions concerning the way policy X is implemented within their organization.	The influence that frontline workers usually perceive themselves as having on decisions concerning the way (new) government policies are implemented within their organization.	A teacher stating that the school leader does not involve teachers structurally in designing the implementation of government policies within the school.
Operational powerlessness	The perceived influence of frontline workers during actual implementation of policy X.	The influence that frontline workers usually perceive themselves as having during the actual implementation of government policies.	A teacher answering 'totally agree' to a survey question asking if autonomy during the implementation of government policies is usually lower than it should be.
Societal meaninglessness	The perception of frontline workers concerning the added value of policy X to policy goal Y.	The perception of frontline workers concerning the added value of contemporary policy to socially relevant goals.	A teacher stating in an interview that contemporary education policy is, in their opinion, not contributing to socially relevant goal A.
Client meaninglessness	Frontline workers' perceptions of the added value of policy X for their own clients.	The perception of frontline workers concerning the added value of contemporary policy for their own clients.	A teacher noting that, overall, contemporary education policy has detrimental effects on their own students' wellbeing.

* The definitions presented in this column are drawn from Tummers (2012).

2.3 GENERAL POLICY ALIENATION MEASUREMENT SCALE

In this section, we report on how we developed an empirically validated measurement scale for general policy alienation. We first briefly introduce the case in which we tested our scale, and then show how we developed the items and collected our sample. We then describe our analysis plan and present the results of the exploratory and confirmatory factor analyses, as well as internal construct and convergent validity tests.

2.3.1 Case

The case we selected for our study is the Dutch secondary education sector. Within this sector, there are around 700, both publicly run and privately run, schools. We selected this case because, in recent decades, the sector has experienced many problems as a result of the reshuffling of authority and responsibilities between the ministerial and the school levels (Pijl & Frissen, 2009). Further, the sector has been characterized by numerous policy changes (Bronneman-Helmers, 2008). These problems were also highlighted by the 2008 Dutch Parliamentary Commission ('Commission Dijsselbloem') that investigated problems with education reforms. The Commission's main conclusion was that the government interfered too often in education. They recommended that schools should have greater autonomy, rather than, as in the past, being mere executors of central government policies. That the findings of the Commission are still relevant is highlighted by a recent report by the Dutch Education Council that stated that teachers have not seen any improvement since the Commission Dijsselbloem report (Onderwijsraad, 2014).

2.3.2 Item generation and expert review

The proposed general policy alienation measurement scale is an adaptation of the validated policy alienation measurement scale (Tummers, 2012). As such, we used the same items (measured on five-point Likert scales), but adjusted them to measure general policy alienation. For instance, in the policy alienation measurement scale the following item is used to measure tactical powerlessness: *In my organization, professionals were not listened to about the introduction of the policy.* To measure general policy alienation, this becomes: *In my organization, professionals are not listened to during the introduction of government policies.* An example item for the meaninglessness dimension of the policy alienation scale is: *The policy is contributing to the welfare of my clients (R).* To measure general policy alienation, this becomes: *In general, government policies contribute to the welfare of my clients (R).*

To further increase content validity (DeVellis, 2003), we asked ten experts to evaluate the adjusted items. We selected these experts for their range of different expertise, including public administration scholars, specialists in electronic surveys, policy officers working at the Ministry of Education, and teachers. Appendix I presents an overview of the items in the general policy alienation scale.

2.3.3 Sample and procedure

The general policy alienation measurement scale was tested using large-scale survey data. These data were collected in June 2013. A nationwide sample of 2.863 secondary teachers, selected through the records of the pension fund for all Dutch government and education employees (ABP), was identified. This ensured that the sample would be sufficiently representative of all Dutch secondary school teachers. All the potential respondents were sent an e-mail with an invitation to voluntarily participate in the questionnaire; and a reminder was sent one week later. In total, 1.096 teachers completed the questionnaire: a response rate of 38 percent. The average age of the respondents was 51 years, and 59 percent were male. National statistics on secondary school teachers in 2013 indicate that the average age is 46 and that 48 percent are male (DUO, 2014). As such, men are overrepresented in our sample, and the respondents were on average older than the population from which they were drawn.

2.3.4 Analysis

In order to establish whether the general policy alienation measurement scale performed as expected, a number of analyses were completed using the latent variable program Mplus (version 6). All parameters were estimated using full information likelihood estimation (FIML) such that all respondents with data on at least one of the variables were included in the analyses. As a first step, we conducted factor analyses. Factor analysis is a statistical method used to determine the number of underlying dimensions contained in a set of observed variables and to identify the subset of variables that corresponds to each dimension. Since the policy alienation scale had been validated in previous studies, the dimensionality of policy alienation was already known, and so a confirmatory factor analysis was in principle sufficient (Brown, 2012). However, since we made minor modifications to each item and previous survey studies using the scale were conducted in the healthcare sector, an exploratory factor analysis was nevertheless conducted. Here, we randomly split the total sample of 1.096 into two (subsample 1: *N*=543; subsample 2: *N*=553). We carried out an exploratory factor analysis using the first subsample, and a confirmatory factor analysis using the second. Osborne and Fitzpatrick (2012) refer to this as internal replication and recommend this approach for determining the extent to which solutions are

likely to be robust. Finally, we conducted tests to establish the construct validity of the general policy alienation scale by comparing the measured construct to other constructs based on hypothesized relationships (DeVellis, 2003). Here we looked at convergent validity: the similarity between measures of theoretically related constructs.

2.3.5 Results of factor analyses

Exploratory factor analysis
An exploratory factor analysis was carried out on our first subsample of 543 teachers. We employed oblique rotation since this enabled us to study both the pattern and the structure matrix. This is a common approach when factors are known to be related (Brown, 2012). Further, given the hypothesized five dimensions of general policy alienation, we allowed Mplus to vary the number of factors to be found from 1 to 5. In assessing the number of factors that best fitted the survey data, we referred to the comparative fit index (CFI), the Tucker-Lewis index (TLI), the root mean square error of approximation (RMSEA), and the standardized root mean square residual (SRMR) (see also Schreiber *et al.*, 2006). Generally accepted cutoff criteria for the CFI and TLI indices are ≥0.95 for a good fit and ≥0.90 for a moderate fit. Similarly, RMSEA values ≤0.06 indicate a good fit and ≤0.08 a moderate one (Brown, 2012). SRMR values ≤0.08 reflect a good fit (Hu & Bentler, 1999). Table 2.2 shows that CFI and TLI increased, and the RMSEA and SRMR decreased, as the number of factors distinguished increased (i.e., the fit improved). As expected, the five-factor structure of general policy alienation best fitted the data. Only when five factors were distinguished, did all the indices achieve at least a moderate fit.

Table 2.2 Fit indexes exploratory factor analysis

Fit index	Number of factors				
	1	2	3	4	5
CFI	0.48	0.74	0.79	0.88	0.95
TLI	0.44	0.69	0.73	0.83	0.92
RMSEA	0.15	0.11	0.10	0.08	0.06
SRMR	0.13	0.08	0.06	0.03	0.03

Confirmatory factor analysis

A confirmatory factor analysis was completed using the second subsample of 553 teachers. Again, we assessed the fit of the model based on the CFI, TLI, RMSEA, and SRMR values. The fit of the hypothesized five-factor model was again good (with CFI, TLI, RMSEA, and SRMR values of 0.92, 0.91, 0.06, and 0.05 respectively). This is a good indication that no further modifications to the model are necessary to measure general policy alienation.

Descriptive statistics

Table 2.3 shows the mean scores of our respondents on general policy alienation and its sub-dimensions. Examining Table 2.3, we see that teachers on average experience considerable policy alienation (mean 3.46). They have the impression that, in general, they do not have the power to influence policies (mean 3.34) and neither do they perceive policies as being meaningful for society or for their own clients (means for societal and client meaninglessness 3.49 and 3.67 respectively). Nevertheless, we should also note that the variation in scores between individual teachers is quite large (with mean scores varying between 1 and 5). This indicates that there are also Dutch secondary teachers who do not experience policy alienation at all.

Table 2.3 Means of general policy alienation and its dimensions

	Mean	Standard deviation	Minimum	Maximum
General policy alienation (1-5)	3.46	0.58	1.39	5
Powerlessness (1-3)	3.34	0.60	1.72	5
1. Strategic	3.71	0.66	1.33	5
2. Tactical	3.07	0.88	1	5
3. Operational	3.22	0.75	1	5
Meaninglessness (4-5)	3.58	0.75	1	5
4. Societal	3.49	0.85	1	5
5. Client	3.67	0.78	1	5

2.3.6 Results of construct validity tests

Internal construct validity

Given that the three powerlessness and the two meaninglessness dimensions all measure the same underlying latent construct (general policy alienation), the factors should correlate. Table 2.4 shows the correlations among the powerlessness and meaninglessness dimensions, and indeed they, as expected, are all positively correlated.

Table 2.4 Internal construct validity

		1	2	3	4	5
1	Strategic powerlessness	-				
2	Tactical powerlessness	0.39*	-			
3	Operational powerlessness	0.45*	0.56*	-		
4	Societal meaninglessness	0.48*	0.30*	0.42*	-	
5	Client meaninglessness	0.49*	0.26*	0.47*	0.77*	-

* $p<0.001$

Convergent validity

Our final test examined the relationship between general policy alienation and theoretically related concepts to test the convergent validity of the scale. If our scale truly measures general policy alienation, it should correlate with scales of related concepts. Here, we examined the correlations of general policy alienation with four related concepts: alienation towards a specific policy program; policy consistency; transformational leadership; and frontline workers' willingness to implement new policies. Below, we describe why we expect a relationship to exist between each of these four concepts and general policy alienation, and whether correlational analyses confirmed the expectations.

Alienation towards a specific policy program

We argued earlier that alienation towards a specific policy (program) is determined by a combination of frontline workers' degree of general policy alienation and their perceptions of the unique characteristics of this specific policy (in terms of both content and process). If this is true, general policy alienation should positively correlate with policy alienation towards a specific policy program. To

estimate this relationship, we asked half of our respondents (randomly selected; *N*=551) to assess the societal and client meaninglessness of a specific recent government policy program, namely 'data-driven teaching' (in Dutch: 'opbreng-stgericht werken'). This program aims to stimulate teachers to make educational decisions based on data, a policy apparently dedicated to the achievement of better student results. Indeed, research in Canada and the US shows that data-driven decision-making in teaching contributes to better student results (e.g., Marsh *et al.*, 2010). The policy program is thus to an extent evidence-based. However, another characteristic of the policy is that it intervenes at the classroom level by prescribing how teachers should teach and organize their lessons. This touches on the sensitive 'what-and-how debate' in the Dutch education sector, which postulates that government should focus on *what* should be taught, and schools (school leaders and teachers) on *how* this should be taught. The data-driven teaching policy program is not aligned with this principle. As such, we would expect these specific characteristics to affect the degree of policy alienation teachers feel towards this policy. However, given the purpose of the correlation analyses - to test the correlation of general policy alienation with related concepts - we do not focus further on this misalignment. As expected, the correlation between frontline workers' general policy alienation and their perceived societal and client meaninglessness of data-driven teaching is positive. This is true for all five dimensions of policy alienation, with the correlation between general and specific policy alienation varying between 0.26 (general tactical powerlessness and policy-specific client meaninglessness) and 0.77 (general client meaninglessness and policy-specific client meaninglessness). This suggests that general and specific policy alienation are indeed related, but distinguishable, concepts. This conceptual distinction would be questionable if the correlation was close to unity.

Policy consistency

The second correlation that we investigated is between general policy alienation and policy consistency, a concept closely related to policy accumulation. Frontline workers are often confronted with new policies, and with new rules, regulations, and organizations that they bring. It takes some time to identify with a new policy program (e.g., Elmore & McLaughlin, 1988) and so being regularly

confronted with new policies could be an important cause of general policy alienation. We would expect that the extent to which teachers have the impression that policies are introduced on an ad-hoc basis and are inconsistent - both over time and in relation to other policy measures - to influence whether they feel connected to these policies, as feeling connected is a process that takes effort and time. As Table 2.5 shows, all five dimensions of general policy alienation are, as expected, negatively related to policy consistency. We see that strategic powerlessness and societal and client meaninglessness are especially correlated with policy consistency (correlations (r) of -0.47, -0.48, and -0.50 respectively). This suggests that policies that are more consistent, implying more consistent policy accumulation, result in lower general policy alienation.

Table 2.5 Convergent validity

General policy alienation dimensions	Policy alienation specific program: data driven teaching^		Policy consistency^^	Transformational leadership^^	Willingness to implement new policies^^
	SM	CM			
1 Strategic powerlessness	0.45*	0.41*	-0.47*	-0.23*	-0.33*
2 Tactical powerlessness	0.31*	0.26*	-0.15*	-0.71*	-0.21*
3 Operational powerlessness	0.34*	0.35*	-0.25*	-0.42*	-0.31*
4 Societal meaninglessness	0.70*	0.61*	-0.48*	-0.30*	-0.47*
5 Client meaninglessness	0.63*	0.77*	-0.50*	-0.28*	-0.51*

* $p<0.001$; ^ N=551; ^^ N=1.096

Transformational leadership

The third correlation investigated was between general policy alienation and transformational leadership. Here we made use of the concise measure of transformational leadership by Carless, Wearing, and Mann (2000). We have two main arguments for expecting this correlation. First, the organizational change literature shows that organizational leaders play a crucial role in the successful management of change. Transformational leaders are able to provide an inspirational vision of the future and encourage others to understand the rationale behind new policies (DeCelles *et al.*, 2013). The second argument is that transformational leadership is characterized by empowering and inspiring behavior, thereby supporting others to take personal responsibility when facing

new challenges (Moynihan *et al.*, 2012) - such as new policies. As can be seen in Table 2.5, all five dimensions of general policy alienation are, as expected, negatively related to transformational leadership. We see that especially tactical and operational powerlessness are negatively correlated with transformational leadership (r=-0.71 and r=-0.42 respectively) - which is in line with the findings discussed above.

Willingness to implement new policies

The fourth correlation investigated was between general policy alienation and willingness to implement new government policies, using the five–item change willingness scale of Metselaar (1997). The assumption is that frontline workers who experience greater general policy alienation will be less willing to implement future policies. We offer two main reasons for this. The first is that, in the change management literature, the 'case for change' notion, which is closely (and negatively) related to the meaninglessness dimension of policy alienation, increases willingness to change. Further, it is well established that influence over decisions related to change – i.e. reduced powerlessness - leads to increased commitment and performance, and less resistance to change (Wanberg & Banas, 2000). Furthermore, Tummers (2011) showed that the degree of policy alienation shown by mental healthcare professionals' towards a specific policy (a new reimbursement policy) negatively influenced their willingness to implement that policy. We would expect a similar correlation between general policy alienation and willingness to implement future policies. Indeed, as can be seen in Table 2.5, all five dimensions of general policy alienation are negatively related to willingness to implement new policies. There are especially strong correlations between both societal and client meaninglessness and the willingness to implement (r=-0.47 and r=-0.51 respectively). This suggests that if frontline workers have the impression that government policies in general contribute to important societal goals and achieve desirable outcomes for their own clients, they will be more willing to implement future government policies - possibly because they expect these future policies to be meaningful for society and for their clients as well.

2.4 DISCUSSION

Policies have a history - they are not developed in a vacuum. When studying the effects of specific policies, it is important to take account of the accumulation of policy programs within a specific sector (In 't Veld, 1989). Studying policies in a vacuum, and ignoring the consequences of their history, fails to deliver a complete picture. The starting point of this study was the argument that frontline workers' earlier experiences with government policies will affect their current predisposition towards policies in general. This predisposition will, in turn, affect their attitudes and behaviors towards new policies. Many studies focus on the attitudes and behaviors of frontline workers in relation to policy implementation - accepting that appropriate attitudes and behaviors are crucial for successful implementation - but often fail to consider the possible consequences of their policy predisposition. In this study, the focus is on this policy predisposition, and we conceptualize and operationalize it. In this, we build on the earlier work by Tummers, Bekkers, and Steijn (2009; 2012) on policy alienation, defined as a psychological disconnection from a policy program. As we are interested in investigating overall policy experiences, we introduce and operationalize the term general policy alienation, which will enable future research to analyze the overall experiences of frontline workers with government policy.

In our study, we theoretically related general policy alienation to the consequences of policy accumulation: the continuous aggregation of policies that historically follow upon each other, and the new rules, regulations, and organizations that result. By studying policy alienation in relation to its historical context, we are not only contributing to the work on policy alienation, we also extend the theoretical work on policy accumulation (In 't Veld, 1989) and related concepts such as policy succession (Hogwood & Peters, 1982) and institutional layering (Thelen, 2004). Our respondents' relatively high scores for general policy alienation show that Dutch secondary school teachers do not in general identify with government policies: they have the impression that they lack sufficient power to influence government policies and they sometimes fail to perceive these policies as meaningful, either for society as a whole or for their own students. Relating this to the concepts of change fatigue and change cynicism, it may be that frontline workers experience something akin to policy fatigue or policy cynicism.

This is not the same as private sector employees developing cynical attitudes that characterize organizational change efforts as just the 'flavor of the month' (Herold *et al.*, 2007), but about frontline workers developing cynical attitudes that characterize new policies as just the 'political flavor of the month'. This is a serious problem, especially for governments, as these frontline workers form a crucial link between formulated and implemented policies and between governments and citizens (Bartels, 2013; Tummers *et al.*, 2015).

We would urge future studies to dig deeper into this topic. This is important for at least two reasons. First, our analyses found some evidence that frontline workers' general policy perceptions are indeed related to their perceptions of a specific new policy program. That is, the analyses showed that frontline workers which have a relatively high level of general policy alienation also show greater alienation towards a specific policy program. This suggests that, if one wants to fully understand frontline workers' attitudes towards a specific new policy, both their perceptions of this policy's characteristics as well as their overall policy perceptions should be simultaneously investigated (along with other relevant variables as discussed in the next paragraph). Excluding either set of perceptions is likely to result in an inability to put forward satisfactory explanations of why frontline workers do, or do not, identify with a specific new policy. Thus, the main advice resulting from this study would be to *bring in policy history.*

Second, the developed and validated measurement scale enables future researchers to quantitatively examine the antecedents and effects of the extent of frontline workers' general policy alienation (reflecting their policy predisposition). Although we conceptually link frontline workers' general policy alienation to the consequences of policy accumulation, we are not implying that general policy alienation is the result *only* of accumulated past policy experiences. We acknowledge that other factors play a role, and future research should address this. Regarding new theoretical avenues, we would first urge future research to further analyze the concept of policy accumulation. What policy characteristics influence the degree to which frontline workers perceive policy accumulation as either positive or negative? Policy accumulation may, for instance, be perceived as negative when the rate of policy change is high (Huy, 2001) or the accumulated policies are inconsistent. We have provided some initial evidence of the latter through our correlational analysis between policy consistency and

general policy alienation: greater perceived policy consistency - an indicator of more continuous policy accumulation - seems to be related to lower general policy alienation. Second, we recommend further investigation of general and specific policy alienation, and particular responses alongside other important antecedents on the policy, organizational, and individual levels. In this way, one could determine which factors have the greatest influence in specific contexts. As noted, organizational socialization and culture may be important predictors. More generally, potentially important factors can be found in the literature on the sociology of professions (Teodoro, 2014), organizational behavior (Vigoda-Gadot & Beeri, 2012), and street-level bureaucracy (Hupe & Buffat, 2014).

Regarding the possible effects of general policy alienation, our convergent validity tests showed that general policy alienation is negatively related to willingness to implement future policies. This could have important consequences for (the study of) change management in the public sector. Kickert (2010) noted that the change management literature is primarily focused on the private sector and that little attention is paid to the way in which public employees react to change. Our measurement instrument is useful for researching public employees' experiences with past, current, and future policy changes and the consequences of these changes. It will enable future research to fill the gap in the literature on change management by specifically applying a public administration perspective (Kuipers *et al.*, 2014). Ultimately, this could contribute to a better understanding of why, despite all the efforts made, many change efforts in the public sector fail.

Despite the progress made, this study has, as all studies, some limitations. The first limitation is that the data used to establish convergent validity are cross-sectional. In recent years, authors, reviewers, and editors of leading public administration journals have become increasingly concerned about the validity of such research. One of the main concerns is that causal inferences are not possible. In our research, we investigated correlations without aiming to make statements about causality. However, especially in light of the relationship found between policy accumulation, (general) policy alienation, and willingness to implement a specific policy, future studies should adopt longitudinal (or experimental) designs to investigate causality. In this way, it could be established whether frontline workers' general policy alienation (at t=0) influences their feeling of policy alienation towards a newly introduced policy program at some later

time (t=1), which in turn could influence their general policy alienation (after the implementation of the policy program, at t=2). Further, it would also enable an assessment of whether perceptions of policy accumulation processes (at t=-1) affect the degree of general policy alienation (t=0). A second limitation is that the organizational context was not included in the analyses. Government policies are implemented in this context, and it is therefore likely to have a significant influence on overall policy perceptions. Although we focused on the government context, the convergent validity tests on general policy alienation and transformational leadership show the importance of organizational leadership and thus organizational context. In future studies, greater attention could therefore be paid to processes at the organizational level.

2.5 CONCLUSIONS

In concluding, we would emphasize that the present study explicitly considered processes of policy accumulation and promoted the notion that, when investigating the formulation and implementation of a specific (new) policy, this policy's past should not be ignored. Future research should take advantage of this, and use the framework to ensure that attention is given to the previous policy experiences of frontline workers. This acknowledges that they bring with them a policy history, and cannot be regarded as 'neutral' implementers. In our opinion, this recognition contributes to a more realistic and context-sensitive research perspective on policy implementation and its effects on frontline workers.

chapter 3

Developing a short measure of general policy alienation

ABSTRACT

Public administration research is becoming increasingly quantitative. As seen in psychological and managerial research, the result is a growing demand for valid and reliable measures. Given the tradition of contextually embedded research in public administration – where research should cover multiple factors to find useful answers to real-life problems – survey research should ideally incorporate many measures. This is driving a need for short measures that do not compromise on validity and reliability. In this study, a short measure of general policy alienation is developed and tested, observing stringent criteria. The analyses on three independent datasets (N=1.183, N=354, and N=933) show that the original multidimensional 26-item measure can be abbreviated to a short five-item measure with limited compromises on validity and reliability. Practical applications and methodological implications regarding both the developed measure and the 10-step procedure used are discussed.

3.1 INTRODUCTION

The implementation of government policies can put frontline workers - such as teachers, police officers and nurses - severely under pressure (Lipsky, 1980; Brehm & Gates, 1991; Maynard-Moody & Musheno, 2003; Hill & Hupe, 2009; Tummers *et al.*, 2009; Kiefer *et al.*, 2015). This, of course, is not a problem in itself; but when frontline workers are confronted with new policies – and the concomitant changes to, among others, rules, administrative procedures and finances – on a regular basis, then this may well cause problems. This applies particularly if the workers do not recognize the added value of these policies (e.g., May & Winter, 2009). As studies on change management have suggested, repeated policy changes may lead to fatigue or apathy of frontline workers (Connel & Waring, 2002). This implies that they become indifferent to whatever new policy (change) is introduced, viewing new policies as just the new 'political flavor of the month' (cf. Herold *et al.*, 2007). This results in suboptimal circumstances for society at large, as public funds are invested in the formulation and implementation of government policies which (apparently) are not supported by frontline workers. Ultimately, this impedes the improvement of public service provision, as this is only likely to be achieved if actors operating at different levels of the system collaborate willingly (Bryson *et al.*, 2015).

To explicitly take account of frontline workers' previous experiences with government policies, Van Engen *et al.* (2016) introduced the concept of general policy alienation. This is defined as an overall cognitive disconnectedness from government policies. In other words, general policy alienation occurs when frontline workers cannot identify with government policies overall. It is not about a specific policy, but about a disinterest in government policies in general; a negative policy predisposition. Building on the policy alienation framework of Tummers (2012), they developed a 26-item measure of general policy alienation. However, given the limited number of items that can be included in a typical survey questionnaire, the 26-item length of the current measure might have limited use for research.

The main goal of this study is therefore to develop a valid and reliable short measure of general policy alienation. First, this creates survey time that researchers can use to measure additional variables (Liden *et al.*, 2015). This makes it

easier to include the measure into surveys. Second, many items may introduce fatigue or boredom among respondents, which may negatively influence the quality of the responses obtained (Crede *et al.*, 2012). Third, a short measure is more likely to be applied in other fields of public administration where general policy alienation is not the core subject matter, but could form a relevant contextual or explanatory factor. For instance, when studying the effectiveness of political and public leaders or the organizational commitment or turnover intentions of public employees. Clearly, however, short measures may compromise validity if not developed using rigorous methods. This problem will be contained by applying the stringent 10-step approach we developed based on guidelines by, among others, DeVellis (2012) and Smith *et al.* (2000).

This article is structured as follows. The next section offers a brief theoretical background on policy implementation, policy alienation, and the development of short measures. Section 3 outlines the method adopted and describes the results of the analyses conducted on three independent large-scale data sets, collected in the Dutch education sector in 2013, 2015, and 2016. The final section presents the conclusions, focusing particularly on recommended future lines of research and the methodological implications for public administration scholars aiming to develop short measures.

3.2 THEORETICAL BACKGROUND

3.2.1 Policy implementation and street-level bureaucracy

For a long time, policy implementation was considered a rather mechanistic activity. Van Meter and Van Horn (1975, p. 450) even described policy implementation as "a series of mundane decisions and interactions unworthy of the attention of scholars". Not surprisingly, early theories of policy formulation and implementation were top-down oriented: administrators were simply expected to carry out the policies as formulated by politicians. However, insights changed as it became clear that implementation may lead to a reformulation of policies, to other outcomes than expected, or even to outright failure (Lipsky, 1980; Sabatier & Mazmanian, 1980; Hill & Hupe, 2009).

Over time, several explanations have been put forward to explain this 'implementation gap'. These include the lack of control and monitoring (Pressman & Wildavsky, 1984), the lack of adequate training opportunities (Kroll & Moynihan, 2015), and too ambiguous policy objectives (Lipsky, 1980). These critiques showed that the earlier, rational models of policymaking and public management were "not effective in practice, nor convincing in theory" (Parsons, 1995, p. 468). As a result, new models were developed that emphasized the complexities of policy implementation and the prominent role fulfilled by frontline workers.

Within the subfield of street-level bureaucracy, researchers such as Maynard-Moody and Musheno (2003), Hill and Hupe (2009) and Tummers, Bekkers, and Steijn (2009) have repeatedly emphasized how the support of frontline workers is crucial to a successful implementation. Yet despite frontline workers' crucial role, there has been little effort to develop and test a framework for this (O'Toole, 2000). In an attempt to fill this gap and to allow for systematic and coherent analyses of frontline workers' experiences with policies, Tummers, Bekkers, and Steijn (2009) developed the policy alienation framework.

3.2.2 Policy alienation

Policy alienation is defined as "a cognitive state of psychological disconnection from the policy program being implemented by a public professional who regularly interacts directly with clients" (Tummers *et al.*, 2009, p. 688). The concept consists of two main dimensions: powerlessness and meaninglessness. Policy powerlessness concerns the degree of influence frontline workers have on shaping the policy introduced by the government. This power may be exercised on the strategic, tactical or operational levels, where it influences, respectively, the national (government) level, the organizational level, and the actual policy implementation at the micro-level. Meaninglessness refers to frontline workers' perceptions of the contribution a policy makes (or fails to make) to some greater purpose. Societal meaninglessness refers to frontline workers' perception of the value that a policy has for socially relevant goals (Tummers *et al.*, 2009). Client meaninglessness reflects frontline workers' perception of the added value for their own clients (e.g., patients, students). Previous research has shown that policy alienation negatively affects implementation willingness. If frontline workers experience more policy alienation towards a policy, they are less will-

ing to implement this policy and less motivated to support it (Tummers, 2012). Research also shows that policy alienation is negatively related to important job aspects for frontline workers, such as job satisfaction (Tummers, 2012). Furthermore, a relationship has been established between policy alienation and coping behavior of frontline workers (Loyens, 2015) and output performance (Thomann, 2015).

Recently, a conceptual distinction between specific policy alienation (cognitive disconnectedness from a specific policy program) and general policy alienation (overall cognitive disconnectedness from government policies) was proposed (Van Engen *et al.*, 2016). Whereas the original policy alienation framework aims to analyze frontline workers' experiences with a *specific policy (program)* in their field, the general policy alienation framework can be used to analyze frontline workers' *overall experiences with government policies.* In this study we focus on the latter. General policy alienation takes account of the fact that government policies are not developed in a vacuum (Hogwood & Peters, 1982) but rather have a history; they are built upon other policies. Insights from change management studies – where terms such as 'change fatigue' and 'change cynicism' are used – show that employees' previous experiences with change significantly affect their openness and willingness to change at a later stage (Bordia *et al.*, 2011). The same mechanism may also apply to frontline workers and their perception of new policies, and this will also influence the effectiveness and legitimacy of these policies.

3.2.3 Short measures

Short – sometimes referred to as unidimensional or global (e.g., Smith *et al.*, 2000; Crede *et al.*, 2012) – and multidimensional measures of the same concept often co-exist (Wright *et al.*, 2013). Multidimensional measures are usually seen as yielding a more nuanced understanding of the different origins or forms of a concept. However, as pointed out in the research on the measurement of public service motivation, the length of a multidimensional measure may limit its use (Coursey & Pandey, 2007; Wright *et al.*, 2013). Besides that, concerns have been raised regarding respondent fatigue or response bias (Crede *et al.*, 2012). Short measures, as compared to multidimensional measures, often yield a more accurate measure of a concept's overall strength (Ironson *et al.*, 1989; Crossley *et al.*, 2007). That is, researchers often use short measures to quickly assess the overall

or general level of a construct without having to identify or include a full range of dimensions of that construct (Crossley *et al.*, 2007): short measures assume that some level of aggregated mental processing occurs as respondents combine their thoughts and feelings regarding various aspects of a multidimensional construct in order to provide a single integrated response (Ironsen *et al.*, 1989).

There are two objections to the development of short measures. The first is that it is virtually always a mistake to try to develop one: any proper assessment of a real life situation is always worth the time. The inevitable loss of validity in return for the time savings is sometimes unjustifiable (Smith *et al.*, 2000). Simulation research in clinical contexts, for instance, showed that shorter tests produce a higher risk of drawing incorrect conclusions about change in individual clients (Kruyen *et al.*, 2013). Therefore, one should be cautious in deciding what loss of validity might be acceptable in light of the research topic. This is especially relevant if measures are used to draw far-reaching conclusions at the individual level, which is, for instance, more common in clinical than in public administration contexts. The second objection, on the other hand, is not directed at short measures per se but rather at the incorrect ways in which they are often constructed. Short measures have frequently been developed without a thorough consideration of validity. From this it follows that useful abbreviated instruments can only be developed if researchers improve the methodology of short measure development (Smith *et al.*, 2000).

In this article, we develop a short measure of general policy alienation. Summarizing the above, the advantage of such a measure is that it allows researchers to use just a few items to assess frontline workers' overall level of general policy alienation. They do not have to include the full range of policy alienation dimensions. In doing this, we follow a 10-step procedure, based on guidelines by DeVellis (2003) and Smith *et al.* (2000), to pay close attention to validity and reliability issues.

3.2.4 Steps to develop a short measure

DeVellis (2012) formulated eight general steps that need to be taken in order to develop a valid measurement instrument. The steps he prescribes are: 1) determine clearly what you want to measure, 2) generate an item pool, 3) determine the format for measurement, 4) have the item pool reviewed, 5) consider the inclusion of validation items, 6) administer the items to a pilot sample, 7) evaluate

the items, and 8) produce the final measure. These steps are generally considered useful guidelines in developing 'normal' measures. To fit our objective of developing a short measure, a number of adaptations were made, mostly based on suggestions by Smith *et al.* (2000).

First, we distinguished between a preparatory and an analysis stage, and by using a different dataset for each we avoided using a single dataset to develop both the original and short form. Second, and again as recommended by Smith *et al.*, we included the requirement to only develop a short measure of a sufficiently validated original measure (step 1). Third, we modified the second step of DeVellis, 'Generate an item pool', to 'Select item pool from original measure', as a selection of original items often forms the short measure (e.g., the work on servant leadership by Liden *et al.*, 2015; and the work on public service motivation, as noted by Wright *et al.*, 2013). Using pre-existing items rather than crafting novel items is not only efficient (Crede *et al.*, 2012), it also allows the short measure to be drawn from – and compared with - (existing) datasets including the original measure. When selecting the items, it should be kept in mind that the selection of items should allow respondents to combine their thoughts and feelings regarding various aspects of a multidimensional construct in order to provide a single integrated response (Ironsen *et al.*, 1989). However, (fit) analyses performed on the selected item pool could reveal a need to develop additional items. Therefore, we included an explicit decision on whether or not to develop additional items (step 4). Furthermore, we included the decision on whether or not to change the format for measurement (step 5), as it could be that the fewer number of items demands a larger variety of answer categories (Dawes, 2010). After these steps have been completed, the proposed short measure should be discussed with experts (step 6). Finally, once the experts have approved the measure, the proposed measure should be included in a new survey questionnaire. With this seventh step, the preparatory stage is concluded.

The second stage of short measure development is the analysis stage, where the goal is to evaluate the proposed measure using a new dataset. We included the assessment of internal consistency reliability (step 8) to determine whether the proposed items in fact address the same underlying construct. We also included construct validity (step 9), to determine whether the measure 'behaves as it should behave' in relation to other variables. Once these steps have been

completed, the final measure is ready (step 10). If the results of the analysis stage are not satisfactory, this stage should be repeated using a new (third) dataset (see Liden *et al.*, 2015). In Appendix II, we provide an overview of the steps taken in this study in order to develop a valid and reliable short measure.

3.3 A SHORT MEASURE OF GENERAL POLICY ALIENATION

Here we apply the ten steps described in the previous paragraph to develop a short measure of general policy alienation.

Step 1. Only develop a short measure of a sufficiently validated original measure

The short measure of general policy alienation is a short form of the validated general policy alienation measure (Van Engen *et al.*, 2016), which is an adaptation of the validated policy alienation measure (Tummers, 2012; used in, among others, Tummers *et al.*, 2012; Thomann, 2015; Kerpershoek *et al.*, 2016; Van der Voet *et al.*, 2017). The preliminary requirement to only develop a short measure of a validated original measure is thus met, although it should be acknowledged that despite the fact that these studies offer evidence of a valid original measure, further cross-national and cross-sectoral validation is recommended. We discuss this limitation more extensively in the discussion section.

Step 2. Determine clearly what you want to measure

A short measure of general policy alienation needs to produce an accurate measure of the concept's overall strength, yet using a smaller number of items. General policy alienation has been defined as a cognitive state of psychological disconnectedness from government policies (Van Engen *et al.*, 2016). It is a mental status quo of frontline workers that reflects their policy predisposition informed by, among other things, accumulated past policy experiences. We invoke general policy alienation as an explanation for observed behavior, such as practical interventions by frontline workers, but also industrial action or reduced prosocial behavior. General policy alienation is thus not equivalent to this behavior but underlies it.

The core of the concept concerns the fact that frontline workers often feel neglected in the set-up and design of relevant policies at the national level, and that these policies do not allow for enough discretionary power. Additionally, they often do not perceive policies as contributing meaningfully to socially relevant goals or as having added value for their own clients, both in the short and the long term. First, this conceptualization suggests that the tactical powerlessness dimension of policy alienation, which pertains to whether frontline workers feel they have the power to influence the actual implementation of government policies within their organization (Tummers *et al.*, 2009), perhaps should not be included in the short measure. From previous research, it is known that characteristics of the organization and the organization leader play an extremely important role in policy implementation success or failure (e.g., Brodkin, 2012). Therefore, it might be that these characteristics should be measured separately and not included in the short measure. The latter then concentrates on the direct interplay between government policies and frontline workers' individual perceptions, and not on the mediating role that organizations may play in this process. Second, this conceptualization implies that the short measure of general policy alienation could either be a unidimensional measure (since short scales usually do not allow researchers to identify the different dimensions of a concept) or a second order, two-factor structure model (i.e., a powerlessness and meaninglessness dimension). Both options will be empirically investigated.

Step 3. Select item pool from original measure

In order to select relevant items for the short measure from the original 26-item policy alienation measure, analyses were conducted on the dataset used to develop and validate the original measure (see Van Engen *et al.*, 2016). The analyses in this study were conducted using *R*-packages 'lavaan' (Rosseel, 2012), 'psych' (Revelle, 2015), and 'semTools' (semTools Contributors, 2016).

Sample

For the first dataset, survey data was collected in June 2013. A nationwide sample of 3.127 school leaders and teachers, selected through the records of the pension fund for all Dutch government and education employees (ABP), was identified. This ensured that the sample was sufficiently representative. All the potential

respondents were sent an e-mail with an invitation to participate in the questionnaire, and a reminder was sent one week later. In total, 1.183 respondents completed the questionnaire: a response rate of 38 percent. The average age of the school leaders in our sample was 55 years, and 76 percent were male. The average age of the teachers in our sample was 51 years, and 59 percent were male. Dutch national statistics of 2013 indicate that the average age of school leaders is 53 and of teachers 45. These statistics also indicate that 71 percent of school leaders were male and 49 percent of teachers were male (DUO, 2016). Men are therefore overrepresented in our sample, and the respondents were on average older than the population from which they were drawn.

Item selection

As a general rule, a short measure requires a total of at least three items for the purpose of accurately estimating internal consistency reliability (Liden *et al.*, 2015). As is common practice in short measure development, we started the item selection for the short measure by identifying the items with the highest item-total correlations. That is, from the original 26-item measure, we selected the items with the highest factor loading for respectively the strategic and operational powerlessness and the societal and client meaninglessness dimensions (all these factor loadings are >0.76). This is an appropriate procedure, as items with the most error variance will be eliminated, resulting in a 'purer' measure of the target construct (Smith *et al.*, 2000). In total, we selected four items that could potentially form the short measure. Furthermore, we decided to include one additional item for societal meaninglessness (factor loading is 0.90). The main argument for this is that previous research has shown this to be an important dimension explaining policy evaluations and implementation willingness (Tummers, 2012). Table 3.1 provides an overview of the proposed short measure of general policy alienation.

Finally, to be able to analyze whether our theory-informed decision to not include an item for tactical powerlessness in the short measure is empirically supported, we also selected the item with the highest factor loading (0.88) for the tactical powerlessness dimension.

Table 3.1 The proposed short measure of general policy alienation

Item	Template	Present study	Dimension
1	Professionals cannot influence the development of policies at the national level (Minister and Ministry of X, national government)	School leaders and teachers cannot influence the development of education policies at the national level (Minister and Ministry of Education, national government)	Strategic powerlessness
2	Generally, I have freedom to decide how to use government policies (R)	Generally, I have freedom to decide how to use government education policies	Operational powerlessness
3	Overall, I think that government policy leads to socially relevant goal A (R)	Overall, I think that government education policy leads to higher educational quality	Societal meaninglessness (1)
4	In general, I think that government policy in the long term will lead to socially relevant goal A (R)	In general, I think that government education policy in the long term will lead to higher educational quality	Societal meaninglessness (2)
5	In general, government policy enables me to better solve the problems of my clients (R)	In general, government education policy enables me to better solve the problems of my students	Client meaninglessness
*Optional**			
6	In my organization, professionals - through working groups or meetings - take part in decisions on executing government policies (R)	In my school, teachers – through working groups or meetings – take part in decisions on executing government education policies	Tactical powerlessness

* This item is included to be able to evaluate whether a five- or a six-item (thus including one additional item for tactical powerlessness) short measure fits the data best.

Step 4. Decide whether it is necessary to develop additional items

Based on analyses indicating acceptable factor loadings and thus internal consistency reliability (see Table 3.3; results for study 1), it appeared unnecessary to develop additional items. Naturally, if the analyses would not have indicated this, this step would have been included in the procedure.

Step 5. Decide whether it is necessary to change the format for measurement

The original general policy alienation measure uses 5-point Likert scales. The reduced number of items for the short measure could demand an increase in answer categories to allow for more variation in responses (Dawes, 2010). However, an analysis of the variance in scores among the 1.183 respondents indicated that a 5-point Likert scale is well-suited to the short measure. Both minimum

and maximum scores of 1 and 5 are represented in the dataset and the standard deviation for all items is approximately 1 (see Table 3.2). For all indicators and latent variables, skewness and kurtosis statistics were >-1 and <1, which is generally considered to be an indicator of a normal distribution (Sheskin, 2011). It thus appeared unnecessary to change the format for measurement.

Step 6. Evaluate face validity: Review item pool with experts

In this stage, the proposed five items were discussed with three public administration scholars, one psychology researcher, one expert in survey research, two policy officers and two teachers (in total nine experts). First, the experts were of the opinion that the five items sufficiently covered the general policy alienation concept. Second, all experts recognized the added value of the reduced number of items. For instance, it was emphasized that answering these five questions would be much less tiresome than answering the original 26 questions (by the teachers) and that researchers would be more likely to include the short measure in their questionnaire, especially if policy alienation is not the main topic of study (by the public administration scholars). Given the positive results of the first phase, the five items were now ready to be included in a new survey.

Step 7. Include proposed short measure in a new survey questionnaire

For the second dataset, survey data were collected in 2015 as part of a policy evaluation study, conducted at the request of the Dutch Ministry of Education. All school leaders of the 97 schools that participated in a policy pilot and school leaders of 700 schools that did not participate received an e-mail inviting them to voluntarily participate in the study. The response rate among the school leaders of schools participating in the pilot was 41%, and of the school leaders of schools not participating in the pilot was 9%. Besides the request to fill in the questionnaire, school leaders were also asked to distribute teacher questionnaires among teachers involved in the policy pilot. Unfortunately, this sampling procedure does not allow us to calculate exact response rates. A total of 57 teachers from pilot schools and 192 teachers from non-pilot schools filled out the questionnaire. The complete sample thus consists of 354 school leaders and teachers.

Step 8. Show internal consistency reliability

Internal consistency reliability is concerned with the homogeneity of the items within a measure (DeVellis, 2012). Table 3.2 shows an overview of the descriptive statistics. First, the results indicate that the two samples used are quite comparable. Second, they indicate that the most prominent difference between samples 1 and 2 is that for all indicators the mean score of sample 1 is higher than the mean score of sample 2. Third, they indicate that the mean scores for the tactical powerlessness item are lower than the mean scores for the other five items.

Table 3.2 Descriptive statistics

Item	Mean		Standard deviation		Minimum		Maximum	
	Study 1	Study 2	Study 1	Study 2	Study 1	Study 2	Study 1	Study 2
1	3.86	3.46	0.97	1.03	1	1	5	5
2	3.19	2.82	0.98	0.89	1	1	5	5
3	3.31	3.17	1.04	1.03	1	1	5	5
4	3.41	3.14	1.01	1.02	1	1	5	5
5	3.86	3.44	0.81	0.94	1	1	5	5
6	2.80	2.22	1.04	0.89	1	1	5	5

To test the internal consistency reliability of the proposed measure, a number of fit indices were analyzed (as recommended by Kline, 2015): the comparative fit index (CFI), the Tucker-Lewis index (TLI), the root mean square error of approximation (RMSEA), and the standardized root mean square residual (SRMR). Generally accepted cut-off criteria for the CFI and TLI indices are ≥0.95 for a good fit and ≥0.90 for a moderate fit. Similarly, RMSEA values ≤0.06 indicate a good fit and ≤0.08 a moderate fit and SRMR values ≤0.06 reflect a good fit (Hu & Bentler, 1999; Brown, 2012). Given the five-point Likert, semi-nominal nature of our data, we used the WLSMV-estimator, which does not assume normally distributed variables and thus provides the best option for modeling the data (Brown, 2012). Here we compared the proposed five-item measure with an alternative six-item measure (i.e., including a tactical powerlessness item), modeled as both a unidimensional and second-order construct. An overview of the results is reported in Table 3.3.

Table 3.3 Fit statistics of the tested models

Study	Type of model	CFI	TLI	RMSEA	SRMR
Proposed five-item short measure					
1	unidimensional	0.99	0.98	0.049	0.030
	second-order	0.99	0.99	0.037	0.019
2	unidimensional	0.99	0.99	0.037	0.025
	second-order	0.99	0.99	0.041	0.025
*Alternative six-item short measure**					
1	unidimensional	0.96	0.94	0.079	0.051
	second-order	0.99	0.99	0.037	0.025
2	unidimensional	0.99	0.98	0.063	0.051
	second-order	0.99	0.98	0.075	0.050

* That is, the five-item measure and one item for tactical powerlessness.

First, Table 3.3 illustrates that the proposed five-item measure, overall, fits the data collected in both study 1 and 2 better than the alternative six-item short measure. For all the tested five-item models, the fit statistics pass the recommended thresholds. This is not the case for the six-item short measure. This supports our theoretical argument to not include an item for the tactical powerlessness dimension in the proposed five-item measure. Furthermore, Table 3.3 illustrates that the difference in fit between the five-item unidimensional and second-order model is very limited. Therefore, to decide whether we propose a unidimensional or second-order model, we move our attention to the standardized factor loadings. For the unidimensional model, they vary between 0.24 and 0.95 (study 1) and 0.22 and 0.95 (study 2) (i.e., factor loadings <0.30). For the second-order model, they vary between 0.40 and 0.94 (study 1) and 0.40 and 0.97 (study 2). Standardized factor loadings thus improve with a second-order model. Summarizing, this provides the empirical evidence that the five-item short measure, modeled as a second-order construct fits the data best.

Furthermore, we assessed the reliability of the proposed measure by examining Cronbach's alpha. Although this statistic has been heavily critiqued (e.g., Sijtsma, 2009), reporting it is still common practice in public administration research. Therefore, while we do report Cronbach's alpha, we do not strongly rely on it for the fit evaluation. We were guided by Nunnally and Bernstein's (1994) suggestion that Cronbach's alpha should be at least 0.70 for acceptable reliability.

The Cronbach's alpha coefficient of the short measure of general policy alienation meets this threshold, with 0.70 (study 1) and 0.78 (study 2).

Finally, to be able to further generalize our findings, we cross-validated the proposed five-item measurement structure using a third dataset. This dataset was collected following the same collection procedure as described in the section 'step 3', heading 'sample'. This resulted in a total sample of 933 (a response rate of 50%), of which 84 school leaders and 849 teachers. All four fit statistics, using the WLSMV-estimator, pass the recommended thresholds with CFI=0.99; TLI=0.97; RMSEA=0.065; SRMR=0.047. Standardized factor loadings vary between 0.40 and 0.95 and Cronbach's alpha is 0.77. As a final check, we conducted multi-group CFA and tested whether the proposed short measure has measurement invariance, also known as measurement equivalence, across groups (Byrne, 2008). A scale is said to have measurement invariance across groups if subjects with identical levels of the latent construct have the same expected raw-score on the measure (Drasgow & Kanfer, 1985). The analyses using data of study 1, where we compared groups based on gender (male versus female) and type of school (publicly versus privately owned), indicate that there is measurement invariance between these groups: the chi-square differences are not significant (for gender p-values are respectively 0.74, 0.16 and 0.31 and p-values for type of school respectively 0.72, 0.23 and 0.42). These findings further support a five-item short measure, modeled as a second-order construct.

Step 9. Show construct validity

Construct validity pertains to whether the underlying (latent) concept, here general policy alienation, is the underlying cause of item covariation. To the extent that a measure is reliable, variation in scores can be attributed to the true score of some phenomenon that exerts a causal influence over all the items (DeVellis, 2012).

Convergent validity

Convergent validity tests assess whether constructs that are expected to be related are in fact related. Here we related the short measure to four variables: perception of a specific policy program, policy consistency, implementation willingness, and trust in government. An overview of the measures used is provided

in Appendix IV. The fit of the comprehensive measurement models meet the recommended thresholds, with fit statistics (using the WLSMV-estimator) being CFI=0.99; TLI=0.99; RMSEA=0.037; SRMR=0.040 for study 1, CFI=0.99; TLI=0.99; RMSEA=0.034; SRMR=0.041 for study 2, and CFI=0.99; TLI=0.99; RMSEA=0.010; SRMR=0.032 for study 3.

Perception of a specific policy program. We argued that frontline workers' general policy perceptions affect their perception of specific new policies. To assess this relationship, we asked respondents in both study 1 and 2 to assess the added value (meaningfulness) of a specific, recent government policy program. For the respondents in study 1 this was 'data-driven teaching', a program meant to encourage teachers to make educational decisions based on data. For the respondents in sample 2 this was the 'development of a new diagnostic test' for students completing lower secondary education (age 13 to 15). We expect a negative relationship between respondents' degree of general policy alienation and their perception of the added value of specific policy programs. This assumption was fully confirmed by the data, as shown by Table 3.4.

Policy consistency. Frontline workers are frequently confronted with new policies and the associated new rules, regulations and organizations. As it always takes some time to identify with a new policy program (e.g., Elmore & McLaughlin, 1988), to be confronted with new policies regularly could be an important cause of general policy alienation. We would expect the extent to which frontline workers perceive policies to be introduced on an ad-hoc basis and to be inconsistent – both over time and in relation to other policy measures – to influence to what extent they identify with these policies. As Table 3.4 shows, the short measure of general policy alienation is indeed negatively related to policy consistency.

Implementation willingness. We assume that frontline workers who experience more general policy alienation will be less willing to implement future policies. We offer two main reasons for this. Firstly, if frontline workers perceive the added value of policies – i.e., reduced meaninglessness – their implementation willingness is higher (Tummers *et al.*, 2009). Secondly, it is well established that influence over decisions related to change – i.e., reduced powerlessness - leads to increased commitment and performance, and less resistance to change (Wanberg & Banas, 2000). Hence, we expect a negative relationship between the

proposed short measure and implementation willingness. As Table 3.4 shows, this is supported by the data.

Trust in government. Earlier studies showed that characteristics of New Public Management (NPM) - the economic and management paradigm from the private sector applied in the public sector (for recent work on NPM in the education sector, see Aoki, 2015) - affect policy alienation (Tummers *et al.,* 2009). Of interest is that the NPM model is, among others, characterized by distrust between principals and agents. Or, as Bouckaert (2012, p. 99) stated, as a result of NPM, "the adage 'trust is good, control is better' was replaced by 'distrust is better, audit is best'". We thus expect a negative relationship between the short measure of general policy alienation and trust in government. This assumption was fully confirmed by the data, as shown by Table 3.4, although the relationship is stronger in study 3 (data collected in 2016) than in study 1 (data collected in 2013).

Finally, we also investigated whether or not structural equation modeling with the short measure produces (approximately) equal results as compared to the original measure. The analyses using data of study 1 (N=1.183) show that this is indeed the case for both policy consistency (respectively β=-0.30 for the short and β=-0.29 for the original measure) and implementation willingness (respectively β=-0.17 for the short and β=-0.22 for the original measure). This provides initial evidence for the success of the short measure in capturing the essence of the original measure.

Table 3.4 Construct validity tests

	Short measure of general policy alienation		
	Study 1	**Study 2**	**Study 3**
School size	-0.01	-0.02	
Type of school *ref=public*	-0.05	-0.03	
Perception specific policy program	-0.51**	-0.44**	
Policy consistency	-0.30**		-0.35**
Implementation willingness	-0.17**		-0.21**
Trust in government	-0.11*		-0.50**

* p<0.05; ** p<0.01. Standardized coefficients from the structural equation modeling are reported.

Discriminant validity

Discriminant validity tests serve to determine that variables that should be un-related indeed are unrelated. Two variables that are not theoretically related to policy alienation are the number of students at a school and whether a school is publicly or privately owned. Table 3.4 shows that this assumption is confirmed by the data.

Step 10. Determine final measure

Given the successful completion of the first nine steps, we now propose a valid and reliable short measure of general policy alienation. This measure should be treated as a first attempt, as more (replication) research is needed in other public domains and countries to further develop it (see the discussion section). An overview of the five items that form the measure is presented in Table 3.2 and Appendix I and IV, where an overview of all measures used in this study is provided.

3.4 DISCUSSION

This article reports the development of a short but valid and reliable measure of general policy alienation. To do so we adopted a systematic 10-step procedure, that may also be helpful for researchers to develop short versions of other measures. This resulted in a five-item measure that can be used to measure frontline workers' overall cognitive disconnectedness (or: connectedness) regarding government polices (Van Engen *et al.*, 2016). The measure allows future researchers to easily assess frontline workers' earlier experiences with government policies and to investigate the (behavioral) effects of this predisposition. The main implications of this, regarding both the measure itself and the systematic procedure used to develop it, is discussed below.

The first main implication is that there are now two validated measures of general policy alienation. Although the current investigation has produced substantial evidence for the success of the short measure in capturing the essence of the original measure, the former cannot serve as a full replacement for the latter: the measures clearly serve different purposes. This implies that future researchers who want to use the policy alienation concept in their research should decide

for themselves which measure is the best choice. A first general rule of thumb is that the research question should always guide the decision. If the aim of a study is to investigate (in-depth) what the antecedents and consequences of policy alienation are, we recommend using the original measure. The same applies if one of the policy alienation dimensions is hypothesized (e.g., operational power-lessness or client meaninglessness) to be either the antecedent or a consequence of observed behavior by frontline workers.

If, however, the aim of a study is to incorporate the effect of frontline workers' overall policy perceptions, then the short measure is recommended. Research on public service motivation reveals that much of the current understanding of this concept is based on studies using a general or global measure of PSM (e.g., Stazyk & Davis, 2015; for an overview see Wright *et al.*, 2013). The short measure of general policy alienation may serve this goal, too, as it is more easily integrated in surveys. Frontline workers' policy predispositions (i.e., their degree of general policy alienation), as crystallized attitudes, might heavily condition the influence of government behavior on their policy evaluations (cf. Tesler, 2015). By captur-ing this, our short measure acknowledges that frontline workers bring with them a history of government policy (changes) and, hence, general ideas about the, for instance, effectiveness and legitimacy. Our measure thus enables the application of a typical public administration perspective in change management and policy implementation research (Kuipers *et al.*, 2014). This application is especially relevant in light of the recent increase in public administration studies borrowing and extending theories from the field of psychology, or simply: the rise of the behavioral public administration (Grimmelikhuijsen *et al.*, 2017). This approach, and the opportunities it creates for both public administration scholars and practitioners, can only be successful if the field further strengthens its quality of measurement (Perry, 2016).

This brings us to the second main implication. In recent years we have witnessed a clear increase in the number of quantitative public administration studies (Groeneveld *et al.*, 2015). In line with this, we have witnessed an increase in the number of measures being developed by public administration scholars (e.g., public leadership roles by Tummers & Knies, 2016 or red tape by Van Loon *et al.*, 2016). In light of the limited number of items that can usually be included in a survey questionnaire and the contextually rich studies that public administra-

tion scholars usually (aim to) conduct, it might be expected that short measures will be developed for these. It has been noted that useful, valid and reliable short measures can only be developed by following strict procedures (Smith *et al.*, 2000). We proposed and used a systematic 10-step procedure for developing our short measure. We hope other researchers will find this procedure useful. Although we are not the first in the public administration field to develop short measures – with the work on public service motivation probably being the most exemplary (e.g., Vandenabeele, 2008; Kim *et al.*, 2013) – we believe that our 10-step procedure offers a good starting point for short measure development.

Finally, this study has some limitations that should be spelled out, which will also suggest valuable lines for future research. The first limitation is that we tested our short measure in only one public sector in one country. Notwithstanding that our samples consist of both school leaders and teachers and were collected at three different points in time, the measure should be used and tested in cross-national survey research in, ideally, multiple sectors (as was done within public service motivation research by Kim *et al.*, 2013). If the measure works satisfactorily in these different contexts, this would serve as additional evidence that the measure performs as it should. The analyses conducted in this study yield only initial evidence for this. We especially recommend future researchers to, first, further investigate the (relative) explanatory power of the two measures: do original and short measures produce the same or similar results, and how is this dependent on the type of relationships investigated, contextual variables, and methodological choices (e.g., Harari *et al.*, 2016)? And, second, further validate the short measure by thoroughly investigating the overlap of the short and the long form, ideally using independent administrations (Smith *et al.*, 2000, p. 105). Besides that, future research should further investigate the interplay of general policy alienation with related (attitudinal) concepts, such as job dissatisfaction, experiences of red tape, or burnout. In this way, the incremental validity of the short measure can be tested (e.g., Brackett & Mayer, 2003), as well as the interaction of the effects of these variables on implementation willingness or organizational performance (Andersen *et al.*, 2016). The final limitation is that of common method bias, since we used variables to test construct validity collected from respondents at the same point in time. Although this is not a major problem given the measure development purpose of this study – we do not aim to claim

a causal relationship –, we do urge future researchers to conduct longitudinal or experimental studies to investigate the (long-term) effect of general policy alienation on policy implementation success and failure and trust in government, public service motivation, and job satisfaction or turnover intention.

3.5 CONCLUSIONS

This study has yielded a short measure of general policy alienation. With only five items, it allows both scholars and practitioners to account for the effects of frontline workers' overall assessment of government policies in their work. We believe this contributes to a more realistic and context-sensitive approach when investigating policy alienation, policy implementation or, more generally, the management of change in the public sector. Furthermore, the systematic 10-step approach that was followed in this study may help others researchers to develop short versions of other measures in public administration research. This strengthens the quality of quantitative (public administration) research by promoting deliberate short scale development, which prevents researchers from creating ad-hoc short measures, that makes it particularly difficult to compare research results and impairs the development of a common body of knowledge. Furthermore, it allows more (short) measures to be included in surveys, while taking full account of validity and reliability issues.

chapter 4

Determining whether consistent government policies lead to greater meaningfulness and legitimacy on the frontline

ABSTRACT

We investigate the impact of policy consistency on frontline workers' perceptions of policy meaningfulness and legitimacy. The results from an experiment involving 779 teachers indicate that policy consistency does have a positive effect on legitimacy and to a lesser extent on meaningfulness. However, the extent depends on policy content and the degree of autonomy. Overall, our findings emphasize the potential positive impact of policy consistency. Although this, to some extent, conflicts with the nature of political decision- and policymaking (i.e., democratically elected governments have been mandated to change policy), our study suggests that policy consistency could be a valuable strategy for governments to strengthen successful policy implementation. This adds a new perspective to the continuing debate within policy implementation and street-level bureaucracy research on how to account for the complex, messy and sometimes contradictory implementation of public policies.

4.1 INTRODUCTION

Teachers, nurses and police officers working on the frontline of public service delivery are often confronted with new policy programs that usually result in them having to implement new rules and regulations. This could influence the way in which they perform their tasks, with established practices being challenged. For the successful implementation of these policies, policymakers are dependent on the willingness of these frontline workers (the term 'street-level bureaucrats' is used interchangeably, e.g., Meyers *et al.*, 1998) to cooperate (Lipsky, 1980; Sabatier & Mazmanian, 1980; Hill & Hupe, 2009; Tummers *et al.*, 2009; Brodkin, 2012; Gofen, 2014; Van Engen *et al.*, 2016). These frontline workers need to tailor the new policies to their clients' needs (Sommer Harrits & Ostergaard Moller, 2014), deal with conflicting demands from different policies (Tummers *et al.*, 2015) and have discretion in doing so (Lipsky, 1980). Research has shown that their actual behaviors during policy implementation does not necessarily align with the policymakers' ambitions (Maynard-Moody & Musheno, 2003; May & Winter, 2009). In effect, frontline workers have the ability to create major difficulties for governments democratically mandated to introduce new policies.

Being continuously confronted with the consequences of political decisions that result in new policy measures – ones that are not necessarily coherent with previous policies – can be challenging for frontline workers. Often, they need to adapt to a new perspective or mind-set (e.g. the introduction of performance management systems in professional organizations; Kerpershoek *et al.*, 2016), cope with budget cuts (Kiefer *et al.*, 2015), all while having to deal with multiple accountabilities at the same time (Hupe & Hill, 2007). Research indicates this is particularly the case when a large number of new policies are introduced (Huy, 2001), when there are conflicting political signals (May & Winter, 2009) and when incompatible goals are set (Boerzel & Van Huellen, 2014).

It is therefore important to understand how frontline workers, as well as other stakeholders, perceive and experience government policies over time. In this article, we focus on the effects of policy consistency. In other words, we study how the continuity of policies over time influences frontline workers. Many, and perhaps rather capricious, inconsistent changes might generate resistance among these workers, which might influence not only the efficiency and effectiveness of

the policies involved but also their legitimacy. Although the street-level bureaucracy literature recognizes the important role of frontline workers in determining the effectiveness and legitimacy of public policy implementation (Lipsky, 1980; Freidson, 2001; Bekkers *et al.*, 2007), public administration and management research still tends to marginalize the perspectives and experiences of those who enact the policy in practice (O'Toole, 2000; DeLeon & DeLeon, 2002; Barrett, 2004; Saetren, 2005; Werts & Brewer, 2015). In particular, the micro-level (psychological) underpinnings of this (Grimmelikhuijsen *et al.*, 2017) have not been well researched although there have been recent notable exceptions (Andersen & Jakobsen, 2017; Raaphorst, 2018; Thomann *et al.*, 2018). This inattention is surprising given that policy implementation is sometimes complex and contradictory, and remains one of the main challenges facing civil servants worldwide (O'Toole, 2004; Moulton & Sandfort, 2017). This matter therefore deserves the attention of public administration and management scholars.

Putting policy into practice is not easy: it requires the investment of scarce funds and time, and not only of the organizations involved but also on a personal level. Frontline workers' 'investment decisions' are constrained, including by budgets, laws, policies, managers, social and professional norms and past experiences (e.g., Lipsky, 1980; Ewalt & Jennings, 2004; Ackroyd *et al.*, 2007; Hupe & Hill, 2007; May & Winter, 2009; O'Sullivan, 2010; Van Engen *et al.*, 2016). Consequently, when frontline workers have to decide whether to put effort into implementing a new policy, the government's past performance in maintaining policies could be an important consideration (cf. White *et al.*, 2013). This suggests that governments should not regard frontline workers as 'neutral' implementers since they bring with them a history of experienced government policy changes and, hence, ideas about the contribution of policies to a greater purpose and their added value ('meaningfulness') and how justified and appropriate these government policies are ('legitimacy'). Our premise is that policy consistency has a positive influence on how frontline workers perceive the policy programs they are required to implement and, vice versa, that policy inconsistency has a negative influence. However, little empirical research has investigated this, and this study aims to fill this gap. Our main research question is formulated as: What is the effect of policy consistency on how frontline workers perceive the meaningfulness and legitimacy of the policies they are required to implement? Given the

apparent importance that frontline workers attach to autonomy (Lipsky, 1980; Maynard-Moody & Musheno, 2003), we also investigate if, and then how, this effect is moderated by experienced autonomy.

This article is structured as follows. The next section offers some theoretical background and introduces our assumptions. Next, we outline the method adopted and describe the experimental design and the results of the experiment. The final section then presents the discussion and conclusions, focusing particularly on the theoretical implications for public administration and public policy scholars, practical implications, and future lines of research.

4.2 FRONTLINE WORKERS AND POLICY PERCEPTIONS

Policy implementation may involve a reformulation of policies that lead to unexpected outcomes or even to outright failure (e.g., Elmore, 1980; Lipsky, 1980; Sabatier & Mazmanian, 1980). Over time, several explanations have been put forward, mostly focusing on street-level bureaucracy, to explain this difference between intended and realized policy (i.e. the 'implementation gap'). These include the lack of control and monitoring (Pressman & Wildavsky, 1984), insufficient training opportunities (Kroll & Moynihan, 2015) and the discretionary power of those who enact the policy (Lipsky, 1980). For many citizens, their encounters with frontline workers are their most immediate and personal experience of state representatives (Lipsky, 1980; Maynard-Moody & Musheno, 2003). Therefore, it is seen as crucial that these workers adhere to the values of fairness, equality and equity when implementing rules that were determined through democratic procedures (Deutsch, 1975). Thus, frontline workers should, regardless of their own ideological beliefs, neutrally and loyally implement public policies (Gruber, 1987). However, numerous studies have shown that the way in which they implement public policies is often influenced by their opinions, values, preferences and world views (Kaufman, 1960; Maynard-Moody & Musheno, 2003), as well as by their perceptions of specific policies (Brehm & Gates, 1997; Meyers & Vorsanger, 2003; May & Winter, 2009; Tummers *et al.*, 2009; Van Engen *et al.*, 2016). Although there seems to be little consensus on how perceptions actually influence behaviors, there does seem to be broad agreement that they

frequently do (e.g., Ajzen & Fishbein, 1980). The transition from perceptions to behaviors is assumed to take place through a cognitive process in which beliefs about what is expected influence behaviors. Research in both public administration and psychology suggests that understanding frontline workers' perceptions of a given policy is relevant, and perhaps even a prerequisite, for understanding policy implementation behaviors (Andersen & Jakobsen, 2017). In this respect, we argue that policy consistency, or the lack thereof, influences the perceptions of frontline workers.

Before we discuss the theoretical mechanism that links policy consistency with policy perceptions, we first define policy consistency. When applied to the public policy domain, policy consistency can be defined as the degree to which government policies are constant and steady over time (based on White *et al.*, 2013; Béland & Powell, 2016; Cayton, 2017). Hence, policy consistency is associated with terms such as certainty, continuity and predictability. Consistency can relate to multiple characteristics of the policy (and associated process), including overall policy approach (e.g., should frontline workers be involved in establishing policies or should policies be decided in a top-down fashion?) and policy direction (e.g., should all students have obligatory Chinese lessons or will these be non-compulsory?). Hence, in our definition of consistency, the approach, direction or other characteristics of policies are immaterial, as long as these are consistently applied. For example, if a specific policy that funds teacher development programs is continued by successive government officials, and thus endures over the passage of time and shifts in party lines, and receives persistent funding (White *et al.*, 2013), this could be seen as a case of policy consistency. The same would be true for a consistent policy that strictly prescribes teachers' desirable classroom behaviors.

In public policy studies, the importance of consistency has been emphasized in two streams of literature. First, in the literature focusing on policy implementation in bureaucracies and its administrative processes (e.g., Dunsire, 1978), it has been argued that a government's desired policy outcomes can sometimes best be obtained through consistent policies rather than through ad-hoc ones (Calvo, 1977; Kydland & Prescott, 1977). The main argument for this is that individuals – in our study, frontline workers – form their own expectations of what will happen in the future based on what has happened in the past (Cagan, 1956). This implies

that when frontline workers form an opinion about a policy, the government's past performance in maintaining previous policies (over a significant period) is an important consideration. If the government has previously shown itself to be unwilling or unable to do so, this may have consequences for frontline workers' expectations regarding the future and, hence, for their present perceptions and behaviors.

Second, there is an extensive literature in political and economic science showing that people, including frontline workers, have a bias in favor of the status quo. That is, when faced with a complex decision, they often prefer to stick with the existing situation (i.e., the policy as it currently is) (Fleming *et al.*, 2010; Arnold & Fleischman, 2013). This 'status quo bias' is shaped by a number of complex and interacting factors, including the economic costs involved in transitioning (e.g., Samuelson & Zeckhauser, 1988). That is, when deciding about investments – such as should a teacher follow a training course to prepare for the implementation of inclusive education (Sharma *et al.*, 2006) or initiate a cross-school informal network (Coburn, 2001) – a prognosis of future policies needs to be made. If the future is uncertain, larger 'risk margins' will be factored into such decisions (White *et al.*, 2013), which may trigger a status quo bias. Policy uncertainty or inconsistency thus affects policy perceptions: how wise is it to support and invest in a policy if the likelihood is that, based on previous experiences, this policy will be abolished or changed, or a new policy will be introduced?

Therefore, we hypothesize that policy consistency has a positive effect on frontline workers' policy perceptions. Based on the discussion above, we first expect policy consistency to positively affect frontline workers' perceptions of a policy as meaningful for reaching important social goals and as valuable for their clients (Freidson, 2001; Van Engen *et al.*, 2016). In other words, policy consistency contributes to the perceived added value of government policies to frontline workers. This is consistent with previous research, which has found that a large number of policy changes (an indicator of policy inconsistency) increases the likelihood that frontline workers will not perceive policies as meaningful in achieving important societal goals (Tummers *et al.*, 2009). This is because it takes some time for frontline workers to identify with a new policy program (e.g., Elmore & McLaughlin, 1988). Further, a lack of consistency makes it hard to understand policies (Brehmer, 1974). Moreover, belief sometimes follows

action (Fullan, 1986) such that frontline workers who initially do not support a certain policy could, by working with this policy, come to recognize its added value and therefore become more supportive. Meaningfulness is important because research has repeatedly shown that meaningfulness and implementation willingness (Matland, 1995; Higgs & Rowland, 2005; Tummers *et al.*, 2012; Van Engen, 2017) and commitment to change (Van der Voet *et al.*, 2017) are strongly connected.

Second, we hypothesize that policy consistency positively affects legitimacy. Legitimacy amounts to a general confidence that the government's power to make binding decisions regarding the policy is justified and appropriate (Dahl, 1998). Where authority rests upon legitimacy, frontline workers will feel an obligation or duty to uphold laws and accept governmental decrees as legal and authoritative (Peters, 1986). It reflects a voluntary willingness to obey policy decisions (Tyler, 2006) and to trust that the government is acting in society's best interests (Levi, 1997). In terms of procedure, observers of public policy are virtually all agreed on the importance of time (Wallner, 2008). Polsby (1984) argues that time facilitates 'policy incubation', a phase in which actors can adopt the idea, adapt it and reshape it, and place it in the ongoing culture. Inconsistent, rapidly changing policies do not allow this, and this may contest their legitimacy among frontline workers. Therefore, we expect policy consistency to strengthen frontline workers' sense of duty to uphold laws and accept, and have confidence in, the governmental decrees. That is, policy consistency heightens perceived legitimacy. Summarizing the above discussion, our first hypothesis is:

Hypothesis 1: Policy consistency (as against policy inconsistency) has a positive effect on how frontline workers perceive a policy's (a) meaningfulness and (b) legitimacy.

The question is of course whether this is true in all circumstances. Street-level bureaucracy research has shown that autonomy, generally defined as the extent that frontline workers have freedom to choose among possible courses of action or inaction (also referred to as discretion), is usually valued highly by frontline workers (e.g., Lipsky, 1980; Maynard-Moody & Musheno, 2003). Their expertise (Johansson, 2012), multiple accountabilities (Hill & Hupe, 2007) and professional

values (Kerpershoek *et al.*, 2016) then allow them to adhere to equity and equality principles when putting policies into practice. Indeed, they often feel that a reasonable amount of autonomy is beneficial in achieving the public values that policies pursue. Furthermore, research indicates that if frontline workers feel that policies do not guarantee an acceptable level of autonomy, then this negatively impacts their policy support (Tummers, 2012). In relation to our first hypothesis, this apparent importance of autonomy for frontline workers raises the question if, and then how, the hypothesized effects of policy consistency on meaningfulness and legitimacy are moderated by the degree of autonomy that frontline workers perceive themselves as having; that is, their ability to choose among alternative behaviors when implementing a policy (Hoogerwerf, 1978). It could be, for instance, that the positive effects of policy consistency on meaningfulness and legitimacy are neutralized or outweighed by perceptions of low autonomy. To evaluate this, we formulate the following hypothesis:

Hypothesis 2: The positive relationships between policy consistency and a policy's (a) meaningfulness and (b) legitimacy are moderated by autonomy. These relationships are stronger if experienced autonomy is high rather than low.

It should be stressed that we do not rule out factors other than policy consistency and autonomy influencing frontline workers' policy perceptions. Indeed, previous studies have highlighted various factors that impact policy perceptions, including the influence of professional culture and organizational socialization (Oberfield, 2010; Hatmaker *et al.*, 2011) and frontline workers' political beliefs (e.g., Riccucci, 2005). Furthermore, personality characteristics, such as psychological reactance and self-efficacy, may also play a role (Bandura, 1977; Brehm & Brehm, 2013). Our goal is, nevertheless, limited to clarifying the effects of policy consistency and autonomy on meaningfulness and legitimacy, rather than to comprehensively explain the latter.

4.3 AN EMPIRICAL TEST FOR POLICY CONSISTENCY EFFECTS

4.3.1 Case

The case we selected to test the hypotheses is the Dutch secondary education sector. The Dutch education system consists of an obligatory eight years of primary education, followed by an obligatory four, five or six years of secondary education (depending on student capacities). The Dutch secondary education sector comprises around 700 schools. All schools are funded by the Dutch national government and have to adhere to the same rules and regulations (EP-Nuffic, 2015). All teachers in secondary education in the Netherlands are public sector workers. From an international perspective, decisionmaking in the Netherlands is the most decentralized of all OECD countries (OECD, 2013). The Dutch Ministry of Education is responsible for the education system as a whole, and is responsible for education quality, efficiency and accessibility (Dutch Ministry of Education, Culture and Science, 2018).

We chose this case for three reasons. First, teachers play a crucial role in delivering services. Second, the sector has experienced many problems in recent decades as a result of the reshuffling of authority and responsibilities between the ministerial and the school levels (Pijl & Frissen, 2009). Third, the sector can be characterized by numerous policy changes (Bronneman-Helmers, 2008). This makes it an appropriate case for investigating the possible effects of policy inconsistency on the perceptions of frontline workers.

4.3.2 Data collection

An experiment was conducted that involved collecting large-scale survey data in June 2016. A nationwide sample of 1.682 secondary school teachers was used. These potential respondents were all members of a large voluntary panel of Dutch public sector employees organized by the Dutch Ministry of the Interior and Kingdom Relations (subsample secondary school teachers). For more information on the panel, which has been used in other studies including Van Loon *et al.*, 2016 and Van der Voet & Vermeeren, 2017, see http://www.internetspiegel.nl. To ensure the representativeness of this large panel, the members were selected using the records of the ABP pension fund that all Dutch government employees are legally obliged to join. All the 1.682 potential respondents were sent a person-

alized e-mail with an invitation to voluntarily participate in the questionnaire. Two reminders were sent. In total, 908 respondents accepted the invitation to participate. The respondents who accepted the invitation did not differ significantly from the respondents who did not accept the invitation in terms of gender, age, function and education level.

To increase the quality of our sample, we asked the respondents to indicate whether they were presently working in secondary education. Sixteen respondents indicated they were not/no longer working in secondary education and six respondents did not provide an answer to this question. A further 20 respondents indicated they were not working as a teacher. These 42 respondents were all removed from the sample. Of the remaining 866 respondents, we further excluded all who did not meet the threshold of providing answers to at least 95% of the survey questions (in total 87). This resulted in a final sample of 779 respondents, a response rate of 46%.

4.3.3 Background characteristics and representativeness

Overall characteristics of the sample are summarized in Table 4.1 ('total sample' column). The average age of our respondents was 54 years; the youngest being 23 and the oldest 69. Of the respondents, 58% were male. Nearly all of our respondents (96%) have at least a (university) degree. On average, our respondents have worked for almost 23 years in secondary education and 8% have managerial responsibilities as section or team heads, i.e. middle managers. We compared the sample characteristics with national statistics on teaching personnel in secondary education for the 2015-2016 school year (DUO, 2016). Most notably, our respondents are on average older than the population (mean age respectively 53.6 versus 44.3) and males were overrepresented in our sample (58% versus 45.7%). Hence, our sample does not fully reflect the population (i.e., Dutch secondary school teachers) we are aiming to study. Therefore, we should be cautious in generalizing our results.

4.3.4 Experiment design

This research uses an experimental approach to explore the effects of policy consistency on frontline workers' perceptions of meaningfulness and legitimacy. Although experiments, by definition, manipulate situations (i.e. situations are

not 'real', which limits ecological validity), they do allow one to isolate and explore causal effects of interest in ways that other methods cannot (Grimmelikhuijsen *et al.*, 2017). In this way, we can get some idea of the causal effects of top-level political and policy decisions, which can subsequently be explored in the field.
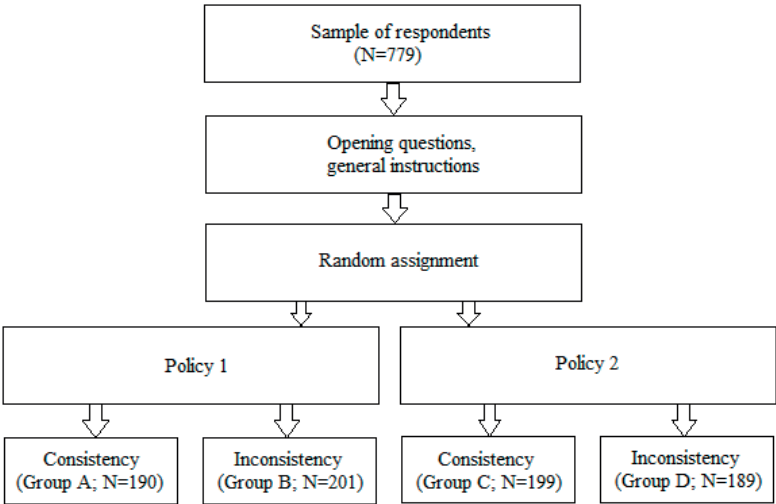

Figure 4.1 Experiment design

Figure 4.1 summarizes the experimental design of this study. First, the respondents were invited to participate in a survey. Upon accepting this invitation, the respondents were randomly assigned one of two policy measures, and to either a consistent or inconsistent outcome. In a typical fully randomized set-up, treatment and control groups have the same characteristics except for the treatment they are given (Remler & Van Ryzin, 2011). Hence, in a typical experimental set-up, respondents would first answer some questions relevant to the experimental treatment (such as general degree of trust in the government, policy consistency, implementation willingness and autonomy) before being allocated. Although we did not follow this procedure, the lack of statistically significant differences across the four groups, summarized in Table 4.1, shows that our groups are statistically equivalents.

Subsequently, the respondents were asked to carefully read a case, and answer some questions about it. Although fictitious, the political and policy decisions that we presented were relevant, authentic and inspired by real policy and political decision-making processes. This improves the ecological validity of the experiment. First, the respondents were asked to imagine that, in the current school year (2016-2017), the Secretary of State for Education (a Junior Minister) introduces a new policy. In the vignette, a rationale, based solely on research, was made for this policy so that respondents could see that there were more than just personal or political arguments in favor of it. Research has shown that policy content affects the way frontline workers perceive policies (e.g., Meyers *et al.*, 1998; May & Winter, 2009; Tummers *et al.*, 2012). Therefore, we evaluated the effect of policy consistency using two different policy cases that varied in terms of topic, policy goal and how prescriptive the government was in achieving these goals. In this way, we acknowledged that policy content might influence the relationships that we were studying. Although evaluating the influence of policy content is not the main goal of our study, including the manipulation of content improves the validity of our experiment. Further, if we find the same relationships between variables with two different policies, we can be more confident in the generalizability of our results than if there are different relationships.

Hence, we designed two fictitious policy measures, with both addressing actual policy challenges in the Dutch secondary education sector and both having direct consequences for the frontline workers and the organizations in which they work. Two prominent policy challenges in the Dutch secondary education at the time of the experiment were the professional development of education professionals and the inequality of education opportunities for children with the same intellectual capabilities but unequal family socioeconomic status (Inspectorate of Education, 2016). Policy 1 therefore focused on professional development, and policy 2 on inequality. Policy 1 consists of a government measure that provides each school with additional funding for professional development. School leaders and teachers at these schools are invited to formulate their own specific goals and to determine how they will to spend the budget. Policy 2, on the other hand, consists of a government measure that introduces a norm that all schools should adhere to in order to reduce inequality, thereby restricting professional leeway.

Table 4.1 Background characteristics and comparison across control and experimental groups

	Total sample	A	B	C	D	χ2/F*	*p* value
Categorical							
Gender						2.20	0.53
Male	58%	58%	62%	55%	57%		
Female	42%	42%	38%	45%	43%		
Education level						7.08	0.63
Appl. University	54%	55%	53%	5856%	51%		
University	42%	42%	42%	41%	42%		
PhD	3%	2%	4%	2%	4%		
Other	2%	1%	2%	1%	3%		
Position						1.20	0.75
Teacher with managing responsibility	8%	8%	9%	7%	7%		
Teacher	92%	92%	91%	93%	93%		
Nominal/ordinal							
Age	53.60 (9.28; 23-69)^	55.08	53.82	52.44	53.09	0.17	0.68
Tenure	22.39 (11.03; 1-45)	24.19	21.40	22.20	21.84	3.93	0.05
Trust in government	2.03 (0.60; 1-4)	1.99	2.05	2.06	2.03	0.06	0.81
Policy consistency	2.01 (0.67; 1-5)	2.02	2.01	2.00	2.03	0.04	0.83
Implementation willingness	3.18 (0.82; 1-5)	3.17	3.25	3.16	3.12	0.12	0.73

* χ2 reported for categorical variables (gender, education level, position); F reported for nominal and ordinal variables; ^ Respectively standard deviation and minimum and maximum scores.

Next, we stated that a new government would be formed in 2017, with a new Secretary of State for Education. As elections for the Dutch House of Representatives were scheduled for March 2017, it was realistic that a new government would be formed in 2017. Then, we indicated that one of the first debates facing the new Secretary of State in the House of Representatives would concern a policy decision of the former Secretary of State. A member of the new House of Representatives makes a statement suggesting that the new Secretary of State should end the policy which was only recently introduced. As such, the statement was encouraging the Secretary of State to act inconsistently (i.e., by discontinuing the policy). The new Secretary of State responds to this suggestion and, depending on the group to which the respondent is randomly assigned, decides either (1) to continue the policy (policy consistency) or (2) to discontinue the policy

(policy inconsistency). To ensure that respondents given the consistency condition would not score higher on meaningfulness and legitimacy than respondents with the inconsistency condition simply because a decision to continue could be interpreted as a positive policy evaluation, we indicated that the reason why the new Secretary of State wants to continue the policy is simply because nothing is yet known about the policy outcomes. Furthermore, it should be noted that the Secretary of State had to decide between continuing or discontinuing the policy as it is. That is, we did not include the option to make changes and then continue with the policy. Finally, the responses of the new Secretary of State provided to the respondents did not differ by more than 5% in the number of words used, and we also ensured that the responses had the same number of sentences. This was to ensure that differences between the groups could not be caused by such language differences rather than the experimental treatments. All the vignettes are provided in Appendix III.

4.3.5 Measures

Unless indicated otherwise, all scales were formulated using five-point Likert-type items. The main measures are discussed below and a summary of all the items for each measure is provided in Appendix IV.

Policy meaningfulness

We evaluated the perceived meaningfulness of the decision of the Secretary of State to (dis)continue the policy with an index that consists of three policy meaningfulness items taken from the policy alienation questionnaire (Tummers, 2012). These items were tailored to meet the specific goal of our study. For instance, the item 'I think that the policy, in the long term, will lead to goal 1' in this study becomes 'I think that the decision of the new Secretary of State, in the long term, will lead to greater professionalization' (with policy 1). A second example is the item 'Overall, I think that the policy leads to goal 1' which, in this study, becomes 'Overall, I think that the decision of the new Secretary of State leads to greater equality' (with policy 2). The Cronbach's alpha for the three-item scale is 0.90. Given that the items used reflect the fact that the two policy measures have different goals means that scores for these measures cannot be directly compared. This is not problematic since we are only interested in establishing

the differences between meaningfulness scores related to the consistency/inconsistency condition.

Legitimacy

A voluntary willingness to obey and accept (authoritative) decisions is connected to legitimacy (e.g., Tyler, 2006), which means that a stated willingness to accept a decision can be treated as an empirical indicator of legitimacy (De Fine Licht, 2014). Therefore, we operationalized legitimacy as policy acceptance, which we measured with three items. The first two items are 'What do you think of the decision of the new Secretary of State?' and 'How willing are you to accept the decision of the new Secretary of State?'. Since these items measure a somewhat passive reaction to a political decision (De Fine Licht, 2014), we included the more active self-reported likelihood of protesting the decision as a third item in the measure: 'How likely do you think it is that you will protest against the decision of the new Secretary of State?' (R). The Cronbach's alpha of this scale was 0.79.

Autonomy

This study measures autonomy with a four-item scale extracted from the 'operational powerfulness' dimension of the general policy alienation questionnaire (Van Engen *et al.*, 2016). The items include 'Generally, I have freedom to decide how to use government policies' and 'Generally, when working with government policies, I can be in keeping with clients' needs'. Here, the Cronbach's alpha was 0.73.

Stimulus control: manipulation check

The consistency manipulation check consisted of the four-item policy consistency measure of Van Engen *et al.* (2016). These items were tailored to match the specific goal of our study. For instance, the item 'To what extent do you have the impression that policy by the Ministry of Education is (a) consistent and (b) focuses on the long term', in this study becomes 'The new Secretary of State is (a) consistent and (b) focuses on the long term'. The Cronbach's alpha of this four-item scale was 0.88.

4.4 RESULTS

In this section, we report the results of the analyses. First, we conducted tests to evaluate whether our experimental manipulation had worked. We expected respondents in the consistency group to score higher on perceived consistency than respondents in the inconsistency group. Here, Table 4.2 shows that this is indeed the case (with the means of the consistency and inconsistency group being 3.22 and 2.21 respectively). This provides sufficient confirmation that the experimental manipulation worked. Next, we also analyzed the mean scores for meaningfulness and legitimacy (using ANOVA). The results are again shown in Table 4.2 and provide evidence that supports hypothesis 1 as the means for meaningfulness and legitimacy are higher in the consistency group (respectively 2.88 and 3.66) than in the inconsistency group (respectively 2.40 and 2.99), In other words, policy consistency, as compared to policy inconsistency, has a positive effect on how frontline workers perceive policy meaningfulness and legitimacy. All these differences are statistically significant.

Table 4.2 Means of manipulation check and dependent variables

	Complete sample^	Consistency group	Inconsistency group	t	Consistency (A)	§Inconsistency (B)	Consistency (C)	Inconsistency (D)	F
					Policy 1		**Policy 2**		
Manipulation check									
Perceived consistency	2.72 (0.96; 1-5)	3.22	2.21	16.87*	3.28	2.03	3.17	2.40	104.16*
Dependent variables									
Meaningfulness	2.64 (0.93; 1-5)	2.88	2.40	7.28*	3.01	2.12	2.76	2.72	35.91*
Legitimacy	3.32 (1.08; 1-5)	3.66	2.99	8.96*	3.90	2.64	3.43	3.38	54.73*

* $p<0.01$. ^Between brackets, respectively standard deviation and minimum and maximum scores.

Table 4.2 further shows that, when confronted with policy 1, respondents in the consistency condition group perceive policy meaningfulness as significantly higher than respondents given the inconsistency condition (means are respectively 3.01 and 2.12; $p<0.01$). These teachers thus believe that the decision to con-

tinue the policy is meaningful in that it will enhance professional development, which, ultimately, is the goal of the policy. Further, these respondents have the impression that the decision to continue the policy is more legitimate than the decision to not continue the policy was perceived by that group of respondents (means are respectively 3.90 and 2.64; $p<0.01$). Similarly, respondents given the consistency condition perceive policy meaningfulness when confronted with policy 2 as slightly higher than respondents seeing the inconsistency condition, but this difference is not statistically significant (means are 2.76 and 2.72 respectively). The same is true in relation to legitimacy: i.e. respondents given the consistency condition score higher on legitimacy than respondents considering the inconsistency condition, but this difference is very small and not statistically significant (means are 3.43 and 3.38 respectively). This shows that policy content affects perceived meaningfulness and legitimacy. As a robustness check, we conducted ANCOVA and included age, gender, tenure, position, trust in government, policy consistency and implementation willingness as covariates. This supported the results presented above.

Our theoretical arguments argue in favor of a moderating effect of policy autonomy on the relationships between consistency and meaningfulness and legitimacy. To further understand this effect and the moderating effect of policy content, we conducted regression analyses. The results are shown in Table 4.3.

First, Table 4.3 shows that consistency has a positive effect on meaningfulness and on legitimacy in model 1 (treatment only) (respectively $\beta=0.52$ and $\beta=0.65$; $p<0.01$). This is in line with the results we presented in Table 4.2 and provides support for hypothesis 1. Second, the results indicate that policy content also affects perceptions of meaningfulness and legitimacy, although this effect is less strong and only statistically significant for meaningfulness ($\beta=0.16$; $p<0.01$). It would thus seem that evaluations of meaningfulness at least partially depend on the specific policy that is (dis)continued. In model 2, we add autonomy. The results indicate that the degree of autonomy also influences meaningfulness and legitimacy: the more autonomy frontline workers experience, the more they feel that policies are meaningful and legitimate. The results show that the effect of autonomy is weaker than the effect of consistency.

Table 4.3 Analyses with meaningfulness (M) and legitimacy (L) as dependent variables

	Model 1: Treatments only		Model 2 Model 1+ discretion		Model 3: Model 2 + interaction consistency and policy content		Model 4: Model 2 + interaction consistency and autonomy		Model 5: Full model	
	M	**L**	**M**	**L**	**M**	**L**	**M**	**L**	**M**	**L**
Consistency *ref=inconsistency*	0.25**	0.31**	0.24**	0.30**	0.47**	0.58**	0.24**	0.30**	0.47**	0.58**
Policy content *ref=policy 1*	0.16*	0.06	0.09*	0.06	0.33**	0.34**	0.09*	0.06	0.33**	0.34**
Autonomy			0.11*	0.11**	0.12**	0.12**	0.04	-0.01	0.05	-0.01
Consistency* Policy content					-0.41**	-0.49**			-0.42**	-0.49**
Consistency* autonomy							0.10	0.18**	0.05	0.18**

** $p<0.01$; * $p<0.05$; N=779. Standardized coefficients are reported.

In models 3 and 4, we added interaction effects to the analyses. Model 3 shows that the relationship between consistency and meaningfulness is significantly moderated by policy content (respectively β=-0.41 and β=-0.49; $p<0.01$): if the interaction term is added to the model, the direct effects of consistency and of policy content become stronger. However, the results from model 4 indicate that the interaction between consistency and autonomy is only statistically significant for legitimacy (β=0.18; $p<0.01$) and not for meaningfulness.

In the fifth, full model, both interactions are included. To more easily understand these interaction effects, we present them in graphical form. The results for meaningfulness and legitimacy are shown in Figures 4.2 and 4.3 respectively. First, for meaningfulness, Figure 4.2 shows that policy consistency results in greater meaningfulness than policy inconsistency. Furthermore, we see differences depending on policy content. Consistency has a stronger and more positive effect with policy 1 (professional development) than with policy 2 (educational inequality). Moreover, the degree of autonomy that a teacher experiences makes a difference. The positive effect of consistency on the meaningfulness of policy 1 is slightly stronger if the respondent experiences high rather than low levels of autonomy. However, with policy 2, we see a different effect: if experiencing low autonomy, policy consistency has a negative effect on meaningfulness; whereas with high autonomy the effect of policy consistency is positive (but small). A similar analysis for legitimacy produces similar but stronger effects than those

found for meaningfulness. Confronted with policy 1, policy consistency has a strong effect on legitimacy, and this effect is slightly stronger for respondents who experience high rather than low autonomy. Confronted with policy 2, the effect of consistency is only positive for those who experience high autonomy and negative for respondents experiencing low autonomy.

It is noteworthy that we have found different effects of consistency on both meaningfulness and legitimacy depending on policy content and the experienced autonomy of the respondents. Can we better understand these results if we look at the fictitious policy measures we introduced? As explained in our experimental design, we included two policy cases differing in topic, policy goal and how prescriptive the government was in the desired approach. Perhaps not surprisingly, the positive effect of consistency is less strong for the policy that has a more top-down approach to what should be done (policy 2 on educational inequality). Particularly for respondents who already experience low autonomy; the effect of consistency is even negative.

Overall, the results of our analyses support hypothesis 1: policy consistency, as expected, has a positive effect on teachers' perceptions of meaningfulness and, particularly, of legitimacy. The results partially confirm hypothesis 2: the positive effect of consistency is stronger if autonomy is high, but only statistically significant for legitimacy. By studying these relationships with respondents who had been confronted with different policies, our results show that the continuation of certain policies (in our experiment: a policy measure that restricts professional leeway and discretion), but not all, has a negative effect on meaningfulness and legitimacy for respondents experiencing low autonomy. This suggests that policy consistency should not be seen as a 'one size fits all' solution.
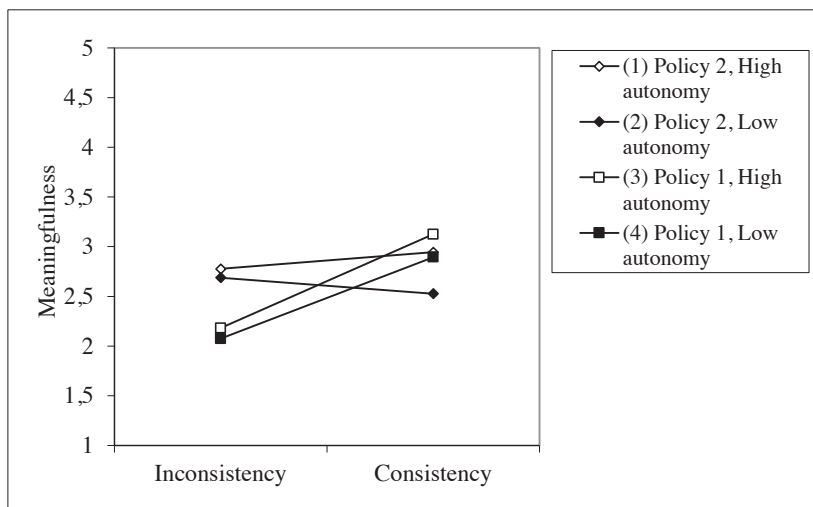
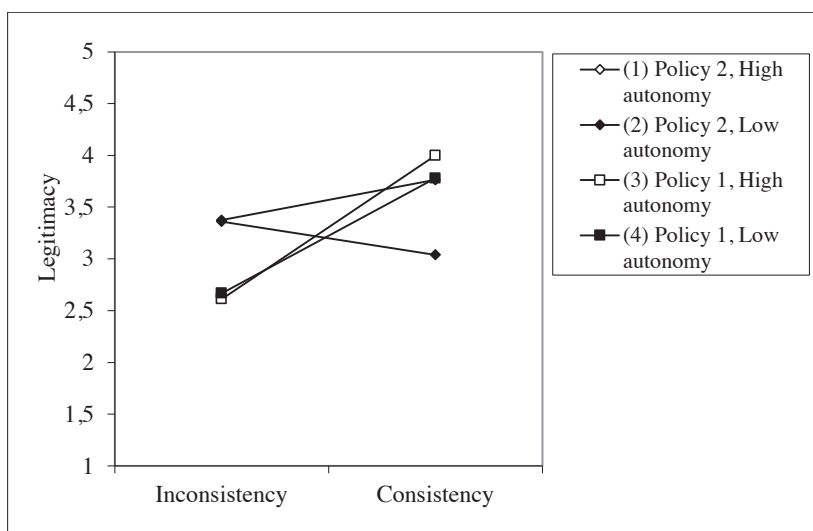**Figure 4.2** Interaction effects for meaningfulness

Figure 4.3 Interaction effects for legitimacy

4.5 DISCUSSION

We have investigated the effect of policy consistency on how frontline workers perceive policy meaningfulness and legitimacy. We found, in line with our expectations, that policy consistency positively affects perceptions of meaningfulness and, particularly, of legitimacy. Nevertheless, our results also indicate that policy consistency should not be regarded as something government should always aim for: greater consistency is not a 'one size fits all' solution or a panacea for all governmental implementation challenges on the frontline.

To single out the effect of policy consistency, and how this is affected by autonomy and policy content, we designed an experiment in which we confronted Dutch teachers with a political decision to continue (indicating consistency) or discontinue (indicating inconsistency) a policy. We also tested how the relationships between policy consistency and both perceived meaningfulness and legitimacy are influenced by autonomy. It would seem that the more autonomy that frontline workers experience, the stronger the positive effect of policy consistency. Furthermore, our results indicate that policy content is a relevant factor to consider when studying the effects of policy consistency, as our findings differ for the two policy measures with which we confronted the teachers.

To summarize, our findings emphasize the potentially positive impacts of policy consistency on perceived meaningfulness and on legitimacy. Furthermore, our results suggest that frontline workers might find policy consistency more important than their own autonomy during policy implementation, although more empirical research is necessary to confirm this impression. Although to some extent at odds with the nature of political decisionmaking and policymaking, our study suggests that aiming for policy consistency might be a useful strategy for governments aiming to improve public service delivery – perhaps even more useful than increasing autonomy –, given its ability to increase policy meaningfulness and government legitimacy among frontline workers. Although frontline workers may not find a specific policy meaningful, or see it as the best way to address societal challenges and create public value, they appear to be more likely to support this policy if they know – possibly from previous experience – that the government is willing and able to maintain this policy over time. Interestingly, our results suggest that frontline workers who experience greater autonomy are

more likely to appreciate consistency than frontline workers who experience less autonomy. This finding illustrates an interesting paradox: although politicians have full democratic and legal authority to introduce inconsistent policies (if supported by a majority in the House of Representatives), this can make it more difficult for administrators to successfully implement these policies. Rapid and inconsistent changes have a negative impact on frontline workers' perceptions of these policies and the government's legitimacy, and may even make them cynical or indifferent (Van Engen *et al.*, 2016). This conclusion aligns with previous studies that have shown that 'what you see (in terms of formal policy) may not be what you get (in terms of policy-as-produced)' (Brodkin, 2012, p. 943) and studies that have concluded that consistency heightens organizational rule-following (Borry *et al.*, 2018).

Naturally, the aim of our study has never been to claim that policies should not be changed. Policies must certainly be flexible and sufficiently responsive to adapt to new technologies, changing circumstances and societal developments (Cayton, 2017). Not least because research has shown that policymakers benefit from being seen to act by their citizens (i.e., potential voters), even if the problem gets worse (Olsen, 2017). However, inconsistent policies may have negative consequences for policy implementation. Although we recognize that policy change can be sensible, we would advise governments who want to change their policy to take the frontline perspective into account when doing so (De Boer & Eshuis, 2018; Lavee *et al.*, 2018). This implication of our study highlights a relevant and unsolved public administration dilemma: what may be regarded as perfectly legitimate and efficient from a top-down perspective may be regarded as entirely illegitimate and inefficient from a bottom-up point of view (Sabatier, 1986; Brodkin, 2012; Gofen, 2014; Alon-Barkat & Gilad, 2016). Nevertheless, the reality is that public values can only be achieved if governments and frontline workers cooperate and align their interests for the sake of society (Bryson *et al.*, 2015). If this is not achieved, and divergent perspectives and behaviors result, core public values might be put at risk. It is crucial that frontline workers adhere to the values of fairness, equality and equity when implementing policies that were decided upon through democratic procedures (Brehm & Gates, 1999).

The main findings of this study suggest a number of relevant future research questions. The first is how do frontline workers respond to and prepare for major

shifts in policy: 'when do they believe the implementation pain is worth the gain?' Unforeseeable events and developments (such as a sudden influx of large numbers of school-age refugees or a growing teacher shortage) will obviously require policy changes, but how can the resulting 'inconvenience' for frontline workers be minimized by practicing due diligence when developing policies? In other words, under what circumstances will frontline workers support policy changes, or what specific actions can governments take so that policy changes meet the criteria of being consistent and logically coherent with previous policies? It would be especially interesting to investigate the effect of policy consistency over time: do the short-term effects we found in this study also hold in the longer term? It could be wise for governments to discontinue a specific policy that frontline workers do not support. However, if governments do this repeatedly, this might trigger 'policy cynicism': "Bend over, here it comes again" (Connel & Waring, 2002).

The second topic for further research, and related to the first, is that future experiments should recognize that policies are often changed or fine-tuned during the implementation process. That is, new policies or policy changes are often intended to either refine or complement already existing policies to adapt them to (un)anticipated implementation circumstances, a lack of results or evolving political needs (Van Gunsteren, 1976; Wildavsky, 1979; Thelen, 2004; Pollitt & Bouckaert, 2011). In this study, we investigated the possible effects on frontline workers of quite a radical policy change. However, would we find similar results if the government decided to implement more incremental policy changes? We would recommend future researchers studying this topic to also pay attention to the 'rules versus principles debate' that postulates that it might be difficult to consistently apply policy if policies lack specific guidance and rules, and are mostly based on principles. The implementation of such policies by frontline workers is inherently inconsistent (e.g., Wüstemann & Wüstemann, 2010). Although we defined policy consistency as consistency over time (i.e., continuity) in this study, it would be relevant to investigate whether we would find similar effects of another subtype of consistency, namely consistency in terms of alignment with other policies (i.e., coherency), on meaningfulness and legitimacy.

The third topic where we see further research as valuable is the influence of frontline workers' personal characteristics (including their political and moral beliefs and their values), as well as the characteristics of the organization they

work for. In this study, we have established that policy consistency influences policy meaningfulness and legitimacy, and that these relationships depend on autonomy. However, elsewhere, it has been shown that meaningfulness and legitimacy also depend on other personal and organizational characteristics (e.g. Tummers *et al.*, 2009; De Fine Licht, 2014). Future research could seek to shed light on this, ideally by conducting a natural field experiment (Grimmelikhuijsen *et al.*, 2017). It could, for instance, be that policy consistency has a stronger effect on frontline workers who are more risk-averse or in organizations where general trust in government is low. It would be welcome if future research could disentangle these effects.

As all studies, this study has its limitations related to both internal and external validity issues. The first limitation is that we conducted an experiment in one sector in one country. Ideally, our experiment would be replicated in other sectors and other countries to assess whether the relationships found in this study also hold elsewhere. A second limitation is that we used a sample made up of volunteers. Although this is not uncommon in public administration studies, there are drawbacks. First, the sample might not necessarily be fully representative of the entire population (as is the case in this study). The second limitation is that we used a survey experiment with hypothetical – albeit realistic – scenarios to assess the effects of policy consistency on perceived policy meaningfulness and on legitimacy. Third, we only investigated the effect of policy consistency on frontline workers, in this case teachers. Although frontline workers are key actors in policy formulation and implementation, we know that other relevant stakeholders, including professional organizations and organizational managers (such as school leaders), should ideally support the introduction or reform of policies, or should at least be non-obstructive (Park & Rethemeyer, 2014; Bryson *et al.*, 2015). Therefore, future research should test whether this study's findings also apply to other stakeholders. Finally, we operationalized legitimacy as policy acceptance. Although this is not uncommon, it should be noted that policy acceptance involves only one aspect of legitimacy. Hence, future research on the relationship between consistency and legitimacy could focus on aspects that move beyond acceptance, including moral and normative approval (Christensen *et al.*, 2016).

4.6 CONCLUSIONS

In conclusion, this chapter has tested the influence of policy consistency on frontline workers' perceptions of policy meaningfulness and the legitimacy of government actions. Specifically, we looked at the effect of policy consistency, in terms of continuity and steadiness, over time. The results of our experimental study show, first, that policy consistency has an overall positive effect on both meaningfulness and perceived government legitimacy and, second, that this effect is enhanced if frontline workers experience greater autonomy. Our findings thus argue in favor of frontline workers having a status quo bias, which is likely to influence the success of new policy implementations. Overall, our study contributes to a better understanding of why frontline workers may create major difficulties for new governments democratically mandated to change policy. At the same time, our study emphasizes the importance of consistency in improving frontline workers' policy perceptions. Yet, it also nuances this statement by showing how frontline workers' evaluations of a policy that is (dis)continued may play a role. Consistency may be less important for frontline workers if they do not support the policy. These findings provide valuable information for governments striving to improve public service delivery.

chapter 5

Evaluating how powerfulness and meaningfulness affect implementation willingness

ABSTRACT

The topic of discretion continues to be hotly debated in policy design and policy implementation. In top-down theories, discretion at the frontline is often seen as a control problem: discretion should be avoided as it can mean that the policy is not implemented as intended. Conversely, bottom-up theories state that discretion can help policy implementers tailor a policy to specific circumstances. However, there has been little systematic research into how the experience of having discretion motivates frontline workers to implement a policy. In this chapter, we conceptualize and test this relationship by combining public administration and motivation literature, using datasets in healthcare and education and large-N set-theoretic configurational analysis. Results robustly show that experiencing discretion is a quasi-necessary condition and, hence, a prerequisite for high implementation willingness. This finding is more in line with bottom-up than with top-down theories. Policy implementers need the freedom to adapt the program to local conditions for being motivated to implement a policy. The evidence encourages scholars and practitioners to move from the question whether frontline workers should be granted discretion to how to best make use of frontline workers' discretion instead.

5.1 INTRODUCTION

"The closer one is to the source of the problem, the greater is one's ability to influence it; and the problem-solving ability of complex systems depends not on hierarchical control but on maximizing discretion at the point where the problem is most immediate."

- Richard Elmore (1979)

Discretion is the freedom to decide what should be done in a particular situation. Repeatedly, research has shown that frontline workers - also referred to as public professionals or street-level bureaucrats - have an important role in the successful implementation of policies as they inevitably retain some degree of discretion (Davis, 1969; Lipsky, 1980; Hupe & Hill, 2007; Maynard-Moody & Musheno, 2012; Gofen, 2014; Barnes & Henly, 2018). However, implementation theory has always held contradictory views on the exact role of discretion (Thomann *et al.*, 2018b). Top-down perspectives treat deviations from the policy-on-paper as a control problem: room for interpretation makes it increasingly likely that policy means and ends will be mismatched (Howlett, 2004, p. 5). Conversely, bottom-up perspectives put frontline workers' discretion at the center stage of policy implementation (Lipsky, 1980; Sabatier, 1986). As the above quotation by Elmore illustrates, from this perspective frontline workers are seen as de facto policymakers. Discretion helps them to tailor a policy to specific circumstances.

Although research has moved on to hybrid, integrative frameworks, the discussion surrounding discretion at the frontline never lost its practical salience for policy design and implementation (Howlett, 2004; Hupe & Hill, 2007; Hupe, 2013). Scholars continue to discuss the reasons why frontline workers use their discretion in more or less beneficial ways for clients and public goals (e.g., Keiser, 1999; Brodkin, 2011; Maynard-Moody & Musheno, 2012; Thomann, 2015). However, little attention has been paid to the implicitly assumed link between frontline workers' discretion and the motivation to implement government policies. This is surprising, given that "research performed in ignorance of the understanding that implementing actors have about their circumstances is likely to miss important parts of the explanation" (O'Toole, 2000, p. 269).

To explore the motivational effects of discretion, this article draws on the logic of the seminal Thomas theorem: "If men define situations as real, they are real in their consequences". (Thomas, 1928, p. 572; see also Lewin, 1986). We focus on the perceived degree of discretion, instead of the objective degree of discretion, and investigate to what extent frontline workers experience discretion. We operationalize perceived discretion via the concept of powerfulness as developed in the policy alienation literature (Tummers, 2011; Loyens, 2015; Thomann, 2015; Van Engen *et al.*, 2016; Van der Voet *et al.*, 2017). Hence, powerfulness is seen as discretion as perceived by frontline workers. We define powerfulness more formally as frontline workers' perceived influence on decisions concerning the policy. The research question of this article is then: *How does powerfulness motivate frontline workers to implement policies?*

Psychologists suggest a positive link between powerfulness and motivation (Gagné & Deci, 2015). However, scholars studying policy implementation have not found a strong, consistent relation between powerfulness and implementation willingness (Tummers, 2011; Loyens, 2015; Thomann, 2015; Van Engen *et al.*, 2016). Contrary to these previous studies, we rely on an asymmetric explanation of policy implementers' motivation: the things that motivate people may be different from those that demotivate them (Herzberg *et al.*, 1959; Matzler & Renzl, 2007). Accordingly, we study two interpretations of the motivational role of powerfulness. The first interpretation argues that powerfulness is quasi-necessary, although on its own not sufficient to motivate employees (Herzberg *et al.*, 1959; Goertz & Starr, 2003; Lammers *et al.*, 2016). The second interpretation is that powerfulness is only motivating when the public policy is consistent with the frontline workers' values and, hence, perceived as meaningful (May *et al.*, 2004; Dias & Maynard-Moody, 2007; Grant & Berry, 2011).

We study these interpretations using two large samples. By doing so, this study makes two contributions to the literature. It adds to theory by clarifying a core aspect of the top-down versus bottom-up debate: is discretion beneficial for policy implementation? It does so by connecting the policy implementation literature with the motivation theory from Herzberg. Methodologically, it uses state-of-the-art tools specifically designed for capturing the hypothesized asymmetric patterns: large-N set-theoretic configurational analysis using fuzzy sets, combined with formal theory evaluation, measures of uncertainty and system-

atic robustness tests (Ragin, 2000; Schneider & Wagemann, 2012; Misangyi *et al.*, 2017).

In the next section we will introduce our theoretical framework and the hypotheses. We then introduce our methods, the research design and the data collected among 1.004 healthcare workers and 1.087 secondary school teachers in the Netherlands. After presenting the results, we conclude and discuss how our results can inform public administration scholars and practitioners.

5.2 THEORETICAL FRAMEWORK

The concept of discretion often serves as an umbrella term for different aspects of bureaucratic practice. In policy implementation research specifically, discretion concerns the extent of freedom that frontline workers have to choose among possible courses of behavior when implementing policies (Davis, 1969; Hupe, 2013). Top-down approaches emphasize the degree of freedom granted by a rule maker to an implementing actor ('discretion-as-granted'; Howlett, 2004). Contrary to this, bottom-up approaches presuppose an inevitable existence of discretion and analyze how the degree of freedom is actually used by frontline workers ('discretion-as-used'; Hupe, 2013).

Next to discretion-as-granted and discretion-as-used, we argue that there is also a key role for discretion-as-perceived: the degree to which frontline workers perceive to possess discretion. According to the Thomas theorem, people often feel and behave based on their perceptions of reality, not on the basis of reality itself (Thomas, 1928). This perspective highlights the importance of policy-related attitudes for frontline policy implementation (Ewalt & Jennings, 2004). The Thomas theorem suggests that discretion-as-used presupposes discretion-as-perceived. Frontline workers should feel that they have discretion before they can actually use it. For instance, a social worker should feel that she can grant an exception to a rule before actually doing this. Street-level bureaucracy scholars have recently begun to explore discretion-as-perceived under the heading of policy powerfulness, meaning the perceived degree of influence that frontline workers have over shaping a policy during its design and implementation (Tummers *et al.*, 2009). This power may be exercised at the strategic, tactical or

operational level. High policy powerfulness thus indicates perceived discretion; the absence of powerfulness (i.e., power*less*ness) indicates a lack of perceived discretion.

We can then connect discretion-as-perceived - here conceptualized as policy powerfulness - to implementation willingness. To actually achieve policy goals, frontline workers should be willing to implement the policy (Ewalt & Jennings, 2004; Van der Voet *et al.*, 2017). High willingness to implement means that frontline workers intend to put effort in executing the policy. Bottom-up theories assume that discretion is positively linked with successful implementation. Note, however, that what exactly success entails might differ from a bottom-up or top-down view. Conformance implementation refers to the degree to which the centrally decided blueprint is implemented from top to down ('implementation success'). From the bottom up, performance implementation means that a policy achieves outcomes that resolve the policy problem at stake ('policy success'; Barrett & Fudge, 1981). Arguably, implementation willingness matters for both conformance and performance implementation.

The positive link between discretion and implementation willingness assumes that policy powerfulness can have a motivational effect on frontline workers. Scholars agree that perceptions can, and often do, influence behavior (e.g., Ajzen & Fishbein, 1980). Experiencing powerfulness is one of the main factors stimulating employees' willingness to support a change (Greenwood *et al.*, 2002). Related to this, the policy alienation framework asserts that as frontline workers' policy powerfulness increases, their support for a policy can increase as well (Tummers *et al.*, 2009). This powerfulness can be experienced at either the national (strategic), organizational (tactical) and client (operational) level, or a combination of these. For instance, if a frontline worker has the impression she – or her colleagues or representatives of a professional organization – is able to influence the content of policies at the national level she is more likely to be motivated to implement the policy (Tummers *et al.*, 2015). This is because it is more likely then that frontline workers' interests and concerns are reflected in the content of the policy.

Next to powerfulness, policy alienation has a meaningfulness dimension. Meaningfulness concerns the perception of the frontline worker that the policy is valuable for society in general (societal meaningfulness) and for the direct clients

of the frontline worker (client meaningfulness). Perhaps contrary to expectation, in empirical tests the relation between powerfulness and implementation willingness appears either as weaker than between meaningfulness and implementation willingness (Van Engen *et al.*, 2016), as ambiguous (Loyens, 2015; Thomann, 2015), or as non-significant (Tummers, 2011).

In light of these puzzling empirical findings, we suggest two alternative interpretations of the motivational link of powerfulness on implementation willingness. Previous research has assumed symmetric effects, where the same change in implementation willingness is expected both when powerfulness is added and when it is taken away. Contrary to this, motivation theory as developed by among else Herzberg *et al.* (1959, see for recent discussions Bassett-Jones *et al.*, 2005; Matzler & Renzl, 2007; Sachau, 2007) suggests the effects of particular motivational factors are asymmetric. It is a fundamental insight from motivation theory (Herzberg *et al.*, 1959) that the things that motivate people are often different from the things that demotivate them. For instance, a low salary makes you dissatisfied. However, a high salary does not automatically make you satisfied. This means that the influence of policy powerfulness might work only, or mainly, in one direction. Thus, the change in implementation willingness might not be of the same magnitude or direction when powerfulness is added as when it is taken away. To detect such patterns, an empirical method is needed that models asymmetric effects. This is why we choose a new, set-theoretic method that enables us to model asymmetric explanatory patterns (Misangyi *et al.*, 2017).

5.2.1 Interpretation 1: Policy powerfulness is a necessary condition

The first interpretation linking powerfulness and implementation willingness builds upon the idea that discretion is a prerequisite for policy success (Matland, 1995). If this is the case, then frontline workers need to feel able to influence the policy to be willing to implement that policy; they need to feel powerful. Hence, powerfulness is a necessary condition for implementation willingness.

Policy implementation literature, especially the studies departing from the bottom-up perspective, suggests that an important factor in this willingness of frontline workers is the extent to which organizations are willing and able to delegate decision-making authority to the frontline (Meier & O'Toole, 2002; Tummers & Bekkers, 2014). This influence may be particularly pronounced in

frontline workers whose expectations of discretion and discretion contradict notions of bureaucratic control (Freidson, 2001). As we study teachers and healthcare workers, this seems to be particularly important. Maynard-Moody and Portillo (2010, p. 259) note, "Street-level workers rely on their discretion to manage the physical and emotional demands of their jobs. They also rely on their discretion to claim some small successes and redeem some satisfaction".

The enabling role of powerfulness for implementation willingness can be traced back to the human relations movement (McGregor, 1960). One of the central tenets of this movement is that employees have a right to give input into decisions that affect their working lives. Employees enjoy carrying out decisions they have helped create – as compared to decisions they have not helped create or were 'forced upon them'. As such, the human relations movement argues that when employees experience discretion during their work, this will positively influence several job indicators, such as implementation willingness, loyalty or responsibility, by fulfilling intrinsic employee needs (for more detailed discussions, see for instance Yukl & Becker, 2006). This mechanism was already proposed by Follet (1924) - her work presaged the rise of the human relations movement - who underscored the importance of leaders having the capacity to increase the sense of power among those led. So that those led, in turn, would be empowered to achieve desired changes at the organizational, community or policy level.

The above argumentation suggests that frontline workers need to feel powerful in order to be willing to implement the policy. However, feeling powerful alone may not be sufficient. Many other factors can influence the willingness of frontline workers to implement a particular policy. This can include resources available in the organization (for instance, is there enough manpower available to make a policy work) or the value of a policy for society and political processes within organizations (O'Toole, 2000; May & Winter, 2009; Thomann, 2015). Hence, frontline workers need to feel powerful, but feeling powerful is not enough. This asymmetric interpretation accounts for the fact that not all frontline workers will use their discretion to contribute to successful implementation. Contrary to a symmetric effect, we hence expect that discretion-as-perceived has an enabling effect for motivating frontline workers (Goertz & Starr, 2003).

Accordingly, we can derive the first hypothesis. In order to be motivated to implement a public policy, frontline workers need to perceive that they have the

power to influence the shaping of a policy program (powerfulness). They should experience this powerfulness at least at either the strategic, tactical or operational level in order to feel motivated for policy implementation (Van Engen *et al.*, 2016). Still, this powerfulness does not by definition result in high implementation willingness. Hence, frontline workers with high implementation willingness are a subset of those frontline workers who experience powerfulness. We hypothesize that *policy powerfulness (either strategic: SP; tactical: TP or operational: OP) is a quasi-necessary, but not sufficient condition for high implementation willingness (W)*. This is shown in Figure 5.1. To formalize this first hypothesis, the backward arrow ← means 'is necessary for' and '+' denotes the logical 'OR'.
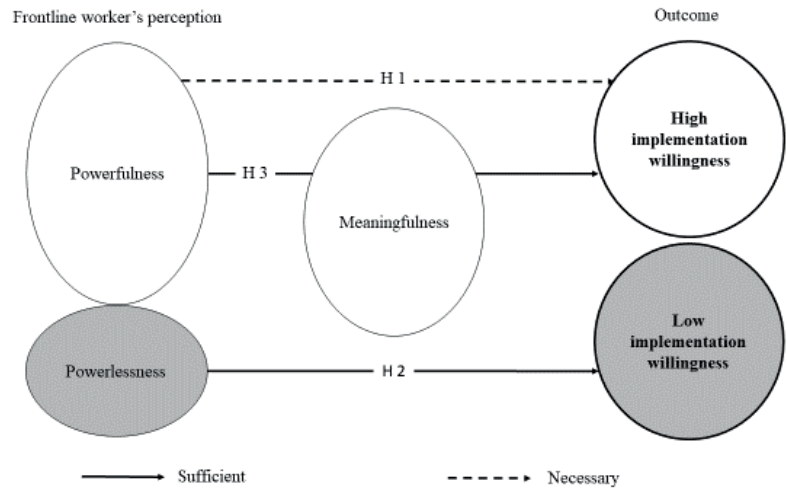
Hypothesis 1: SP + TP + OP ← W

Figure 5.1 Hypotheses

Similarly, we expect that frontline workers who do not feel powerful are typically unwilling to implement government policies. As Figure 5.1 illustrates, if high implementation willingness requires the presence of powerfulness, then the frontline workers who do not feel powerful are a subset of those frontline workers with low implementation willingness. Since powerfulness is indicated by either strategic, tactical or operational powerfulness (or a combination of these three),

all three have to be absent to indicate the absence of powerfulness (Schneider & Wagemann, 2012). Our second hypothesis thus states that *the absence of the combination of strategic, tactical and operational powerfulness is quasi-sufficient for low implementation willingness.* The '*' sign denotes the logical 'AND', while the forward arrow → indicates 'is sufficient for'. The tilde sign '~' denotes the absence of a factor:

Hypothesis 2: ~SP * ~TP * ~OP → ~W

5.2.2 Interpretation 2: Policy powerfulness interplays with policy meaningfulness

The second interpretation takes into account that frontline workers often feel a desire to benefit others with their work (Dias & Maynard-Moody, 2007). They seek to help clients achieve long-term success and analyze the perceived added value of a policy for society. Meaningfulness refers to workers' perceptions of the contribution a policy makes to a greater purpose, such as societal goals (societal meaningfulness), and the added value of the policy for own clients (client meaningfulness) (Tummers *et al.*, 2009). For instance, client meaningfulness is high when a teacher believes that the policy helps her students to improve their learning outcomes. Meaningful work is of critical importance for frontline workers (May *et al.*, 2004; Grant & Berry, 2011) and numerous studies have found a strong and positive correlation between meaningfulness and implementation willingness (Tummers, 2011; Loyens, 2015; Van Engen *et al.*, 2016; Van der Voet *et al.*, 2017).

The bottom-up view acknowledges that policy changes arise from the interaction of policy and setting, and should be consonant with the values of implementing agents (Matland, 1995). If frontline workers experience discretion, they can tailor the policy to the specific situation of the clients, thereby increasing their perception of its meaningfulness. The implementing actors' perceptions, in turn, can be decisive for implementation outcomes. In summary, powerfulness adds to meaningfulness, which in turn fosters implementation willingness (Lipsky, 1980; Matland, 1995; Tummers & Bekkers, 2014).

Hence, our third hypothesis expects that frontline workers who both feel powerful and perceive the policy as meaningful are willing to implement the

policy. This hypothesis does not rule out that high implementation willingness can also result from other factors. As Figure 5.1 illustrates, it simply assumes that frontline workers who both feel powerful and find the policy meaningful are a subset of the frontline workers who are willing to implement the policy. *The combination of policy powerfulness (strategic, tactical, or operational) with policy meaningfulness (societal meaningfulness: SM, or client meaningfulness: CM) is a quasi-sufficient condition for high implementation willingness*:

Hypothesis 3: $(SP + TP + OP) * (SM + CM) \rightarrow W$

It should be noted that these two interpretations are compatible: Powerfulness can be quasi-necessary for implementation willingness (hypothesis 1), and in combination with meaningfulness, quasi-sufficient (hypothesis 3). However, they are not identical: the first interpretation thinks of powerfulness as a prerequisite for implementation willingness (necessity), while the second one assumes that powerfulness in situations of meaningfulness typically results in high willingness to implement (sufficiency). They also represent two different variants of the bottom-up view on discretion-as-perceived. The first interpretation hypothesizes an enabling, but not automatically triggering role of powerfulness for frontline workers' willingness to implement. The second interpretation highlights the decisiveness of implementing actors' perceived meaningfulness of policies, and assumes that the degree of policy meaningfulness interacts with policy power-fulness to trigger implementation willingness. We may find that powerfulness enables, but does not always result in implementation willingness (interpreta-tion 1 supported), while its combination with meaningfulness is not decisive for implementation willingness (interpretation 2 rejected) – or vice versa. Finally, we do not rule out that other factors than powerfulness and meaningfulness influ-ence implementation willingness. Indeed, bottom-up perspectives highlight various factors that can impact policy implementation. Furthermore, the effects of motivating factors can differ between individuals and situations. Our goal is to clarify the motivating role of powerfulness for, rather than comprehensively explain, implementation willingness. In addition, we identify the empirical rel-evance of powerfulness and meaningfulness for explaining implementation willingness (Sachau, 2007).

5.3 METHOD

Above, we have theorized the role of powerfulness for implementation willingness as an asymmetric and non-linear effect. While a variety of techniques can detect non-linear effects (e.g., polynomials; see also Matzler & Renzl, 2007), we use large-N set-theoretic configurational analysis (Ragin, 1987, 2000; Schneider & Wagemann, 2012). We chose this method as it is the only available technique that models three central theoretical features of our framework (software: R packages QCA and SetMethods; Medzihorsky *et al.*, 2017; Dusa, 2018). First, set-theoretic configurational comparative methods are designed to assess subset relations like the ones hypothesized in Figure 5.1 in terms of necessity and sufficiency. Accordingly, high implementation willingness can have different causes than low implementation willingness. Second, they also provide the possibility of equifinality, meaning that various scenarios can result in high or low implementation willingness: many (but not all) roads lead to Rome. This allows for motivations to differ between individuals. Third, conjunctural explanations are possible, capturing that case-specific factors affect implementation willingness in combination rather than in isolation (Schneider & Wagemann, 2012). We need this possibility to test our third hypothesis. Configurational set-theoretic techniques can be applied to a large-N setting (Fiss, 2011). For theory-testing research designs like ours, large case numbers provide for a more robust test of the theory than small samples (Greckhamer *et al.*, 2013; Thomann & Maggetti, 2017).

Given that this method is not widely used in public administration, we shortly explain its rationale (for detailed descriptions, see Fiss, 2011; Schneider & Wagemann, 2012; Thomann & Maggetti, 2017). The set-theoretic method applied focuses on configurations of variables as sets in which cases have membership or not. The attribution of cases to sets is called calibration. Fuzzy sets allow us to account for differing degrees to which frontline workers' perceptions are present. Qualitative anchors determine the stage at which the outcome or condition is deemed fully present (fuzzy value 1), fully absent (fuzzy value 0) and an indifference (or crossover) point at 0.50. Contrary to usual measurement scales, the crossover point establishes the difference in kind. For example, fuzzy values in the set 'high implementation willingness' above 0.50 mean that implementation

willingness is quite high (W), while values below 0.50 indicate that implementation willingness is quite low (~W).

We can think of necessary and sufficient conditions as subset relations. For instance, our first hypothesis states that frontline workers with high implementation willingness are a subset of those frontline workers who feel powerful. Large-N applications integrate probabilistic elements to capture the degree to which a majority of cases correspond to the statement that X is a superset of Y (quasi-necessity; X≥Y), or a subset of Y (quasi-sufficiency; X≤Y) (Ragin, 2000). The analysis of necessity starts with identifying simple conditions that are a superset of (that is: necessary for) the outcome (here: high implementation willingness). If no simple condition proves necessary, further simple conditions can be added disjunctively until necessity is obtained (Thiem, 2014). We interpret those supersets as necessary conditions that make theoretical sense against the background of our hypotheses, and meet the criteria outlined below (cf. Schneider & Wagemann, 2012).

For the analysis of sufficiency, a 'truth table' is constructed. The rows of the truth table indicate all possible combinations. This enables us to attribute the cases accordingly to the truth table and identify empirically unobserved configurations (so-called logical remainders). If all or enough cases' fuzzy set membership in a truth table row is smaller than or equal to its membership in the outcome, then the row is identified as a sufficient configuration for the outcome. For example, if those frontline workers who partly or fully feel strategically, tactically and operationally powerful and think the policy makes sense for clients and for society are also rather or fully willing to implement the policy, then this configuration of attitudes is sufficient for high implementation willingness. The logical minimization process then identifies the shortest possible expression depicting the configurations that imply the outcome - the solution term. This is a straightforward procedure that relies on basic set theory: for example, A*B*C + A*B*~C can be reduced to A*B (Thomann *et al.*, 2018a).

To evaluate our results, we use consistency and coverage measures. The values of these fit indices can range from 0 (low) to 1 (high). Consistency is the extent to which the results are in line with the statements of necessity or sufficiency. For sufficient conditions, consistency is indicated for single truth table rows (raw consistency), for single configurations of, or for the whole solution term.

Table 5.1 Strategies to address errors and evaluate model

	Issue	Definition	Strategy	Application
Possible error sources	*Deviant case & measurement errors*	Errors related to sensitivity to one or more flawed cases	*Frequency thresholds robustness test*	Use of three different frequency thresholds; configurations without a certain frequency are treated as logical remainders
		Sensitivity to changes in raw consistency levels	*Raw consistency robustness test*	Use of three different raw consistency thresholds (criterion: PRI)
	Plausibility & tenability	Limited diversity & contradictions can trigger inferences that are implausible and/or contradictory	*Enhanced Standard Analysis*	Intermediate solution, based on directional expectations and exclusion of contradictory rows and untenable assumptions
	Accuracy	Degree to which observations correspond to set relation	*Consistency*	Necessity: ≥0.90 Sufficiency: ≥0.75
Criteria for model evaluation		Simultaneous subset relations: degree to which the same condition is not simultaneously sufficient for the negated outcome	*Proportional Reduction in Inconsistency (PRI)*	No fixed threshold
	Explanatory power	Empirical relevance of model	*Coverage & Relevance of Necessity*	Necessity: ≥0.60 RoN ≥0.60 (direct calibration) / 0.55 (recoding method) Sufficiency: verbal interpretation Low coverage indicates low explanatory power
	Random errors	Errors that are unpredictable and inconsistent in their magnitude or direction (e.g., because of estimation and personal factors in surveys)	*Probabilistic criteria*	Right-handed Z-Test for proportion of cases with X≥X (necessity), X≤ Y (sufficiency) 0.8: 'almost always'
	Limited empirical diversity	Presence of logical remainders, i.e. truth table rows without enough cases with membership > 0.5	*Limited diversity index* % remainders / logically possible configurations	Models with less limited diversity have a stronger empirical basis
	Ambiguity	Patterns in data are unclear: several equally non-redundant solutions can be derived	*Ambiguity index* (Nr. of equally plausible models)	Unambiguous models are preferred (row dominance applied)

Table 5.1 Strategies to address errors and evaluate model (continued)

	Issue	Definition	Strategy	Application
Criteria for model evaluation	*Robustness*	Terms of enhanced parsimonious solution remain robust across different models that pass consistency threshold 0.75	*Robustness index* Average % of models in which (a subset of) a term appears	More robust models are preferred
	Skewness	Skewed distributions can produce simultaneous subset relations, exacerbate limited diversity, and strongly distort parameters of fit	*Skewness statistics*	% of cases with membership >0.50 in sets is reported Skewness is problematic if the vast majority (>85%) of the cases cluster *in only one* of the four possible intersecting areas of the XY plots with two diagonals

Furthermore, the proportional reduction in inconsistency (PRI) indicates the degree to which a given configuration is not simultaneously sufficient for both the occurrence and the non-occurrence of the outcome. Coverage sufficiency depicts how well the model explains the available empirical information. Raw coverage expresses how much a single configuration covers, and unique coverage indicates how much it uniquely covers. Low coverage means that the model has a limited capacity to explain the outcome. For necessary conditions, coverage expresses their relevance in terms of the condition set not being much larger than the outcome set, and the relevance of necessity (RoN) in terms of the condition being close to a constant (all formulae in Schneider & Wagemann, 2012).

Error management is a salient issue for large-N applications of set-theoretic configurational comparative methods (Maggetti & Levi-Faur, 2013; Thomann & Maggetti, 2017). In the absence of established guidelines, we propose state-of-the-art strategies that complement the traditional parameters of fit to address possible error sources, as shown in Table 5.1. To account for different possible model specifications and to assess robustness, we calculated 54 models, using two calibration techniques (see below) and three different raw consistency and frequency thresholds. The models presented in the paper rank best on eight criteria for model evaluation, see Table 5.1. The rationale underlying the choice of different analytic thresholds and the "best" models for interpretation is outlined in detail in box 5.1.

We assess hypothesis 1 on necessary conditions in Figure 5.2. To assess our hypotheses on sufficient conditions (hypotheses 2 and 3), we apply Ragin's (1987) principles of formal set-theoretic theory evaluation, as extended by Schneider and Wagemann (2012) to account for consistency and coverage. This procedure identifies the proportion of cases that confirm, refute or extend our theoretical expectations. To this end, the scenarios expected (T) and those not expected (~T) in the hypotheses were intersected with the scenarios that were empirically (not) observed (S and ~S). This technique helps us answer three questions. First, which parts of the theory are supported by the findings (T*S and ~T*~S)? Second, in which direction should theory be expanded (~T*S)? Third, which parts of the theory need to be dropped (T*~S)? Table 5.2 summarizes the main analytic steps (on p. 96).

Table 5.2 Main steps of the large-N set-theoretic configurational analysis

Step 1	Analysis of necessity (H1)	Identify the supersets of high implementation willingness for both datasets, using two calibration strategies
Step 2	Analysis of sufficiency	Identify subsets of low and high implementation willingness, using both datasets, two calibration strategies, three different raw consistency thresholds and three different frequency thresholds
Step 3	Model evaluation, analysis of sufficiency	Identify best-performing model for each outcome, dataset and calibration strategy (for criteria, see Table 5.1)
Step 4	Model selection, sufficient conditions	Identify the models with highest explanatory power per dataset and outcome for interpretation
Step 5	Formal set-theoretic theory evaluation (H2 and H3)	Identify how results behave with respect to the hypotheses: which (parts of) the hypotheses are supported, which ones are refuted?

The data, truth tables, directional expectations, conservative and parsimonious solutions, simplifying assumptions, skewness tests, *R* codes for replication, and the results not reported in this are all provided as online supplementary material.[2]

5.3.1 Data

We used two data samples collected in the Netherlands in two sectors (healthcare and education) at two times (2010 and 2013). By analyzing these two datasets, we

---

2   The online Appendix and replication materials are published at dataverse, see http://dx.doi.org/10.7910/DVN/G9PYIV .

both evaluate whether our hypotheses hold for frontline workers implementing a specific policy (dataset 1), and whether the hypothesized relations hold in another policy sector and from a more general perspective (dataset 2). This allows us to adopt a comparative approach and provides a stronger empirical basis to either accept or reject the hypotheses. Still, in examining two case studies, the possibility to make general claims remains limited. This is acknowledged and will be discussed in the discussion section.

Box 5.1 Procedure for model evaluation and selection, analysis of sufficiency

Setting raw consistency thresholds is decisive for determining which conditions are sufficient. Since consistency values strongly depend on the specific dataset, truth table and case distributions, there are no fixed anchors for setting these thresholds (Schneider & Wagemann, 2012; Thomann & Maggetti, 2017). Accordingly, using standardized thresholds is widely considered bad practice (Wagemann *et al.*, 2016). Therefore, we use a context-sensitive strategy that integrates PRI values for determining raw consistency thresholds. Considering the range of PRI values in a truth table, a context-specific critical PRI value was determined. This procedure ensures that raw consistency is set such that simultaneous subset relations – when the same configuration is considered sufficient for both low and high implementation willingness – are avoided (Schneider & Wagemann, 2012). The first raw consistency threshold was set above the first row with a PRI below this critical value; the second threshold was set above the second row with a PRI below that value; and the third threshold, above the third respective row. Hence, the same principle was applied to each analysis, but considering the specificities of the respective truth table.

Tables B2-B7 in the online Appendix report all resulting models and illustrate their robustness. The 'best' models for each dataset, calibration strategy and outcome (high and low implementation willingness) were then identified according to their performance regarding consistency, PRI, coverage, statistical significance, limited diversity, ambiguity, robustness and skewness. These criteria comprehensively capture the main challenges to validity with set-theoretic techniques (Thomann & Maggetti, 2017; Table 5.1). The best model is the one whose average rank on each of these indicators is the highest amongst those models with a minimum consistency of 0.75. Below this threshold, QCA solutions are usually not considered sufficient (Schneider & Wagemann, 2012). The ranking procedure is self-explanatory for consistency, PRI, coverage, Z values and robustness. Additionally, high levels of limited diversity and model ambiguity were punished, by rewarding the lowest levels a ranking of 1; the highest level is attributed the lowest possible rank (e.g., 7 if 7 models pass the consistency threshold); then the second highest level is attributed the second worst rank, and so on. The motivation for this was that limited diversity poses serious threats to inferences with truth table analyses (Thomann & Maggetti, 2017) and model ambiguities indicate that the results are inconclusive (Baumgartner & Thiem, 2017).

This left us with six sufficient models, among which the ones with the highest explanatory power (coverage) were preferred for each outcome and dataset, reported in Table 5.4 and chosen for interpretation.[3] This procedure minimizes the weakness of many large-N set-theoretic configurational analyses, which often suffer from very limited coverage (Wagemann *et al.*, 2016).

---

3   No analysis of sufficiency was possible for dataset 2 using the recoding method.

The 2010 study ('study 1') investigated whether Dutch mental healthcare workers felt alienated from one specific government policy program, namely, the Diagnosis-Related Group (DRG) policy, and their willingness to implement this new policy. The DRG policy was developed by the Dutch government as a means to determine the level of financial reward for mental healthcare provision by stipulating a standard rate for each disorder. The sampling frame consisted of 5.199 professionals who were members of two nationwide mental healthcare associations (see Tummers *et al.*, 2012). Using an e-mail and two reminders, 1.317 returns of the questionnaire were received (25% response). The gender composition of the respondents was 66% female. This is consistent with the Dutch average (69%) for mental healthcare professionals. The average age was slightly higher than that of the mental healthcare professional population (48 versus 44). Common reasons for not participating were a lack of time, retirement, change of occupation, or not working with the DRG policy.

Dataset 2

The 2013 study ('study 2') investigated whether Dutch teachers felt alienated from government education policies in general, and the relationship with their general willingness to implement government policies. The sampling frame consisted of a nation-wide sample of 2.863 teachers working in secondary education, selected through the pension fund for all Dutch employees in government and education (ABP) (Van Engen *et al.*, 2016). Using an e-mail and one reminder, 1.096 returns of the questionnaire were received (38% response). On average the respondents were 51 years old, and 59 percent were male. Dutch national statistics on secondary school teachers in 2013 have shown that the average age is 46, and 48 percent are male. In our sample males were therefore somewhat overrepresented, and the respondents were on average slightly older than the national average. To rule out a non-response bias, we asked the organization managing the sampling frame to analyze whether or not the respondents problematically differed from non-respondents in terms of variables such as age, gender, and occupation. For instance, the results indicated there were no significant differences between the two groups in terms of occupation (respondents with managing responsibilities: 8%; non-respondents: 9%). They also indicated that the arguments non-respon-

dents gave for not participating usually were 'no time', 'forgot the questionnaire' and 'did not open e-mail during response period'. Nevertheless, it is important to highlight that although we argue that our data is fairly representative, it is still possible that some type of response bias could have influenced our results.

5.3.2 Measures

The measures of implementation willingness, powerfulness and meaningfulness were formatted using five-point Likert scales. All measures had adequate Cronbach alphas (ranging between 0.78 and 0.97).

In dataset 1 we measured policy powerfulness (strategic, tactical and operational powerfulness: six indicators) and policy meaningfulness (societal: twelve indicators, client: four indicators) for a specific policy using the policy alienation measurement scales of Tummers (2012). In dataset 2 we measured general policy powerfulness (strategic, tactical and operational powerfulness: six indicators) and general policy meaningfulness (societal and client: four indicators) using the general policy alienation measurement scales of Van Engen *et al.* (2016). Implementation willingness was measured using five indicators corresponding to the validated scale by Metselaar (1997). If necessary, we inverted the positive and negative end of the respective scales, so that high scores always indicate high powerfulness, meaningfulness, and implementation willingness.

5.3.3 Calibration

Indicator variables were calibrated into indicator sets. Set membership requires a statement about a qualitative state: cases are either (more or less) in a set or (more or less) out of a set. The answer categories of Likert scales have a fixed qualitative meaning, which can be directly translated into set membership scores. For example, if a frontline worker answers 'disagree' (score of 2 on 1-5 scale) to the question 'In my organization, professionals could take part in conversations regarding the execution of the policy', then this means that on this item the case 'tactical powerfulness' is rather absent, but not totally absent.

The neutral answer (score of 3) poses a conceptual challenge for calibrating set membership (Wagemann *et al.*, 2016). In box 5.2 we discuss in detail the nature of this challenge and how we address it.

Box 5.2 Procedure to test for different calibration strategies

The neutral answer (score of 3) poses a conceptual challenge for calibrating set membership. Neutral answers could indicate that a frontline worker experiences neither the presence nor the absence of, say, tactical powerfulness (point of indifference). However, cases with a set membership score of 0.50 cannot be attributed to truth table rows, which results in excessive dropout rates and should therefore be avoided (Wagemann *et al.*, 2016). While Likert scales are typically acknowledged to represent ordinal rather than interval-level data (Wirth & Edwards, 2007), the status of neutral answers in the scale and hence also in the set can be disputed. One possible interpretation is that the answer 'neither agree nor disagree' indicates less agreement than 'rather agree', but more agreement than 'rather disagree' – we can treat the answers as scale. However, another possible interpretation is that 'neither agree nor disagree' indicates both 'no agreement' as well as 'no disagreement' – in other words, no presence, of, say, powerfulness at all. Hence, these cases would in fact be 'fully out' of the set of, for example, tactical powerfulness. Different calibration techniques can substantially affect the results of set-theoretic configurational analyses (Skaaning, 2011). To identify the best calibration strategy, we tested for two different commonly used calibration techniques for Likert scales. First, the direct method of calibration uses a logistic function to fit the raw data in-between the three qualitative set membership anchors (Schneider & Wagemann, 2012). This method is very popular in large-N set-theoretic configurational analyses. Typically, the crossover point is set right above the indifferent answers, resulting in set memberships extremely close to 0.50 that can hardly be interpreted in conceptual terms. As Wagemann *et al.* (2016, p. 55) point out: "This is arbitrary and should not become common practice. (...) [it] does not have much to do with a decision about set membership". To avoid this pitfall, we interpret neutral answers as 'fully out' of the set (the cases remain in the sample, but they have a set membership of 0). Answers of 4 (agree) and 5 (fully agree) were recoded into 3 and 4 before calibration. Second, we alternatively treated the answers strictly as a scale using simple recoding technique, which involves the grouping of cases into previously defined set-membership scores (Schneider & Wagemann, 2012). Here, we followed the proposal by Emmenegger *et al.* (2014) (and slightly adapted it to account for degrees of non-membership) and used the calibration anchors shown schematically in Table 5.3.

Our results indicate that in the analysis of sufficiency, the recoding method works better for dataset 1 (the models perform better and explain more cases), while for dataset 2, the direct strategy is more feasible – recoding method leads to distorted parameters of fit that prevent a meaningful analysis of sufficiency. Importantly, however, both calibration strategies attribute indifferent answers as more out than in the set, resulting in the same conceptual meaning and attribution of cases to truth table rows. The differences in the results are thus exclusively due to changes in the parameters of fit. The results of necessity are robust regardless of the calibration strategy. Using the direct strategy for dataset 1 for sufficient conditions leads to the same overall conclusions regarding our hypotheses as with the indirect strategy. For these reasons, we adopted the recoding method for dataset 1 and the direct calibration method for dataset 2 for the results interpreted below.

In short, we conceive of indifferent values as more out than in of the set. To identify the best calibration strategy, we tested for two different commonly used calibration techniques for Likert scales. First, the direct method of calibration uses a logistic function to fit the raw data in-between the three qualitative set membership anchors (Schneider & Wagemann, 2012). Using our data, this commonly applied technique results in set membership scores of 0.05, 0.27, 0.73

and 0.95; indifferent answers were coded as 'fully out'. Second, we alternatively treated the answers strictly as a scale using a simple recoding technique. This technique involves the grouping of cases into previously defined set-membership scores (Schneider & Wagemann, 2012; Emmenegger *et al.*, 2014), see Table 5.3 for an example. Based on assessment of their performance, we adopted the recoding method for dataset 1 and the direct calibration method for dataset 2 for the results interpreted below. Both strategies attribute the same values on the Likert scale as more in/more out of the set, resulting in the same conceptual meaning, but different parameters of fit.

Table 5.3 From Likert scale to indicator sets: an example of recoding method

Likert score	Indicator fuzzy set score
Survey question: 'I intend to put effort into achieving the goals of the DRG policy'	*Set: 'High implementation willingness, indicator 2'*
Completely agree (5)	Highly willing (1)
Agree (4)	Mostly but not highly willing (0.8)
Neutral (3)	Rather unwilling (0.4)
Disagree (2)	Mostly but not fully unwilling (0.2)
Completely disagree (1)	Fully unwilling (0)

Missing values make it impossible to attribute cases to truth table configurations. This is a potential issue since a high share of cases has missing values on at least one indicator set in dataset 1. This is due to the fact that we gave the possibility to indicate 'don't know' for each item in dataset 1 and doing this on one out 39 items already indicates a missing value (60% in dataset 1, 7.8% in dataset 2). Excluding these cases from the analysis would result in a biased sample.

The aggregation strategy will impact the analysis. It needs to avoid such excessive dropout, while ensuring construct validity and avoiding overly skewed condition and outcome sets. The first out of three aggregation options would be building averages across the indicators. Doing so for raw values would negatively affect construct validity: the inclusion of neutral answers (score 3) leads to average values that are difficult to interpret especially since they are numerous. Calculating averages of calibrated sets is equally problematic because it can result in set memberships of 0.50, producing dropouts during truth table analysis. The second and third options are set-theoretic. Using the logical 'AND' as aggregation

strategy (minimum rule) represents a very restrictive conceptualization, as all indicators need to be present simultaneously for an attitude to be present. This results in the excessive dropouts. Moreover, it would produce highly skewed sets that make it impossible to proceed with the analysis of the outcome (Schneider & Wagemann, 2012). For example, in dataset 1, none of the aggregated sets would have more than 5% cases with membership above 0.50.

Accordingly, as the third and in our view superior option, we use the logical 'OR' to aggregate the indicators into the five condition sets. This aggregation strategy conceives of different indicators as functional equivalents that indicate the presence of an attitude (Goetz & Starr, 2003). For instance, it suffices for a frontline worker to have a score on one of the five indicator sets for implementation willingness ('W') to obtain a value for 'W' (maximum rule). This 'optimistic' measure lowers the dropout problem (final N for dataset 1=1.004, dropout 23.8%; for dataset 2=1.087, dropout 0.8%) and produces acceptable levels of skewness that enable an analysis of the outcome. This has consequences in terms of concept validity: the positive memberships in sets represent a wider range of functionally equivalent attitudes, which are assumed to represent the concept. This conceptualization does justice to the wide range of experiences facing frontline workers on the ground.

5.4 RESULTS

We can now test the hypotheses. Table B1 in the online Appendix displays descriptive statistics. They show that overall, the Dutch teachers (study 2) have a more positive attitude than the healthcare workers (study 1). They feel more powerful, perceive the policies as more meaningful, and have higher implementation willingness.

Regarding hypothesis 1, we indeed found that feelings of powerfulness are almost always necessary for high implementation willingness. This holds for both datasets and regardless of the calibration strategy used (see Table A1, online Appendix). This is shown in Figure 5.2. In the Dutch education sector, either strategic, tactical or operational powerfulness is needed for high implementation willingness. Among Dutch healthcare workers, the finding is even stronger:

it is enough for high implementation willingness to either feel powerful at the strategic or operational level, or alternatively, to feel powerful at the operational or tactical level. These results provide strong support for the hypothesis that powerfulness at different levels is a prerequisite for implementation willingness.
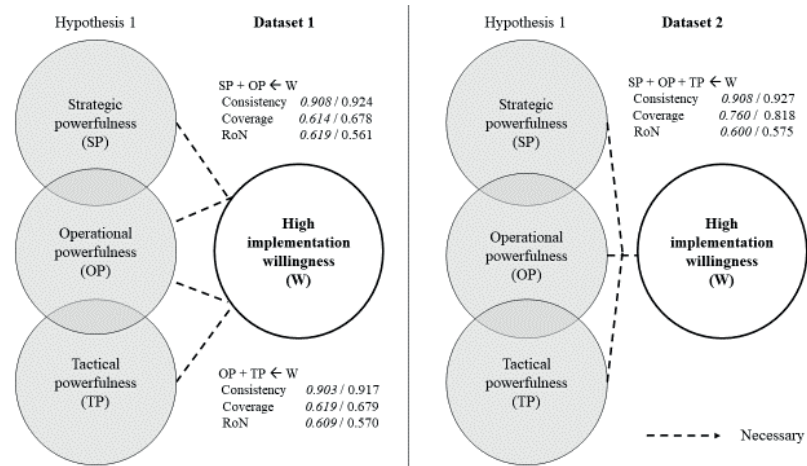

Figure 5.2 Evaluation hypothesis 1

Hypothesis 2 captured a potential consequence of the first hypothesis, namely, that a lack of powerfulness might be quasi-sufficient for low implementation willingness. Table 5.4 reveals three configurations in dataset 1, and five configurations in dataset 2, that are almost always sufficient for low implementation willingness. The Dutch health workers who are unwilling to implement the DRG policy consistently experience low levels of powerfulness and, in path 3, meaningfulness. Conversely, in the education sector, the picture is less clear at first sight: these configurations entail a mix of both positive and negative attitudes. The parameters of fit score well in dataset 1, while in dataset 2, the results are highly consistent, but have a fairly low empirical relevance (coverage).

We indicate the percentage of all cases that display these attitudes with different levels of implementation willingness, and what that means for interpreting the results. For example, in the upper left quadrant, those frontline workers that display these attitudes and have low implementation willingness support the

Table 5.4 Sufficient conditions for implementation willingness (intermediate solution), best-performing models

| | Hypothesis 2: Low implementation willingness (~W) | | | | | | | | Hypothesis 3: High implementation willingness (W) | | | | | | |
| | Dataset 1 Recoding method | | | Dataset 2 Direct calibration | | | | | Dataset 1 Recoding method | | | | Dataset 2 Direct calibration | | |
Configuration	1	2	3	1	2	3	4	5	1	2	3	4	1	2	3
High strategic powerfulness (SP)	○				○	○	○	●		●		●		●	
High tactical powerfulness (TP)		○			○	○	○	○		●	○	●			●
High operational powerfulness (OP)	○	○	○	○		○	●	○	●		●	○	●		
High societal meaningfulness (SM)			○	○	○		●	●			●	●	●	●	●
High client meaningfulness (CM)	●	○	○	●	●	●	○	○	●	●			●	●	●
Consistency	0.865	0.855	0.858	0.872	0.873	0.883	0.883	0.881	0.909	0.935	0.942	0.942	0.939	0.936	0.938
Raw coverage	0.525	0.531	0.573	0.487	0.475	0.419	0.426	0.442	0.488	0.451	0.350	0.313	0.409	0.396	0.415
Unique coverage	0.010	0.017	0.037	0.012	0.013	0.001	0.006	0.012	0.046	0.008	0.022	0.013	0.005	0.006	0.007
Solution consistency	0.818*			0.826**					0.890***				0.935***		
Solution PRI	0.627			0.360					0.727				0.841		
Solution coverage	0.640			0.567					0.565				0.436		

Explanation: This table shows the combinations of conditions (paths) that were quasi-sufficient for low implementation willingness (left-hand side) and high implementation willingness (right-hand side). Black circles indicate the presence of a condition, and white circles its absence. Blank spaces indicate the irrelevance of a condition. The different paths of one model are interchangeable. For example, in dataset one, low strategic powerfulness with low operational powerfulness typically led to low implementation willingness (path 1); alternatively, low operational powerfulness combined with feelings of client and societal meaninglessness also typically did (path 3). Table 1 explains the parameters of fit. For raw consistency thresholds, frequency thresholds and simplifying assumptions applied, see Tables B10, B11, B12, B16, B17, B18 and B19, online Appendix.

*Significant at 0.1 level, **significant at 0.05 level, ***significant at 0.01 level ('almost always sufficient'). All single paths are significantly sufficient at the 0.01 level.

hypothesis; those that have high implementation willingness are 'contradictions', that is, they separate the quasi-sufficient results from perfect sufficiency.

Using set-theoretic theory evaluation to assess hypothesis 2 formally, we find robust support that the absence of tactical, strategic and operational powerfulness implies low implementation willingness in the healthcare sector. This is shown in Table 5.5. However, quite some cases remain unexplained (lower right quadrant). In addition and compatible to what we hypothesized, the absence of operational, but not also tactical and strategic powerfulness in some situations also leads to low implementation willingness (lower left quadrant). Conversely, in the education sector, overall the empirical support for the second hypothesis is so weak that we must reject it. The contradictory cases are empirically more frequent than those instances that directly support the hypothesis (left-hand side of Table 5.5). Here, the solution term only explains a tiny fraction of the observed patterns of low implementation willingness.

Overall, the conclusion for hypothesis 2 is ambiguous. While powerfulness is a quasi-necessary condition for high willingness, the 'flipside' of this argument materializes in the healthcare, but not in the educational sector. While seemingly puzzling, this finding illustrates that the things that motivate people at the workplace can be different from those that demotivate them (see also Schneider & Wagemann, 2012).

Table 5.5 Evaluation of hypothesis 2

		Empirics	
		Detected in solution	**Not detected in solution**
Theory	**Hypothesized**	~SP*~TP*~OP + SP*~TP*~OP *~SM*~CM *~SP*~TP*~OP *~SM*CM +* *~SP*~TP*~OP*CM* 14.1 % / *0.1 %* (~W): support theory 4.8 % / *0.6 %* (W): contradict theory & solution	Empty set *~SP*~TP*~OP *~CM + **SP*~TP*~OP** *~SM*~CM* 0 % / *4.5%* (~W): support theory 0 % / *6.6 %* (W): delimit theory
	Not hypothesized	~OP*(SP*~TP + SP*~SM*~CM + ~SP*TP + TP*~SM*~CM) *~SP*OP*~SM*CM +* *~SP*~TP*OP*SM*~CM* *+ SP*~OP*~SM*~CM +* *SP*~TP*~OP*SM*~CM* *+ TP*~OP*~SM*CM +* *~SP*TP*~SM*CM* 14 % / *1.4 %* (~W): extend theory 8 % / *2.9 %* (W): empirical contradictions	OP + SP*TP*OP*CM + SP*TP*OP*SM + SP*OP + SP*TP*CM + SP*TP*SM + TP*OP *OP*~SM*~CM + TP*OP*~CM +* *OP*SM*CM + SP*OP + SP*OP*SM*CM* *+ TP*OP*SM + SP*~SM*~CM +* *SP*TP*~CM + SP*SM*CM + SP*TP*SM* *+ ~SP*TP*~OP*~CM + TP*~SM*~CM +* *TP*~CM + TP*OP*SM*CM + SP*TP*OP +* *SP*TP*SM*CM + TP*SM* 28.5 % / *21.6 %* (~W): point to overlooked explanations 30.6 % / *62.3 %* (W): support theory
Supports theory		Extends theory	Delimits theory

Based on Schneider and Wagemann (2012, p. 301).
Bold: hypothesized combinations. No italics: dataset 1 (recoding method), *italics*: dataset 2 (direct calibration). *Hypothesis 2*: **~SP*~TP*~OP → ~W**.
Explanation: This table shows how the results behave with respect to hypothesis 2. The upper left quadrant shows those attitudes that were both hypothesized and observed with a set membership > 0.5. The lower left quadrant displays those attitudes that were not expected, but observed empirically, revealing additional explanations for low implementation willingness. The upper right quadrant refers to attitudes that were expected but not observed in the solution. The lower right quadrant displays those attitudes that are neither hypothesized nor covered by the solution.

We indicate the percentage of all cases that display these attitudes with different levels of implementation willingness, and what that means for interpreting the results. For example, in the upper left quadrant, those frontline workers that display these attitudes and have high implementation willingness support the hypothesis; those that have low implementation willingness are 'contradictions', that is, they separate the quasi-sufficient results from perfect sufficiency.

Hypothesis 3 states that the combination of policy powerfulness (strategic, tactical, or operational) and policy meaningfulness (societal or client meaningfulness) is a quasi-sufficient condition for high implementation willingness. Table 5.4 indeed suggests that the combination of high powerfulness and mean-

ingfulness relate to high implementation willingness. Four configurations are very often sufficient for high implementation willingness in the Dutch healthcare sector, and three are almost always sufficient configurations in the education sector. For example, Dutch healthcare workers who feel powerful at the strategic and tactical level and to whom the DRG policy makes sense for the patients typically make efforts to implement the policy. Both models have a good consistency, while its explanatory power (coverage) is quite low in the education sector. The left-hand side and lower right quadrant of Table 5.6 lend full support to our third hypothesis. Powerfulness, in one of its three variants, combined with meaningfulness almost always results in high implementation willingness. This support is empirically stronger in study 2 (education) than in study 1 (healthcare).

However, findings also restrict the hypothesis to certain circumstances. For instance, the upper left quadrant of Table 5.6 shows that in the healthcare sector, the positive motivational role of tactical powerfulness together with meaningfulness often unfolds even in the absence of either strategic or operational powerfulness. In the education sector, regardless of the type of powerfulness typically both societal and client meaningfulness must be present. Conversely, the instances in which hypothesis 3 is rejected both datasets are negligibly rare (upper right quadrant).

In summary, both bottom-up interpretations (hypothesis 1 and 3) of how perceived discretion motivates frontline workers are indeed reflected in our data. Hypothesis 2 is supported for the first dataset (healthcare) but rejected for the second (education). However, for the second interpretation there is also room for improvement, as quite some cases are not explained (23.4% in dataset 1 and 41.7% in dataset 2 point to overlooked explanations). This is not particularly high, as we aimed to explain willingness with just a few indicators and the unexplained variance is quite low. In field studies in social sciences, we should not expect a perfect theory explaining everything. It suggests that powerfulness combined with meaningfulness is only one of several factors that explain frontline workers' high implementation willingness.

Table 5.6 Evaluation of hypothesis 3

	Empirics	
	Detected in solution	Not detected in solution
Theory — Hypothesized	SP*SM*(TP*CM + TP*~OP) + SP*CM*(OP + TP + TP*~OP*SM) + TP*CM*(OP*SM + OP) + OP*SM*(CM + SP*TP*CM + ~SP*TP) + OP*CM + OP*CM*(SP*TP + ~SP*TP*SM + SP* SM) *OP*SM*CM + SP*OP*SM*CM + TP*OP*SM*CM + SP*SM*CM + SP*TP*SM*CM + TP*SM*CM* 9.7 % / *15.7 %* (W): support theory 4.4 %/ *1.2 %* (~W): contradict theory & solution	SP*SM*(OP*~CM + + ~TP*~CM + ~TP*~OP + TP*OP*~CM) + SP*~TP*~OP*CM + ~TP*OP*SM*~CM + ~SP*TP*~OP*SM + ~SP*TP*~OP*CM *OP*SM*~CM + OP*~SM*CM + SP*SM*~CM + SP*~SM*CM + TP*SM*~CM + TP*~SM*CM* 10.3% / *15 %* (W): support theory 6.3 % / *4.3 %* (~W): delimit theory
Theory — Not hypothesized	Empty set *Empty set*	~SP*~TP*~OP*~CM + ~SP*~TP*~OP + ~SP*~TP*~OP*~SM + ~SM*~CM + ~SP*~OP*~SM*~CM + ~TP*~OP*~SM*~CM *~SP*~TP*~OP*~CM + ~SP*~TP*~OP + ~SP*~TP*~OP*~SM + ~SM*~CM + ~SP*~TP*~OP*~SM*~CM* 23.4 % / *41.7 %* (W): point to overlooked explanations 46 % / *22.1 %* (~W): support theory
Supports theory	Extends theory	Delimits theory

Bold: hypothesized combinations. No italics: dataset 1 (recoding method), *italics*: dataset 2 (direct calibration). *Hypothesis 3*: OP*SM + OP*CM + SP*SM + SP*CM + TP*SM + TP*CM → W.
Explanation: This table shows how the results behave with respect to hypothesis 3. The upper left quadrant shows those attitudes that were both hypothesized and observed with a set membership >0.5. The lower left quadrant displays those attitudes that were not expected, but observed empirically. The upper right quadrant refers to attitudes that were expected but not observed in the solution. The lower right quadrant displays those attitudes that are neither hypothesized nor covered by the solution.

5.5 DISCUSSION

The main conclusion of our study is that discretion-as-perceived is a quasi-necessary condition for high implementation willingness. This aligns with Herzberg's motivation theory and suggests an enabling (but not automatically triggering) motivational effect of perceived discretion (Herzberg *et al.*, 1959; Goertz & Starr, 2003). Frontline workers need to feel that they can influence the policy – this is a necessary condition.

Secondly, we have found mixed evidence for the hypothesized more radical 'flipside' of the first interpretation. This result aligns with a classic insight from Herzberg's motivation theory: the things that make people feel satisfied and motivated on the job can be different in kind from the things that make them feel dissatisfied – and this can obviously vary between policy sectors and types of professions (Herzberg *et al.*, 1959; Bassett-Jones *et al.*, 2005; Sachau, 2007).

Thirdly, we also found that - in combination with policy meaningfulness - powerfulness is quasi-sufficient for high implementation willingness. When frontline workers felt that they had both high powerfulness and that the policy was meaningful for society, this strengthened their willingness to implement it (Maynard-Moody & Musheno, 2012; Van der Voet *et al.*, 2017).

Our results encourage scholars to rethink assumptions of implementation theory by moving from a correlational logic to the consideration of asymmetric patterns. By adapting Herzberg *et al.*'s (1959) seminal, fundamentally asymmetric two-factor theory of motivation to the context of frontline implementation, we are able to refine policy implementation theory. The important role of power-fulness could be uncovered by modeling asymmetric effects via a methodology specifically designed to test these (Ragin, 1987, 2000; Schneider & Wagemann, 2012). Our analysis sheds more light on the puzzling results of previous studies, which assumed symmetric, correlational patterns (Tummers, 2011; Van Engen *et al.*, 2016). The strong and robust asymmetric effect of powerfulness that we detected simply escaped the attention of these studies because their designs are unable to detect such asymmetric relationships (Schneider & Wagemann, 2012). This has helped us to identify discretion-as-perceived as a necessary prerequisite for high implementation willingness. Accordingly, implementation theory might fruitfully turn toward more asymmetric and complexity-oriented models of policy in practice (Raab *et al.*, 2015; Misangyi *et al.*, 2017; Thomann *et al.*, 2018a).

A number of caveats apply for this study. First, apart from powerfulness and meaningfulness, additional factors such as caseloads, interactions, and resources influence frontline workers' implementation willingness (e.g., Sabatier, 1986; O'Toole, 2000; May & Winter, 2009). Second, although we analyzed two large-N datasets, we should be careful to generalize these findings to frontline workers in other policy domains or countries. Third, while applying an 'optimistic' measure of our dependent and independent variables helped us reducing drop-out and

countering the skewness of the data, future research should study whether our results also hold applying 'pessimistic' measures, ideally using large datasets in multiple sectors and countries where cases with missing values can be completely deleted from the dataset. Fourth, although there is a fairly strong correlation between intended behavior and actual behavior (Sheeran & Orbell, 1988; Randall & Wolf, 1994; Armitage & Connor, 2001), future studies could measure behavior more directly. Fifth, it should be noted that common method bias could be a problem in our study, since we used the same data source to measure the variables under study (powerfulness, meaningfulness, implementation willingness). It is recommended that future researchers studying the relationship between powerfulness and implementation willingness apply stronger designs and techniques to establish causal inference. We recommend the use of field, lab or survey experiments.

5.6 CONCLUSIONS

Despite the fundamental theoretical debate on the role of discretion and its relevance for policy design and implementation, to date there has been little empirical research to assess the behavioral assumptions underlying this debate. Our study is the first large-N empirical illustration lending robust support to a bottom-up view on discretion as an inevitable and potentially beneficial aspect of frontline implementation. We find that possibilities to participate in and influence public policies are a *prerequisite* for frontline workers to be willing to implement the policy. However, this is not enough. It is not sufficient. Other factors, including perceiving the policy as meaningful for society and clients, are needed to truly increase the willingness to implement of frontline workers.

Our study contributes to clarifying the behavioral underpinnings of the top-down versus bottom-up debate on discretion (Sabatier, 1986; Hupe, 2013; Thomann *et al.*, 2016). The question whether frontline workers should be granted discretion continues to be hotly debated not only in research on policy implementation, but also on policy, regulatory and organizational design (e.g., Howlett 2004; Chun & Rainey, 2005). Our findings lend substantial support to a bottom-up view of street-level bureaucrats as problem-solvers who crucially

need the freedom to adapt the program to local conditions. Conversely, they lend very little support to top-down assertions that high levels of discretion often or predominantly have a negative impact on policy implementation – at least not at the perceived, motivational level.

The link between implementation willingness and actual implementation behavior - which was not analyzed here - will continue to provide fertile grounds for further exploration (see e.g., Brodkin, 1997; Chun & Rainey, 2005; Gofen, 2014). Committed implementers are a crucial factor for successful policy implementation (May & Winter, 2009). Our contribution lies in showing that the overwhelming majority of those frontline workers with high implementation willingness also experience high levels of discretion. This should encourage scholars and practitioners to move beyond the question whether frontline workers should be granted discretion: our answer to this question is yes.

The more salient question seems to be how to make best use of frontline workers' discretion to encourage behavior that eventually contributes to the achievement of policy goals. Discretion appears as a defining contextual feature of street-level bureaucratic work that changes the daily experiences shared by frontline workers. This emphasizes the importance of future research that singles out how a context of more or less discretion affects frontline workers' actual behavior, and under which specific circumstances.

Finally, systematic comparative empirical assessment of street-level bureaucracy theory like ours demonstrate the potential of large-N comparisons over different policy contexts to facilitate theoretical progress in this field (O'Toole, 2000). A micro-level perspective is useful to evaluate the underlying psychology and mechanisms of frontline implementation (Grimmelikhuijsen *et al.*, 2017). It provides valuable information to policymakers and managers engaged in shaping the macro- and meso-level contexts of street-level bureaucracy, in their continuous quest to improve public service delivery.

chapter 6

Conclusions and discussion

6.1 CONCLUSIONS

Teachers, healthcare workers and police, as well as other public employees working at the frontline of public service delivery, are often confronted with new policy programs that, usually, lead to new rules and regulations that have to be implemented. As we explained in the introductory chapter, the fact that these 'frontline workers' are often confronted with new policies is, of course, in itself not problematic - democratically elected governments have the mandate to do so (Dunsire, 1978; Barrett, 2004). However, it can influence the way in which frontline workers perform their tasks, as grown practices may be challenged – repeatedly.

What has been neglected in the literature, so far, is the fact that frontline workers' experiences with new policies should not be studied in isolation. So, we have argued throughout this thesis that policies have a history. This results in frontline workers having a certain policy predisposition. This underscores the need for not only investigating frontline workers experiences with specific policy programs, such as the introduction of a new school curriculum, but also how they identify with government policies in general and, thus, whether or not they experience general policy alienation. Not taking the latter into account might result in a failure to understand why the implementation of new government policies is, or is not, supported by frontline workers. Hence, our main research question is:

How can the general policy alienation of frontline workers be conceptualized and measured, what are its causes and what is its influence on implementation willingness?

Before answering this question in the general conclusion, we first synthesize the results of the four empirical chapters.

6.1.1 Synthesizing the results

In the first study (chapter 2), we introduced the concept of general policy alienation, and defined it as the *overall* experience of frontline workers with government policies. We showed general policy alienation should be conceptualized as having two dimensions, namely: powerlessness (rather, frontline workers should

feel 'powerful' and feel they have the power to influence government policies at multiple levels); and, meaninglessness (rather, frontline workers should perceive policies as 'meaningful' and feel that government policies have added value for both society and their own clients).

Furthermore, theoretically we related general policy alienation to the consequences of policy accumulation, i.e. the continuous aggregation of policies that historically follow upon each other, and the new rules, regulations, and organizations that result. Despite the fact that some of our respondents did not experience any policy alienation at all – clearly, there are Dutch teachers that feel powerful and have the impression government policies are meaningful – the average scores on general policy were quite high. These scores indicate that Dutch secondary school teachers, in general, did not identify with government policies. They have the impression that they lack sufficient power to influence government policies at the national, organizational and personal level. Besides that, a significant number failed to perceive these policies as meaningful, either for society as a whole, or for their own students – or both. Relating these findings to the concepts of change fatigue and change cynicism, it may be that frontline workers experience something akin to policy fatigue or policy cynicism. This is not the same as private sector employees developing cynical attitudes that characterize organizational change efforts as just the 'flavor of the month' (Herold *et al.*, 2007), but rather it is about frontline workers developing cynical attitudes that characterize new policies as just the 'political flavor of the month'. This is a serious problem, especially for governments, as these frontline workers form a crucial link between formulated and implemented policies and, hence, between governments and citizens (Bartels, 2013; Tummers *et al.*, 2015).

In addition, our analyses found initial evidence that frontline workers' general policy perceptions were indeed related to their perceptions of a specific new policy program. That is, the analyses showed that frontline workers that had a relatively high level of general policy alienation also perceived specific policy programs (in our study: data-driven teaching) as less meaningful. This suggests that, if one wants to fully understand frontline workers' attitudes towards a specific new policy, both their perceptions of this new policy's characteristics, as well as their overall policy perceptions, should be investigated; ideally simultaneously. Excluding either set of perceptions is likely to result in an inability to put

forward satisfactory explanations of why frontline workers do, or do not, identify with a specific new policy.

Thus, the main advice resulting from this study would be to bring in policy history, and use it to shed light on frontline workers current experiences with policies.

Finally, the developed and validated measurement scale enables future researchers to quantitatively examine the antecedents and effects of general policy alienation. Although we conceptually link frontline workers' general policy alienation to the consequences of policy accumulation, we are not implying that general policy alienation is the result only of accumulated past policy experiences. Still, we have provided some initial evidence of the latter through our correlational analysis between policy consistency and general policy alienation. Greater perceived policy consistency - an indicator of more continuous policy accumulation - seems to be related to lower general policy alienation.

In the second study (chapter 3), we developed a short, but valid, and reliable measure of general policy alienation using three independently collected datasets. To do so, we adopted a systematic 10-step procedure that may also be helpful for researchers to develop short versions of other measures. This resulted in a five-item measure to gauge frontline workers' overall cognitive disconnectedness (or: connectedness) regarding government polices (Van Engen *et al.*, 2016). The measure allows future researchers to easily assess frontline workers' earlier experiences with government policies and to investigate the (behavioral) effects of this predisposition.

There are two main implications. The first is that there are now two validated measures of general policy alienation (a short and a long version). Although the current study has produced substantial evidence for the success of the short measure in capturing the essence of the original measure, the former cannot serve as a full replacement for the latter; the measures clearly serve different purposes. This implies that future researchers who want to use the policy alienation concept in their research should decide for themselves as to which measure is the best choice. If the aim of a study is to incorporate the effect of frontline workers' overall policy perceptions, then the short measure is recommended. Research on public service motivation reveals that much of the current understanding of this concept is based on studies using a general or global measure of PSM (e.g.,

Stazyk & Davis, 2015; for an overview see Wright *et al.*, 2013). The short measure of general policy alienation may serve this goal, too, as it is more easily integrated in surveys. We hope this will contribute to the application of the general policy alienation concept by public administration scholars world-wide.

This brings us to the second main implication. In recent years we have witnessed a clear increase in the number of quantitative public administration studies (Groeneveld *et al.*, 2015). In line with this, we have witnessed an increase in the number of measures being developed by public administration scholars (e.g., public leadership roles by Tummers & Knies, 2016 or red tape by Van Loon *et al.*, 2016). In light of the limited questions that can usually be included in a survey and the contextually rich studies that public administration scholars usually (aim to) conduct, it might be expected that short measures will be developed for these. Smith *et al.* (2000) noted that useful, valid and reliable short measures can only be developed by following strict procedures. We proposed and used a systematic 10-step procedure for developing our short measure. We hope other researchers will find this procedure useful. Although we are not the first in the public administration field to develop short measures, – with the work on public service motivation probably being the most exemplary (e.g., Vandenabeele, 2008; Kim *et al.*, 2013) – we believe that our 10-step procedure offers a good starting point for valid and reliable short measure development.

In the third study (chapter 4), we investigated the effect of policy consistency on how frontline workers' perceived policy meaningfulness and legitimacy. We designed a survey experiment in which we manipulated consistency (i.e., consistent versus inconsistent), as well as the policy topic (i.e., professional development of teachers versus education inequality). We found that, in line with our expectations, policy consistency positively affected perceptions of meaningfulness and, particularly, legitimacy. Apparently, frontline workers considered policies to have more added value and to be more legitimate if they were more consistent. This is possibly because it simply takes some time to identify with policies.

Furthermore, we tested how the relationship between policy consistency and meaningfulness and legitimacy was moderated by discretion. This is important, because street-level researchers have repeatedly shown that having discretion is of the utmost importance for frontline workers, as well as it being a defining characteristic of their work. We found that the effect of policy consistency on mean-

ingfulness and legitimacy was, indeed, affected by discretion; although this effect was not particularly strong, nor always statistically significant. Finally, we found that whether consistency leads to more meaningfulness and legitimacy is also influenced by the type of policy that is (dis)continued. Our results suggest that it was not necessarily the case that the continuation of a specific policy was always valued positively by frontline workers. Our results indicate that this depends on the type of policy under study, such as the main problem it aims to address, as well as the type of policy instruments adopted. In our study, this is underscored by the finding that the continuation of a policy that restricts professional leeway has a negative effect on teachers that experience low discretion.

To summarize, our findings underscore the potential positive impact of policy consistency on perceived meaningfulness and legitimacy. Although our study is to some extent at odds with the nature of political decision- and policymaking, it suggests that keeping an eye on policy consistency might be a useful strategy for governments to improve public service delivery by increasing policy support among frontline workers. Although frontline workers may not find a specific policy meaningful, or the best way to deal with societal challenges and create public value, they appear to be more likely to support this policy if they know – possibly from previous experience - that the government is willing and able to maintain this policy over time.

In the fourth and final empirical study (chapter 5), we adopted - contrary to the other chapters in this thesis - an asymmetric approach to study the relationship between powerfulness, meaningfulness and implementation willingness. Specifically, we tried to establish the motivating effect of powerfulness for implementation willingness, and how this depends on meaningfulness. We believed an asymmetric approach could be helpful, because it allowed us to detect whether the influence of powerfulness and meaningfulness might work only, or mainly, in one direction. In other words: the change in implementation willingness might not be of the same magnitude or direction when powerfulness is added as to when it is taken away (which is the case when we assume symmetric effects). This might help explain why in quantitative empirical studies, the relationship between powerfulness and implementation willingness appears to be not as strong as the literature on discretion during policy implementation suggests.

Our research results indicated, first, that powerfulness is a quasi-necessary condition for high implementation willingness. We concluded from analyses of two datasets collected in the Dutch education and healthcare sector that the majority of frontline workers who feel powerful also have high implementation willingness. Second, we tested whether powerlessness (i.e., the opposite of powerfulness) is a quasi-necessary condition for low implementation willingness. We found mixed evidence for this, which aligns with a classic insight from motivation theory; the things that make people feel satisfied and motivated can be different in kind from the things that make them feel dissatisfied (Herzberg *et al.*, 1959). So, while powerfulness can result in high implementation willingness, the 'opposite' is not automatically true as well, i.e. that powerlessness results in low implementation willingness. Third, and again in line with our assumptions, we found that, in combination with policy meaningfulness, powerfulness is quasi-sufficient for high implementation willingness. In other words, when frontline workers felt that they had both high powerfulness and that the policy was meaningful for society, this strengthened their willingness to implement it.

In summary, this study lends robust support to a bottom-up view on discretion as an inevitable and potentially beneficial aspect of frontline implementation, as we find that possibilities to participate in and influence public policies are, apparently, a prerequisite for frontline workers to be willing to implement the policy. In doing so, it shows street-level scholars how it can be useful sometimes to move from a correlational logic to the consideration of asymmetric patterns when studying policy implementation and frontline workers' critical role in successfully achieving this.

6.1.2 General conclusion

Researchers, traditionally studying bureaucracies and policymaking from a top-down perspective, have started to acknowledge the inevitability of the 'human factor' and, hence, the fact that individual preferences and personal standards play a role in discretionary decisionmaking at the frontline (Lipsky, 1980). Within the top-down perspective, this is generally regarded a problem of control. For instance, adopting a principal-agent approach, Brehm and Gates (1997) studied how those lower in order carried out requests from higher order principals. Likewise, the literature on policy implementation has focused mostly on the vertical

dimension of public government (Hupe *et al.*, 2015), where the central question is more or less how practices at the frontline align with policies-as-formulated (Brodkin, 2015).

The field of street-level bureaucracy focuses on bureaucrats at the frontline of government decisionmaking and implementation. Typical characteristics of these 'frontline workers' were that they had direct contact with citizens on a daily basis, and had considerable discretion in making decisions. Both these characteristics made them a relevant scholarly subject (Raaphorst, 2018b), because public policies inherently allocate scarce resources (Easton, 1965). Discretion is not only inevitable - policies, rules, and laws are simply never specific enough (Hoag, 2011) -, it is also necessary because frontline workers need to be responsive to individual needs (Evans, 2010). For instance, a police officer can decide whether or not to impose an on-the-spot-fine (Lipsky, 1980), regardless of the targets policymakers have set. Thus, policymakers are highly dependent on frontline workers. This explains why they can and *do* cause government's problems when they do not act in line with their policies (Brehm & Gates, 1999). In the current study, we were not particularly interested in whether this is desirable or not (one can easily think of arguments pro and con), but rather how our questioning could better understand the considerations of frontline workers when confronted with (new) policies.

Some years back, Tummers (2012) had a comparable interest and noticed "although prominent policy implementation scholars have emphasized the crucial role of implementers identifying with the policy, few have developed and tested a framework for analysing this topic" (O'Toole, 2000). Therefore, he developed the concept of policy alienation to analyze systematically and coherently to what extent frontline workers identified with specific government policies (Tummers *et al.*, 2009).

Our study clearly draws from this work, yet shows the added value of making a conceptual distinction between specific and general policy alienation. We define the latter as "an overall cognitive disconnectedness from government policies" (Van Engen *et al.*, 2016) and argue this distinction between frontline workers' specific and overall policy experiences is relevant for at least three reasons. First, we see that the experiences of frontline workers with new policies are often studied in isolation (e.g., Handley & Howell-Moroney, 2010; Sager *et*

al., 2014); thus ignoring the fact that these policies were and are not developed in a vacuum (Hogwood & Peters, 1982). Second, policy experiences should be understood in terms of their 'history'. We refer to this as policy accumulation; i.e. the continuous aggregation of policies that follow each other. Third, this distinction allows researchers to account for the fact that frontline workers might not support a specific new policy at all, but overall do support government policies in their field – or, obviously, the other way around.

All the reasons outlined above, underscore how adopting such a viewpoint may contribute to a more realistic and nuanced understanding of policy implementation success and failure. In line with the earlier work on policy alienation (Tummers, 2012), general policy alienation can best be conceptualized as having two main dimensions: powerlessness and meaninglessness. This is logical, because with these dimensions, it is acknowledged that to support a policy, frontline workers should at least feel that they have the power to influence government policies at the national, organizational and micro-level, as well as have the idea that the policies have added value for both society and clients. If these conditions are not present, it is more likely frontline workers will feel alienated from policies, both currently, as well as in the future.

With the help of the two general policy alienation measures we developed, this study firstly indicates that general policy alienation is related to frontline workers' perceptions of specific policies and their implementation willingness (Van Engen *et al.*, 2016). This means that, if frontline workers experience a higher degree of general policy alienation, they are also less likely to support a specific new policy introduced by the government. No matter how positive they may be about the new policy, their previous policy experiences affect them. In this example, this is in a negative way, although the opposite is logically also possible. We believe this illustrates how the combination of, and the interaction between, general and specific policy experiences is the better way to study policy implementation. This approach is, at least, more accurate and realistic than studying them in isolation. As such, this dissertation adopted a bottom-up approach that allows for the study of the broader context of behavior at the frontline. This is in line with, among others, Lipsky (1980), Maynard-Moody and Musheno (2003; 2012), Hill and Hupe (2009), Gofen (2014), and Evans (2015).

Our results suggest that frontline workers value policy meaningfulness more highly than powerfulness, as we found that implementation willingness was, apparently, more strongly-related to the former than to the latter. However, it should be noted that this statement should be nuanced based on our study where we adopted an asymmetrical approach to the relationship between powerfulness and implementation willingness (chapter 5). Powerfulness is also important for implementation, yet more as a prerequisite than as a determining factor (Thomann *et al.*, 2018).

Our study, furthermore, indicates that an important driver of alienation, besides policy accumulation, can be policy consistency. Although we did not find support that frontline workers always favored consistency over inconsistency – sometimes they detested a policy so much, they simply want it to be stopped –, our research indicates consistency is, overall, positively valued. This is illustrated by the following quote provided by a teacher in our 2016 survey: "*They are simply not interested in consistency and stability. Each new Minister has his own ideas and immediately gets rid of his predecessor's policies. I simply beg them: stay away from what is going well.*" If frontline workers are continuously confronted with government policies that they do not support, this could make them resistant and to view new policies as just the 'political flavor of the month' (cf. Herold *et al.*, 2007). This is likely to have a negative effect on their implementation willingness and, hence, on successful policy implementation.

However, it should be clear, that we do not argue that policy accumulation and the introduction of new policies is undesirable in itself. Rather, we believe that rapid, inconsistent policy changes may affect how frontline workers perceive and enact policies. In particular, we found a strong relationship between consistency and legitimacy. This suggests that government actions in terms of consistency of public policy may influence the degree to which these actions are perceived as justified and appropriate. Recent studies (for an overview, see Mintrom & Luetjens, 2017) have indicated that discussions of public value have emphasized three important aspects: delivering meaningful services, achieving preferred social outcomes, and maintaining trust and legitimacy. The findings of our study suggest that frontline workers, at least to a certain extent, have the impression that achieving social outcomes (i.e., meaningful policies) and maintaining trust and legitimacy are contested.

To conclude, in terms of policy implementation, policymakers (still) do not always get what they want. Hence, it is not surprising that policy implementation is still one of the main challenges for civil servants worldwide (O'Toole, 2004) and one of the key theoretical and empirical puzzles for public administration scholars (Hupe, 2014; Sandfort & Moulton, 2015; Ansell *et al.*, 2017). This explains the continuing debate in academia and practice on how to account for the complex, messy, and, sometimes, contradictory implementation of public policies (e.g., Young & Lewis, 2015; Algemene Rekenkamer, 2017; Siciliano *et al.*, 2017; Tjeenk Willink, 2017). This study contributes to this debate by further investigating the role of frontline workers, and how they perceive and enact government policies from a bottom-up, historical and quantitative perspective.

6.2 DISCUSSION

6.2.1 Academic contributions

This thesis contributes to policy implementation and street-level bureaucracy literature, scale development in public administration research, and the study of education policy.

Contribution to policy implementation and street-level bureaucracy literature

Street-level bureaucracy research - as well as the strongly connected field of policy implementation research - has evolved since Pressman and Wildavsky (1973) and Lipsky (1980). However, several theoretical and methodological issues continue to exist. These include the specification of the dependent variable (Hupe, 2013), and the problem of the 'too many variables' (Goggin, 1986) on the side of potentially explanatory factors at the strategic, tactical and operational level. It has been noted that to make the study of street-level bureaucracy "both generalizable and comparative" is an issue in its own right (Hupe *et al.*, 2015, p. 376). We aimed with this quantitative study to (partially) solve these disputes, by testing, among others, implementation theories using large datasets.

The first contribution of this study is that we highlighted that frontline workers' policy perceptions should be understood in their historical context. Siciliano

et al. (2017, p. 889) considered this perspective relevant and stated "personal beliefs that frontline workers hold toward a particular policy prior to implementation and the social processes and interactions that influence the formation of those beliefs are areas that have received less attention in research examining frontline bureaucratic behaviors". We developed the concept of general policy alienation to do this building on the work on policy alienation by Tummers, Bekkers, and Steijn (2009; 2011; 2012). In line with these studies, we found that both powerfulness and meaningfulness matter for policy implementation, but – if one has to choose - that meaningfulness seems to be the decisive factor. Furthermore, our study suggests that the combination of general evaluations of government policy and characteristics of the new policy to be implemented is the ultimate combination in explaining implementation willingness – obviously, if combined with other personal, organizational and societal characteristics.

Furthermore, our findings underscore the potential positive impact of policy consistency. This aligns with the literature adopting a rational perspective on policymaking, and the literature emphasizing the status quo bias of frontline workers (e.g., Fleming *et al.* 2010; Arnold & Fleischman, 2013). Our study, although to some extent at odds with the nature of political decision- and policymaking (Hill & Hupe, 2009; Head & Alford, 2015; Beland & Howlett, 2016), suggests that keeping an eye on policy consistency might be a useful strategy for governments to improve public service delivery, via increased policy support among frontline workers. Although frontline workers may not find a specific policy meaningful, or the ultimate way to deal with societal challenges or create public value, they appear to be more likely to support this policy if they know – perhaps from previous experience - that the government is willing and able to maintain this policy over time.

This illustrates an interesting paradox: Although politicians have full democratic and legal authority to introduce inconsistent policies (if, of course, supported by a majority in the House of Representatives), it can make it more difficult for administrators to successfully implement these policies. Interestingly in this regard is a recent study of Olsen (2017), who found that citizens evaluated policymakers more positively by their actions, rather than by their inactions – regardless of the outcome. Changing policy from this point of view is a potentially positive choice of action for policymakers; as it might result in more

positive evaluations of citizens (i.e., potential voters). Changing policy from our study's point of view, is perhaps not always the best option, as it might result in less positive evaluations of frontline workers. Thus, based on these findings, the challenge for politicians and governments seems how to balance these different interests and perspectives when initiating, formulating and implementing their policies (Howlet *et al.*, 2015).

Finally, it should be clear that the aim of our study was not to claim that policies should not be changed. Policies must certainly be flexible enough to adapt to new technologies, changing circumstances and societal developments. For instance, research indicated that 'big data' is here to stay, and will be reflected in policies (Giest, 2017). Besides that, noncompliance of frontline workers, and subsequent governmental responses, should also be understood as a source of policy changes and an interactive, ongoing process, in which noncompliance may gain social acceptance (Gofen, 2015). Thus, policies should be fluid, not rigid. However, inconsistent policies may have negative consequences for policy implementation. This implication of our study highlights a relevant and, as yet, unsolved public administration dilemma, namely: what may be regarded as perfectly legitimate and efficient from a top-down point of view, may be regarded as entirely illegitimate and inefficient from a bottom-up point of view (Sabatier, 1986; Brodkin, 2012; Gofen, 2014; Alon-Barkat & Gilad, 2016).

However, the reality is that public values can only be achieved if governments and frontline workers cooperate and align their interests for society's sake (Bryson *et al.*, 2015). If this is not achieved, and divergent perspectives and behavior arise, core public values are put at risk. Hence, it is crucial that frontline workers adhere to the values of fairness, equality, and equity when implementing policies that were decided upon through democratic procedures (Brehm & Gates, 1999). Governments, on the other hand, have the responsibility to create the circumstances in which frontline workers may do so.

Contribution to scale development in public administration research

Public administration research is becoming increasingly quantitative. As seen in psychological and managerial research, the result is a growing demand for valid and reliable measures. However, it has been noted that the field of public administration lags behind other social sciences (Perry, 2016). This is problematic,

because valid and reliable measures can only be developed by following strict procedures (Smith *et al.*, 2000).

In this study, two measures of general policy alienation were developed and tested, observing stringent criteria. This is important because frontline workers' policy predispositions (i.e., their degree of general policy alienation), as crystallized attitudes, might heavily condition the influence of government behavior on their policy evaluations (cf. Tesler, 2015). By capturing this, our measures acknowledged that frontline workers brought with them a history of government policy (changes) and, hence, general ideas about their effectiveness, legitimacy, equity and manageability. Thus, our measures enabled the application of a typical public administration perspective in change management and policy implementation research (Kuipers *et al.*, 2014). This application was especially relevant in light of the recent increase in public administration studies borrowing and extending theories from the field of psychology, or simply: the rise of the behavioral public administration (Grimmelikhuijsen *et al.*, 2017).

We developed a 26-item measure first and then, second, a short compromised five-item measure, based on the original measure. We conducted analyses on three independent datasets. These showed that the original multidimensional 26-item measure can be abbreviated to a short five-item measure with limited compromises on validity and reliability. Developing such a measure is relatively new to public administration research. We hope that the systematic 10-step approach may help others researchers to develop short versions of other measures in public administration research. This strengthens the quality of quantitative (public administration) research by promoting deliberate short-scale development. This also prevents researchers from creating ad-hoc short measures that makes it difficult to compare research results and impairs the development of a common body of knowledge. We believe this is particularly relevant in light of the notion that the public administration discipline has relatively little 'homegrown' concepts and public administration researchers infrequently develop scales themselves (Grimmelikhuijsen *et al.*, 2017).

Contribution to education policy literature

A significant amount of research, both in educational (Klassen & Chiu, 2010; You *et al.*, 2017) and public administration research (Grissom *et al.*, 2016; Janssen,

2016), has been conducted on the job to assess (dis)satisfaction, burn-out and intention of teachers to leave and to posit factors that potentially explain this. Furthermore, several studies have focused on the question how teachers perceive government policies (Matlock *et al.*, 2016) or how new public management practices take over education systems (Aoki, 2015) and what consequences this might have for teachers. For instance, Matlock *et al.* (2016) studied US teachers' views on the common core state standards (i.e., national standards and assessments) and its implementation, which the authors consider the most significant change in American schools' history. This study particularly aimed to address teachers' views and support toward this policy. They measured these items with an 66-item instrument they developed themselves, where it would have been perfectly possible – given the topic under study – to apply the (general) policy alienation framework. The author is unaware of any systematic framework to analyze general experiences of teachers with government policies. Despite the fact that this could clearly be a relevant contextual factor when studying, for instance, teachers' job satisfaction, burn-out or intention to leave.

Our study convincingly shows how the (general) policy alienation framework can be a fruitful instrument for researchers that investigate education policy, or what consequences education policy might have on day-to-day experiences of teachers, as well as to make better sense of education policy implementation, and how teachers perceive and enact policies. Although the general policy alienation framework was developed within the public administration discipline, we hope also researchers from the education research discipline will find the framework and the 'historical perspective' it adopts useful.

6.2.2 Limitations

As with all studies, this study has limitations. This section discusses three limitations that resulted from methodological choices and choices in the research design.

Selection of case study

The majority of the research presented in this thesis (excluding the sample of healthcare workers we used in chapter 5) was based on data collected in the Dutch secondary education sector. We provided three main arguments as to

why this sector is a relevant case to study general policy alienation, namely: the important role school leaders and teachers play in delivering public services, the fact that there has been a lot of reshuffling of authority between government and schools, and that there has been a large number of policy changes. We expect this to be quite similar in other public sectors where policymakers are heavily dependent on implementing organizations for their policy's success, such as the healthcare or the safety domain where similar implementation challenges have been witnessed (e.g., Gofen, 2015; McDermott *et al.*, 2015).

However, we cannot exclude the fact that specific characteristics of our research context influenced our research results. For instance, it could be that two defining characteristics of the Dutch secondary education sector, namely: the combination of a relatively decentralized sector and the relatively large number of policy changes, has an effect on the degree of general policy alienation we report. It could be that the degree of general policy alienation is lower in relatively centralized education sectors or in sectors where a smaller number of policy changes is introduced – or the other way around. Nevertheless, the assumptions we tested in this thesis are grounded in street-level bureaucracy and policy implementation theory and cohere with the findings of previous studies (e.g., Lipsky, 1980; Maynard-Moody & Musheno, 2003; Hill & Hupe, 2009; Tummers, 2012). Therefore, although we should be cautious in generalizing our results, we are confident in the results we present in this thesis, namely: that frontline workers experience such a thing as general policy alienation, and that this influences how they perceive and enact newly introduced policies. However, it should be clear that the strength of the relationships we present may be different for, as examples, teachers in the USA or for police officers in the Netherlands.

Causal inference

Not all findings we present in this thesis allow for causal inference. This is not a major problem, given that this is not our main research aim in all empirical chapters. In chapter 4, on the other hand, establishing a causal relationship between policy consistency and policy perceptions was our main research aim. Therefore, we conducted a survey experiment among Dutch teachers to allow us to draw conclusions about causality (James *et al.*, 2017). The results provided support for the hypothesized causal relationship, namely: that policy consistency results in

more policy meaningfulness and stronger government legitimacy. In chapter 2 and 3, where we studied the correlations between policy alienation and related concepts, already we had established initial evidence of such relationships. Yet, our main goal there was not to prove a causal relationship, but rather to study whether our measures behave as they should behave with theoretically (un)related concepts.

Also, in chapter 5, establishing a causal relationship was not our main research aim. What we did was to investigate the motivational effect of policy powerfulness and meaningfulness for implementation willingness from an asymmetric logic. However, our findings do suggest such a causal relationship – or at least, do not exclude such a relationship. Obviously, we urge future research to establish causality for the relationships under study in these chapters. Although it should be clear that, not in all cases and under all circumstances, an experimental approach will be helpful (Van Engen, 2019).

Policy implementation success is not policy success

This study investigated policy implementation and what factors may contribute to successful policy implementation by frontline workers. This is relevant, because if a policy is not implemented, it cannot be evaluated as to whether or not this policy contributes to solving the societal challenges it aims to address. However, it should be noted that successful policy implementation does not equal policy success. Rather, we believe a distinction should be made between political success, policy implementation success and policy success, as proposed by Marsh and McConnell (2010). Although it should be clear that these types of success are (strongly) connected. For instance, in the case of Dutch secondary education, a political success could be that the House of Representatives supports a new policy proposal by the Minister of Education that aims to introduce a new curriculum that has a motivating effect on students and better prepares them for the next step in their school career. Policy implementation success, then, could be that this new curriculum is developed, tested and, ultimately, used in all schools.

Finally, we may consider this policy a success – which is basically only possible if the policy is successfully implemented – if research shows that the new curriculum indeed increases student motivation, as well as proves that students are better prepared for the next step in their school career. Besides that, different

viewpoints exist on what successful policy implementation actually entails. From a governments' perspective, for instance, this may be that frontline workers do exactly what governments want them to do. However, what you see, in terms of formal policy, is not always what you get, in terms of policy-as-produced (Brodkin, 2012). It is important to take these distinctions into account while reflecting on this study's results. At the same time, it suggests interesting avenues for future research.

6.2.3 An agenda for future research

We distinguish three relevant themes for future research based on our study's research findings and implications, as well as the limitations outlined above.

Revival of policy implementation research

First, we urge for a revival of policy implementation research. Policy implementation research, traditionally, was rather qualitative (e.g., Pressman & Wildavsky, 1973; Lipsky, 1980; Sandfort, 2000, Maynard-Moody & Musheno, 2003; Hill & Hupe, 2009). The rise of the 'behavioral public administration' (Grimmelikhuijsen *et al.*, 2017) and the growing focus on quantitative studies and research methods in public administration, opens up interesting opportunities for policy implementation and street-level bureaucracy research (as is illustrated by, among others, Andersen & Jakobsen, 2017; Raaphorst, 2018a; Schott *et al.*, 2018).

In one of the empirical chapters of this study, we conducted a survey experiment, showing how policy consistency may contribute to more policy meaningfulness and more government legitimacy. Another study showed how the majority of frontline workers who experienced high implementation willingness, also experienced high discretion. Hence, our study illustrates how a quantitative approach offers the opportunity to test theoretical propositions drawn from qualitative implementation and street-level bureaucracy research, including, for example, how frontline workers deal with uncertainties related to information and interpretation problems in interacting with citizen-clients (Raaphorst, 2018a), or whether the explicit treatment of public value creation as a policy goal can improve the fit between original policy intentions and the delivery of public services (Mintrom & Luetjens, 2017). We believe this development will bring on policy implementation research further, by complementing macro-level 'grand

implementation theories' with their micro-level underpinnings (Grimmelikhuijsen *et al.*, 2017; Van Engen, 2019).

Relevance of field experiments and multilevel research

Second, we recommend the use of field experiments and multilevel research when studying policy alienation and its consequences for successful policy implementation. Although we used a survey experiment to establish a relationship between policy consistency, discretion and perceived policy meaningfulness and government legitimacy, we would recommend replicating such findings with field or laboratory experiments. The clear advantage of field experiments, as compared to survey experiments, is the real-life context in which ecological validity is naturally guaranteed (Grimmelikhuijsen *et al.*, 2017; James *et al.*, 2017).

We recommend also further study of the link between policy perceptions and actual behavior. The limitations section highlighted how policy implementation success does not equal policy success. The same is true for frontline workers' policy perceptions and their behavior: perceptions do not equal behavior. Although it is generally acknowledged that perceptions affect behavior, and we showed how policy perceptions affect implementation willingness, we cannot be fully sure that - and if so, how - policy alienation influences implementation behavior. This should be studied in the field.

We also believe it would be relevant to study general policy alienation from a multilevel perspective. The literature on organization socialization (e.g., Oberfield, 2010; Hatmaker *et al.*, 2011; Vigoda-Gadot & Beeri, 2012; Teodoro, 2014) suggests that the way frontline workers perceive government policies is influenced by the organization in which they work. What characteristics make organizations – besides organizational leadership and tactical powerlessness, which we study in this thesis – either 'policy welcoming' or 'policy resistant', and are these characteristics manipulable?

Study policy alienation in multiple countries and public domains

Third, we deem it important that general policy alienation is studied in other countries and public sector domains to complement the Dutch secondary education sector studies. Although we are aware that this is not a highly original suggestion, we believe this is important to further develop the policy alienation

concept. As we have witnessed with research on red tape (e.g., Van Loon *et al.*, 2016) or public service motivation (e.g., Kim *et al.*, 2013), it is important that the (general) policy alienation concept is used by public administration scholars world-wide to show its potential added value for our discipline. The replication of our study's findings would improve the general applicability of the results presented in this research. Although we expect our findings to apply to other sectors where policymakers are highly dependent of frontline workers who have significant discretion in doing their work, we have not explicitly tested this. It would improve the feasibility of the general policy alienation concept if future research addressed these concerns. We recommend two particular avenues for future research.

First, we recommend comparative research in different countries to study how general policy alienation may be dependent on the specific education context. How do particular characteristics influence the degree of alienation? For instance, it is known that the Dutch education sector is relatively decentralized (EP-Nuffic, 2015), whereas the education sector in Singapore is relatively centralized (Dimmock & Tan, 2016). It would be relevant to find out if the average general policy alienations scores between teachers working in these countries differ and, if so, how this might be related to the degree of (de)centralization in the sector – as well as other key characteristics, such as initial teacher training programs, salary versus teaching obligations, and the ratio of full-time versus part-time working teachers.

Second, we recommend research in multiple domains in order to study how general policy alienation may be dependent on specific job characteristics. For instance, comparative studies can be conducted in the education, health and safety sector. Do teachers, on average, find government policies more meaningful than nurses? Or, do nurses, on average, more strongly have the idea that they have more or less influence on policy content at the national level than, say, police officers? The ultimate question, then, is how these differences can be explained, including how this may be dependent on job positions (manager, versus middle manager, versus frontline worker) (e.g., Maynard-Moody & Musheno, 2003; May & Winter, 2009; Brodkin, 2011).

6.2.4 Implications for practice

The findings of this study have three main implications for practice.

Learn from the past

First, this study underscores that, in order to more comprehensively understand policy implementation success and failure, attention should be paid to a policy's history. Specifically, we argue that taking frontline workers previous policy experiences into account better explains the (un)successful implementation of new policies. This insight may urge politicians and civil servants, when introducing a new policy – or already when they consider introducing a new policy – to take this history into account from the start. For instance, this awareness may motivate them to evaluate how a new policy aligns or disaligns with previously introduced policies or to estimate whether frontline workers may start to feel 'overwhelmed' by the number of policy changes announced.

Throughout this thesis, we provided several examples of policy implementation failure. It is not a challenging task to find more examples: "Very often, political decisions... are at odds with the implementation possibilities" (Tjeenk Willink, 2017). Our study aimed to contribute to a better understanding of implementation failure and the role frontline workers have in this failure. Ultimately, these insights may contribute to improved policy formulation and implementation processes, as well as more policy alignment.

Deliberately apply consistent policies

Second, this study marks the relevance of policy consistency for frontline workers. Although we do not find support for the statement that policy consistency is always considered the best option from frontline workers' point of view, we do find support for the statement this is a factor frontline workers take into account when evaluating (a series of) policy measures. Our study suggests that policy proposals and changes that follow upon each other quite quickly can have a detrimental effect on policy implementation willingness. It is important that policymakers are aware of this, as this awareness may help them develop a more consistent policy program. Implementing policies is not easy, but usually requires significant effort. If frontline workers can be relatively sure their investment 'is worth it', this may increase their implementation willingness.

Interestingly, our results indicate that frontline workers do not always support the continuation of policies. If frontline workers do not support the content of a policy, they might welcome its abolishment. In our case study, for instance, this was recently the case with the pulling back of an obligatory diagnostic test and the obligatory participation in a so-called register of teachers (Algemeen Dagblad, 2018). A significant number of teachers responded quite positively to these decisions, for instance: "The decision of the Minister of Education is the only correct way... Teachers were not involved in [the register's] development." (NOS, 2018). This is, in a way, understandable, yet, at the same time, increases the likelihood that teachers in the future will adopt a 'wait-and-see attitude'. From that point of view, it might be better for a government's long-term implementation success to continue policies, despite their lack of support among frontline workers, or, perhaps, even when they do not fully align with the responsible Minister's political or policy preferences.

Measure (general) policy alienation in a representative, nation-wide survey

Finally, this study results in two measurement instruments that can be freely used by practitioners to evaluate policy support among frontline workers. Depending on the specific practical issues and research questions at hand, they can apply either the long or the short measure. This may provide relevant information for governments, for instance, on the perceived added value of a policy or whether or not it is relevant to extend the participation opportunities for frontline workers. This will allow politicians, public managers and civil servants to evaluate if and how (general) policy support evolves over time, especially if these measures are used repeatedly over time. If this is combined with the study of the experiences of frontline workers with the implementation of specific new policies (see Tummers, 2012), this will result in insightful information on policy support among frontline workers. This information can be used to refine, replace or abolish policies. We recommend governments in particular to use such surveys to detect frontline workers who, either do not feel alienated from government policies at all, or have very low alienation scores. Our study indicates there are Dutch secondary school teachers who experience little to no policy alienation. We believe it would be interesting to look into the characteristics and experiences of these frontline

workers, because they can be helpful in formulating and framing policies and implementation strategies.

6.3 TO CONCLUDE

In conclusion, this thesis has investigated the relationship between general policy alienation and policy implementation. Our research findings show clearly that both academics and practitioners should pay attention to the fact that frontline workers' are not neutral implementers. They bring with them a history of government (policy) changes that affect how they perceive and behave in relation to new policies. Therefore, we believe a distinction must be made between specific and general policy experiences, as this will help to better understand policy implementation failure – or, ideally, success. Furthermore, our results underscore the potential added value of policy consistency for frontline workers. We show that policy consistency increases perceived policy meaningfulness, as well as government legitimacy. Therefore, it might be a useful strategy for government to improve their policy implementation via more committed implementers. Finally, we shed light on the ongoing discussion on discretion in policy design and implementation research. Our results show that the majority of frontline workers who feel powerful have high implementation willingness. We urge scholars and practitioners to move away from the question as to whether frontline workers should be granted discretion and on to how to best make use of frontline workers' discretion instead.

Our findings obviously are contributing a new angle on the continuing debate within policy implementation and street-level bureaucracy research as to how to account for the complex, messy, and, sometimes, contradictory implementation of public policies.

REFERENCES

A

Abbott, A. (1988). *The System of the Professions*. Chicago: The University of Chicago Press.

Ackroyd, S., Kirkpatrick, I. & Walker, R.M. (2007). Public management reform in the UK and its consequences for professional organization: A comparative analysis. *Public Administration*, 85(1), 9-26.

Ajzen, I. & Fishbein, M. (1980). *Understanding attitudes and predicting social behavior*. Englewood Cliffs, NJ: Prentice Hall.

Algemeen Dagblad (2018). Minister ziet af van invoering verplicht lerarenregister. *AD*. Retrieved August 20, 2018, from https://www.ad.nl/binnenland/minister-ziet-af-van-invoering-verplicht-lerarenregister~aaaa3b9f/

Algemene Rekenkamer (2017). *Resultaten verantwoordingsonderzoek 2016 bij het Ministerie van Onderwijs, Cultuur en Wetenschap*. Den Haag: Algemene Rekenkamer.

Alon-Barkat, S. & Gilad, S. (2016). Political control or legitimacy deficit? Bureaucracies' symbolic responses to bottom-up public pressures. *Policy & Politics*, 44(1), 41-58.

Andersen, L.B., Heinesen, E. & Pedersen, L.H. (2016). Individual performance: From common source bias to institutionalized assessment. *Journal of Public Administration Research and Theory,* 26(1), 63-78.

Andersen, S.C. & Jakobsen, M. (2017). Policy positions of bureaucrats at the front lines: Are they susceptible to strategic communication? *Public Administration Review*, 77(1), 57-66.

Anderson, J.E. (1975). *Public policy-making*. New York: Praeger.

Ansell, C., Sorensen, E. & Torfing, J. (2017). Improving policy implementation through collaborative policymaking. *Policy and Politics,* 45(3), 467-489.

Aoki, N. (2015). Institutionalization of new public management: The case of Singapore's education system. *Public Management Review*, 17(2), 165-186.

Armitage, C.J. & Conner, M. (2001). Efficacy of the theory of planned behaviour: A meta-analytic review. *British Journal of Social Psychology*, 40(4), 471-499.

Arnold, G. & Fleischman, F.D. (2013). The influence of organizations and institutions on Wetland policy stability: The Rapanos case. *Policy Studies Journal*, 41(2), 343-362.

B

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review,* 84(2), 191-215.

Barnes, C.Y. & Henly, J.R. (2018). "They are underpaid and understaffed": How clients interpret encounters with

street-level bureaucrats. *Journal of Public Administration Research and Theory*, 28(2), 165-181.

Barrett, S.M. (2004). Implementation studies: Time for a revival? Personal reflections on 20 years of implementation studies. *Public Administration,* 82(2), 249-262.

Barrett, S. & Fudge, C. (1981). *Policy and action: Essays in the implementation of public policy.* London: Methuen.

Bartels, K.P.R. (2013). Public encounters: The history and future of face-to-face contact between frontline workers and citizens. *Public Administration,* 91(2), 469-483.

Bassett-Jones, N. & Lloyd, G.C. (2005). Does Herzberg's motivation theory have staying power? *Journal of Management Development,* 24(10), 929-943.

Baumgartner, M. & Thiem, A. (2017). Model ambiguities in configurational comparative research. *Sociological Methods and Research*, 46(4), 954-987.

Bekkers, V.J.J.M., Dijkstra, G., Edwards, A. & Fenger, M. (Eds.) (2007). *Governance and the democratic deficit: Assessing the democratic legitimacy of governance practices.* Aldershot: Ashgate Publishing, Ltd.

Beland, D. & Howlett, M. (2016). How solutions chase problems: Instrument constituencies in the policy process. *Governance,* 29(3), 393-409.

Berkovich, I. (2011). No we won't! Techers' resistance to educational reform. *Journal of Educational Administration*, 49(5), 563-578.

Boerzel, T.A. & Van Huellen, V. (2014). One voice, one message, but conflicting goals: Cohesiveness and consistency in the European neighbourhood policy. *Journal of European Public Policy,* 21(7), 1033-1049.

Bordia, P., Restubog, S.L.D., Jimmieson, N.L. & Irmer, B.E. (2011). Haunted by the past: Effects of poor change management history on employee attitudes and turnover. *Group and Organization Management,* 36(2), 191-222.

Borry, E., DeHart-Davis, L., Kaufmann, W., Merritt, C.C., Mohr, Z. & Tummers, L.G. (2018). Formalization and consistency heighten organizational rule following: Experimental and survey evidence. *Public Administration*, 96(2), 368-385.

Bouckaert, G. (2012). Trust and public administration. *Administration*, 60(1), 91-115.

Brackett, M.A. & Mayer, J.D. (2003). Convergent, discriminant, and incremental validity of competing measures of emotional intelligence. *Personality and Social Psychology Bulletin*, 29(9), 1147-1158.

Brehm,. S.S. & Brehm, J.W. (2013). *Psychological reactance: A theory of freedom and control.* New York: Academic Press.

Brehm, J. & Gates, S. (1999). *Working, shirking, and sabotage: Bureaucratic response to a democratic public.* Ann Arbor: University of Michigan Press.

Brodkin, E.Z. (1997). Inside the welfare contract: Discretion and accountability

in state welfare administration. *The Social Service Review,* 71(1), 1–33.

Brodkin, E.Z. (2011). Policy work: Street-level organizations under new managerialism. *Journal of Public Administration Research and Theory,* 21(2), 253-277.

Brodkin, E.Z. (2012). Reflections on street-level bureaucracy: Past, present, and future. *Public Administration Review*, 72(6), 940-949.

Bronneman-Helmers, R. (2008). *Tijd voor onderwijs: Vijftien jaar onderwijs-vernieuwingen in Nederland.* Den Haag: Sociaal en Cultureel Planbureau.

Brown, T.A. (2012). *Confirmatory factor analysis for applied research.* New York: Guilford Press.

Bryson, J.M., Crosby, B.C. & Middleton Stone, M. (2015). Designing and implementing cross-sector collaborations: Needed and challenging. *Public Administration Review*, 75(5), 647-663.

Byrne, B. (2008). Testing for multigroup equivalence of a measuring instrument: A walk through the Process. *Psicothema*, 20(4), 872-882.

C

Carless, S.A., Wearing, A.J. & Mann, L. (2000). A short measure of trans-formational leadership. *Journal of Business and Psychology*, 14(3), 389-405.

Cagan, P. (1956). The monetary dynamics of hyperinflation. In M. Friedman (Ed.), *Studies in the quantity Theory of Money* (pp. 25-115). Chicago: University of Chicago Press.

Calvo, G.A. (1978). On the time consistency of optimal policy in a monetary economy. *Econometrica*, 46(6), 1411-1428.

Castle, S. (2016). Junior doctors' strike in England disrupts care for thousands. *NY Times*. Retrieved March 28, 2017, from https://www.nytimes.com/2016/01/13/world/europe/doctors-strike-national-health-service.html?_r=0

Cayton, A.F. (2017). Consistency versus responsiveness: Do members of congress change positions on specific issues in response to their districts?' *Political Research Quarterly*, 70(1), 3-18.

Choueiki, A. (2016). Behavioral insights for better implementation in government. *Public Administration Review,* 76(4), 540-541.

Christensen, T., Laegreid, P. & Rykkja, L.H. (2016). Organizing for crisis management: Building governance capacity and legitimacy. *Public Administration Review*, 76(6), 887-897.

Chun, Y.H. & Rainey, H.G. (2005). Goal ambiguity and organizational performance in US federal agencies. *Journal of Public Administration Research and Theory*, 15(1), 529-557.

Coburn, C.E. (2001). Collective sensemaking about reading: How teachers mediate reading policy in their professional communities. *Educational Evaluation and Policy Analysis*, 23(2), 145-170.

Connell, J. & Waring, P. (2002). The BOHICA syndrome: A symptom of cynicism towards change initiatives? *Strategic Change*, 11(7), 347-356.

Cooper, P.J. (2017). "Without which nothing": Public law as the sine qua non of public administration. *Administration & Society,* 49(5), 547-568.

Coursey, D.H. & Pandey, S.K. (2007). Public service motivation measurement: Testing an abridged version of Perry's proposed scale. *Administration & Society*, 39(5), 547-568.

Crede, M., Harms, P., Niehorster, S. & Gaye-Valentine, A. (2012). An evaluation of the consequences of using short measures of the big five personality traits. *Journal of Personality and Social Psychology*, 102(4), 874-888.

Crossley, C.D., Bennett, R.J., Jex, S.M. & Burnfield, J.L. (2007). Development of a global measure of job embeddedness and integration into a traditional model of voluntary turnover. *Journal of Applied Psychology*, 92(4), 1031-1042.

D

Dahl, R.A. (1998). *On democracy.* New Haven, CT: Yale University Press.

Davis, K.C. (1969). *Discretionary justice: A preliminary inquiry.* Baton Rouge, LA: Louisiana State University Press.

Dawes, J.G. (2008). Do data characteristics change according to the number of scale points used? An experiment using 5 point, 7 point and 10 point scales. *International Journal of Market Research*, 51(1), 1-20.

De Boer, N. & Eshuis, J. (2018). A street-level perspective on government transparecny and regulatory performance: Does relational distance matter? *Public Administration Review*, early online view: doi: 10.1111/padm.12517.

De Fine Licht, J. (2014). Policy area as a potential moderator of transparency effects: An experiment. *Public Administration Review,* 74(3), 361–371 .

DeCelles, K.A., Tesluk, P.E. & Taxman, F.S. (2013). A field investigation of multilevel cynicism toward change. *Organization Science,* 24(1), 154-171.

DeLeon, P. & DeLeon, L. (2002). What ever happened to policy implementation? An alternative approach. *Journal of Public Administration and Research Theory,* 12(4), 469-492.

DeVellis, R.F. (2003, 2012). *Scale development: Theory and applications.* London: Sage.

Deutsch, M. (1975). Equity, equality, and need: What determines which value will be used as the basis of distributive justice? *Journal of Social Issues*, 31(3), 137-149.

Dias, J.J. & Maynard-Moody, S. (2007). For-profit welfare: Contracts, conflicts, and the performance paradox. *Journal of Public Administration Research and Theory*, 17(2), 189-211.

Dimmock, C. & Tan, C.Y. (2016). Explaining the success of the world's leading education systems: The case of Sin-

gapore. *British Journal of Education Studies,* 64(2), 161-184.

Drasgow, F. & Kanfer, R. (1985). Equivalence of psychological measurement in heterogeneous populations. *Journal of Applied Psychology*, 70(4), 662-680.

Dunsire, A. (1978). *Implementation in a bureaucracy*. Oxford: Martin Robertson.

DUO (2014, 2015, 2016, 2017). *Onderwijspersoneel vo in personen*. Groningen: Dienst Uitvoering Onderwijs.

Duşa, A. (2018). *QCA with R: A comprehensive resource*. New York: Springer International Publishing.

Dutch Inspectorate of Education (2016). *Staat van het Onderwijs 2014/2015*. Utrecht: Inspectie van het Onderwijs.

E

Education International (2014). Five thousand more Norwegian teachers to join month-long strike. *Education International*. Retrieved December 31, 2014, from http://www.ei-ie.org/en/news/news_details/3173

Elliot, P. (1972). *The sociology of professions*. London: Macmillan.

Ellis, K. (2011). Street-level bureaucracy' revisited: The changing face of frontline discretion in adult social care in England. *Social Policy & Administration,* 45(3), 221-244.

Elmore, R.F. (1979). Backward mapping: Implementation research and policy decisions. *Political Science Quarterly*, 94(4), 601-616.

Elmore, R.F. & McLaughlin, M.W. (1988). *Steady work. Policy, practice, and the reform of American education*. Santa Monica, CA: The Rand Corporation.

Emmenegger, P., Schraff, D. & Walter, A. (2014). QCA, the truth table analysis and large-N survey data: The benefits of calibration and the importance of robustness tests. *COMPASSS Working Paper 2014-79.*

EP-Nuffic (2015). *Education system the Netherlands*. The Hague: The Dutch Organisation for Internationalisation in Education.

Etzioni, A. (1969). *The semi-professions and their organization: Teachers, nurses, social workers*. London: Collier-Macmillan.

Evans, T. (2010). *Professional discretion in welfare services: Beyond street-level bureaucracy*. London: Ashgate.

Ewalt, J.A.G. & Jennings, E.T. (2004). Administration, governance, and policy tools in welfare policy implementation. *Public Administration Review*, 64(4), 449–462.

F

Fiss, P.C. (2011). Building better causal theories: A fuzzy set approach to typologies in organization research. *Academy of Management Journal*, 54(2), 393-420.

Fiss, P.C., Sharapov, D. & Cronqvist, L. (2013). Opposites attract? Opportunities and challenges for integrating large-N QCA and econometric

analysis. *Political Research Quarterly*, 66(1), 191-198.

Fleming, S.M. (2010). Overcoming status quo bias in the human brain. *PNAS*, 107(13), 6005-6009.

Follet, M.P. (1924). *The creative class*. New York: Longmans, Green.

Freidson, E. (2001). *Professionalism, the third logic*. Cambridge: Cambridge University Press.

Fullan, M.G. (1986). Improving the implementation of educational change. *School Organization,* 6(3), 321-326.

G

Gabe, J., Bury, M. & Elston, M.A. (2004). *Key concepts in medical sociology*. London: Sage.

Gagné, M. & Deci, E.L. (2015). Self-determination theory and work motivation. *Journal of Organizational Behavior*, 26(4), 331-362.

Giest, S. (2017). Big data for policymaking: Fad or fasttrack? *Policy Sciences*, 50(3), 367-382.

Goertz, G. & Starr, H. (Eds.) (2003). *Necessary conditions. Theory, methodology and applications*. New York: Rowman & Littlefield.

Gofen, A. (2014). Mind the gap: Dimensions and influence of street-level divergence. *Journal of Public Administration Research and Theory*, 24(2), 473-493.

Gofen, A. (2015). Reconciling policy dissonance: Patters of governmental response to policy noncompliance. *Policy Sciences*, 48(1), 3-24.

Goggin, M.L. (1986). The "too few cases/ too many variables" problem in implementation research. *Political Research Quarterly*, 39(2), 328-347.

Grant, A.M. & Berry W. (2011). The necessity of others is the mother of invention: Intrinsic and prosocial motivations, perspective taking, and creativity. *Academy of Management Journal*, 54(1), 73-96.

Greckhamer, T., Misangyi, V.M. & Fiss, P.C. (2013). The two QCAs: From a small-N to a large-N set theoretic approach. *Research in the Sociology of Organizations*, 38, 49-75.

Greenwood, R., Suddaby, R. & Hinings, B.C.R. (2002). Theorizing change: The role of professional associations in the transformation of institutionalized fields. *The Academy of Management Journal*, 45(1), 58-80.

Grimmelikhuijsen, S.G., Jilke, S., Olsen, A.L. & Tummers, L.G. (2017). Behavioral public administration: Combining insights from public administration and psychology. *Public Administration Review*, 77(1), 45-56.

Grimmelikhuijsen, S.G. & Meijer, A.J. (2015). Does Twitter increase perceived police legitimacy? *Public Administration Review*, 75(4), 598-607.

Grissom, J.A., Viano, S.L. & Selin, J.L. (2016). Understanding employee turnover in the public sector: Insights from research on teacher mobility. *Public Administration Review,* 76(2), 241-251.

Groeneveld, S., Tummers, L.G., Bronkhorst, B., Ashikali, T. & Van Thiel, S. (2015).

Quantitative methods in public administration: Their use and development through time. *International Public Management Journal*, 18(1), 61-86.

Gruber, J.E. (1987). *Controlling bureaucracies: Dilemmas in democratic governance.* Berkeley and Los Angeles, CA: University of California Press.

H

Handley, D.M. & Howell-Moroney, M. (2010). Ordering stakeholder relationships and citizen participation: Evidence from the community development block grant program. *Public Administration Review*, 70(4), 601-609.

Harari, M.B., Herst, D.E.L., Parola, H.R. & Carmona, B.P. (2016). Organizational correlates of public service motivation: A meta-analysis of two decades of empirical research. *Journal of Public Administration Research and Theory*, 27(1), 68-84.

Hatmaker, D.M., Park, H.H. & Rethemeyer R.K. (2011). Learning the ropes: Communities of practice and social networks in the public sector. *International Public Management Journal*, 14(4), 395-419.

Head, B.W. & Alford, J. (2015). Wicked problems: Implications for public policy and management. *Administration & Society,* 47(6), 711-739.

Herold, D.M., Fedor, D.B. & Caldwell, S.D. (2007). Beyond change management: A multilevel investigation of contextual and personal influences on employees' commitment to change. *Journal of Applied Psychology*, 92(4), 942-951.

Herzberg, F., Mausner, B. & Snyderman, B.B. (1959). *The motivation to work.* New York: Wiley.

Hill, M.J. & Hupe, P.L. (2009). *Implementing public policy: Governance in theory and practice.* London: SAGE.

Hinkin, T.R. (1998). A brief tutorial on the development of measures for use in survey questionnaires. *Organizational Research Methods*, 1(1), 104-121.

Hogwood, B.W. & Peters, B.G. (1982). The dynamics of policy change: Policy succession. *Policy Sciences,* 14(3), 225-245.

Hoogerwerf, A. (1978). Bureaucratisering en overheidsbeleid. In L. Rademaker (Ed.), *Sociale problemen 1: Sociale ontwikkelingen* (pp. 51-77). Utrecht/Antwerpen: Uitgeverij Spectrum.

Howlett, M. (2004). Beyond good and evil in policy implementation: Instrument mixes, implementation styles, and second generation theories of policy instrument choice. *Policy and Society*, 23(2), 1-17.

Howlett, M., Mukhersjee, I. & Woo, J.J. (2015). From tools to toolkits in policy design studies: The new design orientation towards policy formulation research. *Policy and Politics*, 43(2), 291-311.

Hu, L. & Bentler, P.M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural*

Equation Modeling - a Multidisciplinary Journal, 6(1), 1-55.

Hupe, P. (2014). What happens on the ground: Persistent issues in implementation research. *Public Policy & Administration,* 29(2), 164-182.

Hupe, P. (2013). Dimensions of discretion: Specifying the object of street-level bureaucracy research. *Der moderne staat–Zeitschrift für Public Policy, Recht und Management*, 6(2), 425-440.

Hupe, P.L. & Buffat, A. (2014). A public service gap: Capturing contexts in a comparative approach of street-level bureaucracy. *Public Management Review*, 16(4), 548-569.

Hupe, P.L. & Hill, M.J. (2007). Street-level bureaucracy and public accountability. *Public Administration*, 85(2), 279-299.

Hupe, P.L., Hill, M.J. & Buffat, A. (2015). *Understanding street-level bureaucracy*. Bristol: Policy Press.

Huy, Q.N. (2001). Time, temporal capability, and planned change. *Academy of Management Review*, 26(4), 601-623.

I

In 't Veld, R.J. (1989). *De verguisde staat.* The Hague: VUGA.

Ironson, G.H., Smith, P.C., Brannick, M.T., Gibson, W.M. & Paul, K.B. (1989). Construction of a job in general scale: A comparison of global, composite, and specific measures. *Journal of Applied Psychology*, 74(2), 193-200.

J

James, O., Jilke, S.R. & Van Ryzin, G.G. (2017). *Experiments in public management research: Challenges and contributions*. Cambridge (UK): Cambridge University Press.

Janssen, T. (2016). *HRM en prestaties in het primair onderwijs*. PhD dissertation. Rotterdam: Erasmus University Rotterdam.

Johansson, V. (2012). Negotiating bureaucrats. *Public Administration,* 90(4), 1032-1046.

K

Kaufman, H. (1960). *The forest ranger: A study in administrative behavior*. Washington, DC: Resources for the Future.

Keiser, L.R. (1999). State bureaucratic discretion and the administration of social welfare programs: The case of social security disability. *Journal of Public Administration Research and Theory*, 9(1), 87-106.

Kenning, P. (2008). The influence of general trust and specific trust on buying behaviour. *International Journal of Retail and Distribution Management*, 36(6), 461-476.

Kerpershoek, E., Groenleer, M. & De Bruijn, H. (2016). Unintended responses to performance management in Dutch hospital care: Bringing together the managerial and professional perspectives. *Public Management Review*, 18(3), 417-436.

Kickert, W.J.M. (2010). Managing emergent and complex change: The case of

Dutch agencification. *International Review of Administrative Sciences*, 76(3), 489-515.

Kiefer, T., Hartley, J., Conway, N. & Briner, R.B. (2015). Feeling the squeeze: Public employees' experiences of cutback-and innovation-related organizational changes following a national announcement of budget reductions. *Journal of Public Administration Research and Theory*, 25(4), 1279-1305.

Kim, S., Vandenabeele, W., Wright, B.E., Andersen, L.B., Paolo Cerase, F., Christensen, R.K., Desmarais, C., Koumenta, M., Leisink, P. & Liu, B. (2013). Investigating the structure and meaning of public service motivation across populations: Developing an international instrument and addressing issues of measurement invariance. *Journal of Public Administration Research and Theory*, 23(1), 79-102.

Klassen, R.M. & Chiu, M.M. (2010). Effects on teachers's self-efficacy and job satisfaction: Teacher gender, years of experience, and job stress. *Journal of Education Psychology,* 102(3), 741-756.

Kline, R.B. (2015). *Principles and practice of structural equation modeling.* New York: Guilford.

Korman, A.K. (1974). *The psychology of motivation*. Oxford (UK): Prentice Hall.

Kroll, A. & Moynihan, D.P. (2015). Does training matter? Evidence from performance management reforms. *Public Administration Review*, 75(3), 411-420.

Kruyen, P.M., Emons, W.H.M. & Sijtsma, K. (2013). On the shortcomings of shortened tests: A literature review. *International Journal of Testing*, 13(3), 223-248.

Kuipers, B.S., Higgs, M., Kickert, W., Tummers, L.G., Grandia, J. & Van der Voet, J. (2014). The management of change in public organizations: A literature review. *Public Administration*, 92(1), 1-20.

Kydland, F. & Prescott, E. (1977). Rules rather than discretion: The inconsistency of optimal plans. *Journal of Political Economy*, 85, 473-90.

L

Lammers, J., Stoker, J.I., Rink, F. & Galinsky, A.D. (2016). To have control over or to be free from others? The desire for power reflects a need for autonomy. *Personality and Social Psychology*, 42(4), 498-512.

Lavee, E., Cohen, N. & Nouman, H. (2018). Reinforcing public responsibility? Influences and practices in street-level bureaucrats' engagement in policy design. *Public Administration*, 96(2), 333-348.

Levi, M. (1997). *Consent, dissent, and patriotism.* New York: Cambridge University Press.

Lewin, K. (1936). *Principles of topological psychology*. New York: McGraw-Hill.

Liden, R.C., Wayne, S.J., Meuser, J.D., Hu, J., Wu, J. & Liao, C. (2015). Servant leadership: Validation of a short

form of the SL-28. *The Leadership Quarterly*, 26(2), 254-269.

Lines, R. (2000). Influence of participation in strategic change: Resistance, organizational commitment and change goal achievement. *Journal of Change Management*, 4(3), 193-215.

Lipsky, M. (1980). *Street level bureaucracy*. New York: Russell Sage Foundation.

Loyens, K. (2015). Law enforcement and policy alienation: Coping by labour inspectors and federal police officers. In P. Hupe, M. Hill and A. Buffat (Eds.), *Understanding street-level bureaucracy* (pp. 99-114). Bristol: Policy Press.

M

Maggetti, M. & Levi-Faur, D. (2013). Dealing with errors in QCA. *Political Research Quarterly* 66(1), 198-204.

Mallon K. (2006). *Renewable energy policy and politics: A handbook for decision making*. London: Earthscan.

Marsh, J.A., McCombs, J.S. & Martorell, F. (2010). How instructional coaches support data-driven decision making policy implementation and effects in Florida middle schools. *Educational Policy*, 24(6), 872-907.

Marsh, D. & McConnell, A. (2010). Towards a framework for establishing policy success. *Public Administration,* 88(2), 564-583.

Matland, R.E. (1995). Synthesizing the implementation literature: The ambiguity-conflict model of policy implementation. *Journal of Public Administration Research and Theory*, 5(2), 145-174.

Matlock, K.L., Goering, C.Z., Endacott, J., Collet, V.S., Denny, G.S., Jennings-David, J. & Wright, G.P. (2016). Teachers' views of the Common Core State Standards and its implementation. *Educational Review,* 68(3), 291-305.

Matzler, K. & Renzl, B. (2007). Assessing asymmetric effects in the formation of employee satisfaction. *Tourism Management*, 28(4), 1093-1103.

May, P.J. (2015). Implementation failures revisited: Policy regime perspectives. *Public Policy and Administration*, 30(3), 277-299.

May, D.R., Gilson, R. & Harter, L. (2004). The psychological conditions of meaningfulness, safety and availability and the engagement of the human spirit at work. *Journal of Occupational and Organizational Psychology*, 77(1), 11-37.

May, P.J. & Winter, S.C. (2009). Politicians, managers, and street-level bureaucrats: Influences on policy implementation. *Journal of Public Administration Research and Theory*, 19(3), 453-476.

Maynard-Moody, S.W. & Musheno, M.C. (2003). *Cops, teachers, counselors: Stories from the front lines of public service*. Ann Arbor: University of Michigan Press.

Maynard-Moody, S.W. & Musheno, M.C. (2012). Social equities and inequities in practice: Street-level workers as agents and pragmatists. *Public*

Administration Review, 72(1), 16-23.

Maynard-Moody, S.W. & Portillo, S. (2010). Street-level bureaucracy theory. In Robert Durant (Ed.), *Oxford handbook of American bureaucracy* (pp. 252-277). Oxford: Oxford University Press.

McDermott, A.M., Hamel, L.M., Steel, D., Flood, P.C. & McKee, L. (2015). Hybrid healthcare governance for improvement? Combining top-down and bottom-up approaches to public sector regulation. *Public Administration*, 93(2), 324-344.

McGregor, D. (1960). *The human side of enterprise*. New York: Wiley.

Medzihorsky, J., Oana, I.E., Quaranta, M. & Schneider, C.Q. (2016). Set-Methods: A package companion to set-theoretic methods for the social science*s. R* Package Version 2.0, http://cran.r-project.org/package=SetMethods

Meier, K. & O'Toole, L. (2002). Public management and organizational performance: The effect of managerial quality. *Journal of Policy Analysis and Management*, 21(4), 629-643.

Metselaar, E.E. (1997). *Assessing the willingness to change: Construction and validation of the DINAMO*. PhD dissertation. Amsterdam: VU Amsterdam.

Meyers, M.K., Glaser, B. & MacDonald, L.K. (1998). On the front lines of welfare delivery: Are workers implementing policy reforms? *Journal of Policy Analysis and Management,* 17(1), 1-22.

Meyers, M.K. & Vorsanger, S. (2003). Street-level bureaucrats and the implementation of public policy. In B.G. Peters & J. Pierre (Eds.), *Handbook of public administration* (pp. 245-255). Thousand Oaks, CA: Sage.

Mintrom, M. & Luetjens, J. (2017). Creating public value: Tightening connections between policy design and public management. *Policy Studies Journal*, 45(1), 170-190.

Misangyi, V.F., Greckhamer, T., Furnari, S., Fiss, P.C., Crilly, D. & Aguilera, R. (2017). Embracing causal complexity: The emergence of a neo-configurational perspective. *Journal of Management*, 43(1), 255-282.

Moynihan, D.P., Wright, B.E. & Pandey, S.K. (2012). Working within constraints: Can transformational leaders alter the experience of red tape? *International Public Management Journal*, 15(3), 315-335.

N

NOS (2016). Opnieuw grote demonstratie tegen plan van Poolse regering. *NOS*. Retrieved April 28, 2017, from http://nos.nl/artikel/2144012-opnieuw-grote-demonstratie-tegen-plan-van-poolse-regering.html

NOS (2018). Omstreden lerarenregister voorlopig van de baan. *NOS*. Retrieved September 16, 2018, from https://nos.nl/artikel/2235951-omstreden-lerarenregister-voorlopig-van-de-baan.html

Nunnally, J.C. & Bernstein, I.H. (1994). The assessment of reliability. *Psychometric Theory*, 3(1), 248-292.

Nutting, P.A., Crabtree, B.F., Miller, W.L., Stange, K.C., Stewart, E. & Jaen, C. (2011). Transforming physician practices to patient-centered medical homes: Lessons from the national demonstration project. *Health Affair*, 30(3), 439-445.

O

Oberfield, Z.W. (2010). Rule following and discretion at government's frontlines: Continuity and change during organization socialization. *Journal of Public Administration Research and Theory*, 20(4), 735-755.

Olsen, A.L. (2017). Responding to problems: actions are rewarded, regardless of the outcome. *Public Management Review,* 19(9), 1352-1364.

Onderwijsraad (2012). *Geregelde ruimte.* Den Haag: Onderwijsraad.

Onderwijsraad (2014). *Onderwijspolitiek na de commissie-Dijsselbloem.* The Hague: Onderwijsraad.

Open Science Collaboration (2015). Estimating the reproducibility of psychological science. *Science*, 349(6251), online view: doi: 10.1126/science.aac4716

Osborne, J.W. & Fitzpatrick, D.C. (2012). Replication analysis in exploratory factor analysis: What it is and why it makes your analysis better. *Practical Assessment, Research and Evaluation* 17(15), 1-8.

O'Sullivan, T. (2010). *Decision making in social work*. London: Palgrave Macmillan.

O'Toole, L.J. (2000). Research on policy implementation: Assessment and prospects. *Journal of Public Administration Research and Theory*, 10(2), 263-288.

O'Toole, L.J. (2004). The theory-practice issue in policy implementation research. *Public Administration*, 82(2), 309-329.

P

Park, H.H. & Rethemeyer, R.K. (2014). The politics of connections: Assessing the determinants of social structure in policy networks. *Journal of Public Administration Research and Theory,* 24(2), 349-379.

Parsons, W. 1995. *Public policy*. Eldershot, UK: Edwar Elgar.

Page, S.B., Stone, M.M., Bryson, J.M. & Crosby, B.C. (2015). Public value creation by cross-sector collaborations: A framework and challenges of assessment. *Public Administration,* 93(3), 715-732.

Persson, M., Esaiasson, P. & Gilljam, M. (2013). The effects of direct voting and deliberation on legitimacy beliefs: An experimental study on small group decision making. *European Political Science Review,* 5(3), 381–399.

Perry, J.L. (2016). Public administration needs to become a player in the ratings and rankings business. *Public*

Administration Review, 76(5), 697-698.

Peters, G.B. (1986). *American public policy: Promise and performance*. Chatham, NJ: Chatham House Publishers, Inc.

Pierson, P. (2000). Increasing returns, path dependence, and the study of politics. *American Political Science Review*, 94(2), 251-267.

Pijl, S.J. & Frissen, P.H.A. (2009). What policymakers can do to make education inclusive. *Educational Management Administration and Leadership*, 37(3), 366-377.

Pollitt, C. & Bouckaert, G. (2011). *Public management reform: A comparative analysis: New public management, governance, and the neo-Weberian state*. Oxford: Oxford University Press.

Polsby, N. (1984). *Political innovation in America*. New Haven, CT: Yale University Press.

Precey, M. (2015). Teacher stress levels in England 'soaring', data shows. *BBC News*. Retrieved April 25, 2017, from http://www.bbc.com/news/education-31921457

Pressman, J.L. & Wildavsky, A.B. (1984). *Implementation: How great expectations in Washington are dashed in Oakland*. Berkeley, CA: University of California Press.

R

Raab, J., Mannak, R.S. & Cambré, B. (2015). Combining structure, governance, and context: A configurational approach to network effectiveness. *Journal of Public Administration Research and Theory*, 25(2), 479-511.

Raaphorst, N. (2018a). How to prove, how to interpret and what to do? Uncertainty experiences of street-level tax officials. *Public Management Review*, 20(4), 485-502.

Raaphorst, N. (2018b). *Uncertainty in bureaucracy: Toward a sociological understanding of frontline decision making*. PhD dissertation. Rotterdam: Erasmus University Rotterdam.

Ragin, C.C. (1987). *The comparative method: Moving beyond qualitative and quantitative strategies.* Chicago/London: University of California Press.

Ragin, C.C. (2000). *Fuzzy-set social science*. Chicago/London: University of Chicago Press.

Randall, D.M. & Wolff, J.A. (1994). The time interval in the intention-behaviour relationship: Meta-analysis. *British Journal of Social Psychology*, 33(4), 405-418.

Rayner, J., Williams, H.M., Lawton, A. & Allinson, C.W. (2011). Public service ethos: Developing a generic measure. *Journal of Public Administration Research and Theory*, 21(1), 27-51.

Remler, D.K. & Van Ryzin, G.G. (2011). *Research methods in practice: Strategies for description and causation*. Thousand Oaks, CA: SAGE Publications.

Revelle, W. (2015). Psych: Procedures for personality and psychological

research. R package version 1.5.8, Northwestern University, http://www.personality-project.org/r/psych-manual.pdf.

Riccucci, N.M. (2005). Street-level bureaucrats and intrastate variation in the implementation of temporary assistance for needy families policies. *Journal of Public Administration Research and Theory,* 15(1), 89-111.

Riccucci, N.M., Van Ryzin, G.G. & Lavena, C.F. (2014). Representative bureaucracy in policing: Does it increase perceived legitimacy? *Journal of Public Administration Research and Theory,* 24(3), 537-551.

Rosseel, Y. (2012). Lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, 48(2), 1-36.

S

Sabatier, P.A. (1986). Top-down and bottom-up approaches to implementation research: A critical analysis and suggested synthesis. *Journal of Public Policy*, 6(1), 21-48.

Sabatier, P.A. & Mazmanian, D. (1980). The implementation of public policy: A framework of analysis. *Policy Studies Journal*, 8(4), 538-560.

Sachau, D.A. (2007). Resurrecting the motivation-hygiene theory: Herzberg and the positive psychology movement. *Human Resource Development Review*, 6(4), 377-393.

Saetren, H. (2005). Facts and myths about research on public policy implementation: Out-of-fashion, allegedly dead, but still very much alive and relevant. *The Policy Studies Journal,* 33(4), 559-582.

Samuelson, W. & Zeckhauser, R. (1988). Status quo bias in decision making. *Journal of Risk and Uncertainty*, 1(7), 7-59.

Sager, F., Thomann, E., Zollinger, C., van der Heiden, N. & Mavrot, C. (2014). Street-level bureaucrats and new modes of governance: How conflicting roles affect the implementation of the Swiss ordinance on veterinary medicinal products. *Public Management Review*, 16(4), 481-502.

Samuelson, W. & Zeckhauser. R. (1988). Status quo bias in decision making. *Journal of Risk and Uncertainty*, 1(7), 7-59.

Sandfort, J.R. (2000). Moving beyond discretion and outcomes: Examining public management from the front lines of the welfare system. *Journal of Public Administration Research and Theory*, 10(4), 729-756.

Sandfort, J. & Moulton, S. (2015). *Effective implementation in practice: Integrating public policy and management.* San Francisco: Jossey-Bass.

Schott, C., Van Kleef, D.D. & Steen, T.P.S. (2018). The combined impact of professional role identity and public service motivation on decision-making in dilemma situations. *International Review of Administrative Sciences*, 84(1), 21-41.

Schreiber, J.B., Nora, A., Stage, F.K., Barlow, E.A. & King, J. (2006). Reporting structural equation modeling and confirmatory factor analysis results:

A review. *Journal of Educational Research*, 99(6), 323-337.

Schneider, C.Q. & Wagemann, C. (2012). *Set-theoretic methods for the social sciences. A guide to qualitative comparative analysis*. New York: Cambridge University Press.

Schwoerer, C.E., May, D.R., Hollensbe, E.C. & Mencl, J. (2005). General and specific self-efficacy in the context of a training intervention to enhance performance expectancy. *Human Resource Development Quarterly*, 16(1), 111-129.

Seeman, M. (1959). On the meaning of alienation. *American Sociological Review*, 24(6), 783-791.

semTools Contributors (2016). SemTools: Useful tools for structural equation modeling. R package version 0.4-14, https://CRAN.R-project.org/package=semTools.

Sharma, U., Forlin, C., Loreman, T. & Earle, C. (2006). Pre-service teachers' attitudes, concerns and sentiments about inclusive education: An international comparison of novice pre-service teachers. *International Journal of Special Education,* 21(2), 80-93.

Sheeran, P. & Orbell, S. (1998). Do intentions predict condom use? Meta-analysis and examination of six moderator variables. *British Journal of Social Psychology*, 37(2), 231-250.

Sheskin, D.J. (2007). *Handbook of parametric and nonparametric statistical procedures*. London/New York: Chapman & Hall/CRC.

Siciliano, M.D., Moolenaar, N.M., Daly, A.J. & Liou, Y. (2017). A cognitive perspective on policy implementation: Reform beliefs, sensemaking, and social networks. *Public Administration Review,* 77(6), 889-901.

Sijtsma, K. (2009). On the use, the misuse, and the very limited usefulness of Cronbach's alpha. *Psychometrika*, 74(1), 107-120.

Simon, H.A. (1955). A behavioral model of rational choice. *The Quarterly Journal of Economics,* 69(1), 99-118.

Skaaning, S. (2011). Assessing the robustness of crisp-set and fuzzy-set QCA results. *Sociological Methods and Research*, 40(2), 391-408.

Smith, G.T., McCarthy, D.M. & Anderson, K.G. (2000). On the sins of short-form development. *Psychological Assessment*, 12(1), 102-111.

Sommer Harrits, G. & Østergaard Møller, M. (2014). Prevention at the front line: How home nurses, pedagogues, and teachers transform public worry into decisions on special efforts. *Public Management Review*, 16(4), 447-480.

Stazyk, E.C. & Davis, R.S. (2015). Taking the 'high road': Does public service motivation alter ethical decision making processes? *Public Administration*, 93(3), 627-645.

Steeh, C. (1981). Trends in nonresponse rates. *Public Opinion Quarterly*, 45(1), 40-57.

T

Teodoro, M.P. (2014). When professionals lead: Executive management,

normative isomorphism, and policy implementation. *Journal of Public Administration Research and Theory*, 24(4), 983-1004.

Tesler, M. (2015). Priming predispositions and changing policy positions: An account of when mass opinion is primed or changed. *American Journal of Political Scienc*e, 59(4), 806-824.

Thelen, K. (2004). *How institutions evolve: The political economy of skills in Germany, Britain, the United States, and Japan.* New York: Cambridge University Press.

Thiem, A. (2014). Navigating the complexities of qualitative comparative analysis case numbers, necessity relations, and model ambiguities. *Evaluation Review*, 38(6), 487-513.

The Guardian (2017). Brazil army takes over state's security as 100 killed amid police strike. *The Guardian.* Retrieved November 18, 2017, from https://www.theguardian.com/world/2017/feb/09/brazil-police-strike-deaths-espirito-santo-vitoria

Thomann, E. (2015). Is output performance all about the resources? A fuzzy-set qualitative comparative analysis of street-level bureaucrats in Switzerland. *Public Administration*, 93(1), 177-194.

Thomann, E., Hupe, P.L. & Sager, F. (2018a). Serving many masters: Public accountability in private policy implementation. *Governance*, 31(2), 299-319.

Thomann, E., Lieberherr, E. & Ingold, K. (2016). Torn between state and market: Private policy implementation and conflicting institutional logics. *Policy and Society*, 35(1), 57-69.

Thomann, E. & Maggetti. M. (2017). Designing research with Qualitative Comparative Analysis (QCA): Approaches, challenges and tools. *Sociological Methods & Research*, early online view: doi:10.1177/0049124117729700

Thomann, E., Van Engen, N.A.M. & Tummers, L.G. (2018b). The necessity of discretion: A behavioral evaluation of bottom-up implementation theory. *Journal of Public Administration Research and* Theory, early online view: doi:10.1093/jopart/muyo24

Thomas, W. (1928). *The child in America: Behavior problems and programs.* New York: A.A. Knopf.

Tjeenk Willink, H.D. (2017). Eindverslag informateur Tjeenk Willink. *Bureau woordvoering kabinetsformatie.* Retrieved October 5, 2017, from https://www.kabinetsformatie2017.nl/documenten/verslagen/2017/06/27/eindverslag-informateur-tjeenk-willink

Triandafyllidou, A. (2003). Immigration policy implementation in italy: Organisational culture, identity processes and labour market control. *Journal of Ethnic and Migration Studies*, 29(2), 257-297.

Tummers, L.G. (2011). Explaining the willingness of public professionals to implement new policies: A policy alienation framework. *International Review of Administrative Sciences* 77(3), 555-581.

Tummers, L.G. (2012). Policy alienation of pubic professionals: The construct and its measurement. *Public Administration Review,* 72(4), 516-525.

Tummers, L.G. & Bekkers, V.J.J.M. (2014). Policy implementation, street-level bureaucracy, and the importance of discretion. *Public Management Review,* 16(4), 527-547.

Tummers, L.G., Bekkers, V.J.J.M. & Steijn, A.J. (2009). Policy alienation of public professionals. *Public Management Review,* 11(5), 685-706.

Tummers, L.G., Bekkers, V., Vink, E. & Musheno, M. (2015). Coping during public service delivery: A Conceptualization and systematic review of the literature. *Journal of Public Administration Research and Theory*, 25(4), 1099-1126.

Tummers, L.G. & Knies, E. (2016). Measuring public leadership: Developing scales for four key public leadership roles. *Public Administration,* 94(2), 433-451.

Tummers, L.G., Steijn, A.J. & Bekkers, V.J.J.M. (2012). Explaining the willingness of frontline workers to implement public policies: Content, context, and personality characteristics. *Public Administration,* 90(3), 716-736.

Tweede Kamer (2008). *Tijd voor onderwijs: parlementair onderzoek onderwijsvernieuwingen.* Tweede Kamer, vergaderjaar 2007–2008, Kamerstuk 31007, nr. 6.

Tyler, T.R. (2006). Psychological perspectives on legitimacy and legitimation. *Annual Review of Psychology,* 57, 375–400.

V

Van der Toorn, J., Tyler, T.R. & Jost, J.T. (2011). More than fair: Outcome dependence, system justification, and the perceived legitimacy of authority figures. *Journal of Experimental Social Psychology,* 47(1), 127–138.

Van der Voet, J., Steijn, A.J. & Kuipers, B.S. (2017). What's in it for others? The relationship between prosocial motivation and commitment to change among youth care professionals. *Public Management Review,* 19(4), 443-462.

Van der Voet, J. & Vermeeren, B. (2017). Change management in hard times: Can change management mitigate the negative relationship between cutbacks and the Organizational Commitment and Work Engagement of Public Sector Employees? *American Review of Public Administration*, 47(2), 230-252.

Van Engen, N.A.M. (2017). A short measure of general policy alienation: Scale development using a 10-step procedure. *Public Administration*, 95(2), 512-526.

Van Engen, N.A.M. (2019). Quantitative street-level bureaucracy research. In P. Hupe (Ed.), *Research handbook on street-level bureaucracy: The ground floor of government in context.* Cheltenham: Edward Elgar Publishing.

Van Engen, N.A.M., Tummers, L.G., Bekkers, V.J.J.M. & Steijn, A.J. (2016). Bringing history in: Policy accumulation and general policy alienation. *Public Management Review*, 18(7), 1085-1106.

Van Gunsteren, H.R. (1976). *The quest for control.* London: Wiley.

Van Haperen, T. (2015). Rekenfalen. *Website Ton van Haperen*. Retrieved August 19, 2018, from http://tonvanhaperen.com/index.php?option=com_content&view=article&id=366:resultaten-en-doelstellingen&catid=36:actueel&Itemid=1

VanLandingham, G., & Silloway, T. (2016). Bridging the gap between evidence and policy makers: A case study of the Pew-MacArthur results first initiative. *Public Administration Review,* 76(4), 542-546.

Van Loon, N.M., Leisink, P.L.M., Knies, E. & Brewer, G.A. (2016). Red tape: Developing and validating a new job-centered measure. *Public Administration Review*, 76(4), 662-673.

Van Meter, D.S. & Van Horn, C.E. (1975). The policy implementation process: A conceptual framework. *Administration & Society*, 6(4), 445-488.

Vandenabeele, W. (2008). Development of a public service motivation measurement scale: Corroborating and extending Perry's measurement instrument. *International Public Management Journal*, 11(1), 143-167.

Veronesi, G. & Keasey, K. (2015). Patient and public participation in the English NHS: An assessment of experimental implementation processes. *Public Management Review*, 17(4), 543-564.

Vigoda-Gadot, E. & Beeri, I. (2012). Change-oriented organizational citizenship behavior in public administration: The power of leadership and the cost of organizational politics. *Journal of Public Administration Research and Theory*, 22(3), 573-596.

W

Wagemann, C., Buche, J. & Siewert, M.B. (2016). QCA and business research: Work in progress or a consolidated agenda? *Journal of Business Research*, 69(7), 2531-2540.

Wallner, J. (2008). Legitimacy and public policy: Seeing beyond effectiveness, efficiency, and performance. *Policy Studies Journal* 36(3), 421-443.

Wanberg, C.R. & Banas, J.T. (2000). Predictors and outcomes of openness to changes in a reorganizing workplace. *Journal of Applied Psychology*, 85(1), 132-142.

Werts, A.B. & Brewer, C.A. (2015). Reframing the study of policy implementation: Lived experience as politics. *Educational Policy,* 29(1), 206-229.

White, W., Lunnan, A., Nybakk, E. & Kulisic, B. (2013). The role of governments in renewable energy: The importance of policy consistency. *Biomass and Bioenergy,* 57, 97-105.

Wildavsky, A.B. (1979). *Speaking truth to power: The art and craft of policy analysis*. Boston, MA: Little-Brown.

Wirth, R.J. & Edwards, M.C. (2007). Item factor analysis: Current approaches

and future directions. *Psychological Methods*, 12(1), 58-79.

Wright, B.E., Christensen, R.K. & Pandey, S.K. (2013). Measuring public service motivation: Exploring the equivalence of existing global measures. *International Public Management Journal*, 16(2), 197-223.

Y

You, S., Kim, A.Y. & Lim, S.A. (2017). Job satisfaction among secondary school teachers in Korea: Effects of teachers' sense of efficacy and school culture. *Educational Management Administration & Leadership,* 45(2), 284-297.

Young, L. (2015). Teacher dispute is about more than pay. *Seattle Times*. Retrieved November 18, 2017, from https://www.seattletimes.com/opinion/letters-to-the-editor/teacher-dispute-is-about-more-than-pay/

Young, T. & Lewis, W.D. (2015). Educational policy implementation revisited. *Educational Policy*, 29(1), 3-17.

Yukl, G.A. & Becker, W.S. (2012). Effective empowerment in organizations. *Organization Management Journal*, 3(3), 210-231.

APPENDIX I: GENERAL POLICY ALIENATION MEASURES

This Appendix includes the original and short measure of general policy alienation. Five-point Likert scales were used with all the items.

General policy alienation measure (26-item measure)

Table I.1 Overview of general policy alienation items

Dimension		General policy alienation item
Strategic powerlessness	1	In my opinion, <u>professionals</u> have too little power to influence government <u>policies</u>
	2	We, as <u>professionals,</u> are completely powerless during the introduction of government <u>policies</u>
	3	<u>Professionals</u> cannot influence the development of policies at the national level (Minister and Ministry of <u>X,</u> National Government)
	4	On a national level, <u>professionals</u> can influence how <u>policies</u> are set up (R)
	5	<u>Professionals</u>, through their professional associations, actively help in drawing up the design of government <u>policies</u> (R)
	6	Politicians, during the design of <u>policies</u>, do not listen to <u>professionals</u> at all
Tactical powerlessness	7	In my <u>organization</u>, it is especially <u>professionals</u> who decide how government <u>policies</u> are implemented (R)
	8	In my <u>organization</u>, <u>professionals</u> – through working groups or meetings – take part in decisions on executing government <u>policies</u> (R)
	9	The management of my <u>organization</u> should involve <u>professionals</u> far more in the execution of government <u>policies</u>
	10	<u>Professionals</u> are not listened to during the introduction of government <u>policies</u> in my <u>organization</u>
	11	In my <u>organization</u>, <u>professionals</u> take part in conversations regarding the execution of government <u>policies</u> (R)
	12	I and my fellow <u>professionals</u> are completely powerless during the introduction of government <u>policies</u> in my <u>organization</u>
Operational powerlessness	13	Generally, I have freedom to decide how to use government <u>policies</u> (R)
	14	Generally, when working with government <u>policies,</u> I can be in keeping with <u>clients</u>' needs (R)
	15	Generally, working with government <u>policies</u> feels like a harness in which I cannot easily move
	16	Generally, when working with government <u>policies</u>, I have to adhere to tight procedures

	17	Generally, government <u>policies</u> allow me to sufficiently tailor them to the needs of my <u>clients</u>
	18	Generally, government <u>policies</u> allow me to make my own judgments (R)
Societal meaninglessness	19	In general, I think that government <u>policy</u> in the long term will lead to <u>socially relevant goal A</u> (R)
	20	In general, I think that government <u>policy</u> in the short term will lead to <u>socially relevant goal A</u> (R)
	21	In general, I think that government <u>policy</u> has already led to <u>socially relevant goal A</u> (R)
	22	Overall, I think that government <u>policy</u> leads to <u>socially relevant goal A</u> (R)
Client meaninglessness	23	In general, government <u>policy</u> enables me to better solve the problems of my <u>clients</u> (R)
	24	In general, government <u>policy</u> contributes to the welfare of my <u>clients</u> (R)
	25	In general, government <u>policy</u> enables me to help <u>clients</u> more efficiently (R)
	26	Overall, I think government <u>policy</u> is ultimately favorable for my <u>clients</u> (R)

Note: In the present study, the general terms (underlined) are replaced by specifics: <u>professionals</u> by teachers, <u>X</u> by Education, <u>policy(ies)</u> by government education policy(ies), <u>organization</u> by school, <u>clients</u> by students, <u>policy</u> by education policy, <u>socially relevant goal A</u> by higher educational quality

Short measure of general policy alienation (5-item measure)

Table I.2 Overview of general policy alienation items

Item	Template	Present study
1	<u>Professionals</u> cannot influence the development of <u>policies</u> at the national level (Minister and Ministry of <u>X</u>, national government)	School leaders and teachers cannot influence the development of education policies at the national level (Minister and Ministry of Education, national government)
2	Generally, I have freedom to decide how to use government <u>policies</u> (R)	Generally, I have freedom to decide how to use government education policies
3	Overall, I think that government <u>policy</u> leads to <u>socially relevant goal A</u> (R)	Overall, I think that government education policy leads to higher educational quality
4	In general, I think that government <u>policy</u> in the long term will lead to <u>socially relevant goal A</u> (R)	In general, I think that government education policy in the long term will lead to higher educational quality
5	In general, government <u>policy</u> enables me to better solve the problems of my <u>clients</u> (R)	In general, government education policy enables me to better solve the problems of my students

APPENDIX II: 10-STEP PROCEDURE TO DEVELOP SHORT MEASURES

This Appendix provides an overview of the 10-step procedure we used to develop a short measure of general policy alienation. This procedure is mostly based on guidelines and suggestions of DeVellis (2003) and Smith *et al.* (2000).

Table II.1 10-step procedure to develop short measures

Steps	
Preparatory stage	
1	Only develop a short measure of a sufficiently validated original measure
2	Determine clearly what you want to measure
3	Select item pool from original measure
4	Determine whether it is necessary to develop additional items
5	Determine whether it is necessary to change the format for measurement
6	Evaluate face validity: Review item pool with experts
7	Include proposed short measure in a new survey questionnaire
Analysis stage (evaluate proposed measure)	
8	Show internal consistency reliability
9	Show construct validity a) Convergent b) Discriminant
10	Determine final measure
Optional	Repeat steps 4-10 if the analysis stage does not provide satisfactory or easy to interpret results

APPENDIX III: EXPERIMENT (TRANSLATED FROM DUTCH TO ENGLISH)

Upon accepting the invitation to participate, respondents were randomly assigned one of the two policies and, subsequently, one of the two possible response options.

Policy 1

Please imagine that Sander Dekker, the current Dutch Secretary of State for Education, decided in 2016 that schools, from school year 2016-2017 onwards, should receive additional funding for professional development. Schools are free to decide how to spend these funds to improve the quality of education (for instance on courses, advanced electronic equipment or supplementary educational support). This is because research has shown that it is necessary that educational staff (school leaders and teachers) professionalize. One of the reasons why professional development was stagnating was that there was insufficient funding available for schools.

In 2017, a new government is inaugurated. The new Secretary of State for Education is [name[4]].

One of the first debates that [name] has in the House of Representatives is about the budget for professional development in the education sector. During this debate, a Member of the House of Representatives states that:

"I believe that schools should not receive a fixed professional development budget that they can spend how they like. I feel it is the government's task to decide where professional development is most strongly needed and which programs should be offered to and financed for schools. We cannot leave this to the schools' discretion. Therefore, I propose that you immediately stop offering these funds to schools."

---

4   We used common Dutch names; not the name of a real politician.

Next, the new Secretary of State responds:

<u>Response 1(indicating policy consistency)</u>
"Chairman, it has been proposed to abolish the professional development budget. However, this measure was only introduced by my predecessor in the last school year. Therefore, we do not yet know whether this measure will have the desired effect. Hence, it seems illogical to already abolish it. The policy will therefore be maintained."

<u>Response 2 (indicating policy inconsistency)</u>
"Chairman, it has been proposed to abolish the professional development budget. However, this measure was only introduced by my predecessor in the last school year. Therefore, we do not yet know whether this measure will have the desired effect. Nevertheless, I agree that it seems preferable to abolish it. The policy will therefore not be maintained."

Policy 2

Please imagine that Sander Dekker, the current Dutch Secretary of State for Education, decided in 2016 that a fixed percentage of schools, from school year 2016-2017, should be [type A[5]] schools. This measure applies to both new and existing secondary schools. This is because research by the Dutch Inspectorate of Education has shown that education inequality is increasing. One of the causes is the increase in the number of [type B] schools and a decrease in the number of [type A] schools.

In 2017 a new government is inaugurated. The new Secretary of State for Education is [name].

---

5   Type A schools offer all three levels of secondary education (ranging from vocational training to pre-university education); type B schools only offer one level of secondary education. This implies that type B schools will be less diverse and that it will be harder for students to switch between levels, an ability which is particularly relevant in terms of educational opportunities for students at the lowest level.

One of the first debates that [name] has in the House of Representatives is about the equality measure. During this debate, a Member of the House of Representatives states that:

"I believe that schools should be free to decide what type of school they want to be. I feel that the current measure restricts the freedom of education. We can leave this to the schools' discretion. Therefore, I propose you immediately stop this equality measure."

Next, the new Secretary of State responds:

Response 1 (indicating policy consistency)
"Chairman, it has been proposed to abolish the equality measure. However, this measure was only introduced by my predecessor in the last school year. Therefore, we do not yet know whether this measure will have the desired effect. Hence, it seems illogical to already abolish it. The policy will therefore be maintained."

Response 2 (indicating policy inconsistency)
"Chairman, it has been proposed to abolish the equality measure. However, this measure was only introduced by my predecessor in the last school year. Therefore, we do not yet know whether this measure will have the desired effect. Nevertheless, I agree that it seems preferable to abolish it. The policy will therefore not be maintained."

APPENDIX IV: OVERVIEW OF ALL OTHER MEASURES USED

This Appendix complements Appendix I and includes all the other measures used in this study. Please note that templates are used in some measures (underlined words). Templates allow researchers to adapt items to their specific situation by replacing general phrases with more specific ones: ones that fit the context of their research. All items are formatted as five-point Likert scales, unless otherwise stated.

Discretion (i.e., operational powerfulness) (Van Engen *et al.*, 2016)

1. Generally, I have freedom to decide how to use government policies
2. Generally, when working with government policies, I can be in keeping with clients' needs
3. Generally, working with government policies feels like a harness in which I cannot easily move (R)
4. Generally, government policies allow me to sufficiently tailor them to the needs of my clients

Implementation willingness (Tummers, 2012, based on Metselaar, 1997)

1. I try to convince colleagues of the benefits that government policies will bring
2. I reduce resistance among colleagues regarding government policies
3. I make time to implement government policies
4. I make an effort to implement government policies successfully

Legitimacy (cf. De Fine Licht, 2014)

1. What do you think of the decision of the new Secretary of State?
2. How willing are you to accept the decision of the new Secretary of State?
3. How likely do you think it is that you will protest against the decision of the new Secretary of State? (R)

Policy alienation (Tummers, 2012)

Strategic powerlessness

1. In my opinion, professionals had too little power to influence the policy

2. We <u>professionals</u> were completely powerless during the introduction of the <u>policy</u>

3. <u>Professionals</u> could not at all influence the development of the <u>policy</u> at the national level (Minister and Ministry of <u>X</u>, National Government)

4. On a national level, <u>professionals</u> could influence how the <u>policy</u> was set up (R)

5. <u>Professionals</u>, through their professional associations, actively helped to think with the design of the <u>policy</u> (R)

6. Politicians did not, during the design of the <u>policy</u>, listen to the <u>professionals</u> at all

Tactical powerlessness

1. In my <u>organization</u>, especially <u>professionals</u> could decide how the <u>policy</u> was to be implemented (R)

2. In my <u>organization</u>, <u>professionals</u> have, through working groups or meetings, taken part in decisions on the execution of the <u>policy</u> (R)

3. The management of my <u>organization</u> should have involved the <u>professionals</u> far more in the execution of the <u>policy</u>

4. <u>Professionals</u> were not listened to over the introduction of the <u>policy</u> in my <u>organization</u>

5. In my <u>organization</u>, <u>professionals</u> could take part in discussions regarding the execution of the <u>policy</u> (R)

6. I and my fellow <u>professionals</u> were completely powerless in the introduction of the <u>policy</u> in my <u>organization</u>

Operational powerlessness (discretion)

1. I have freedom to decide how to use the <u>policy</u> (R)

2. While working with the <u>policy</u>, I can be in keeping with the <u>client's</u> needs (R)

3. Working with the <u>policy</u> feels like I am in a harness in which I cannot easily move

4. When I work with the <u>policy</u>, I have to adhere to tight procedures

5. While working with the <u>policy</u>, I cannot sufficiently tailor it to the needs of my <u>clients</u>

6. While working with the <u>policy</u>, I can make my own judgments (R)

Societal meaninglessness

1. I think that the <u>policy</u>, in the long term, will lead to <u>goal X</u> (R)
2. I think that the <u>policy</u>, in the short term, will lead to <u>goal X</u> (R)
3. I think that the <u>policy</u> has already led to <u>goal X</u>(R)
4. Overall, I think that the <u>policy</u> leads to <u>goal X</u> (R)

Please note that in chapter 4, only item 1, 2 and 4 were used.

Client meaninglessness

1. With the <u>policy</u> I can better solve the problems of my <u>clients</u> (R)
2. The <u>policy</u> is contributing to the welfare of my <u>clients</u> (R)
3. Because of the <u>policy</u>, I can help <u>clients</u> more efficiently than before (R)
4. I think that the <u>policy</u> is ultimately favorable for my <u>clients</u> (R)

Policy consistency (Van Engen *et al.*, 2016)

To what extent do you have the impression that policy by the Ministry of <u>X</u>

1. … is consistent
2. … focuses on the long term
3. … is driven by 'the issues of the day' (R)
4. … expresses long-term vision

Transformational leadership (Carless *et al.*, 2000)

My leader….

1. …communicates a clear and positive vision of the future [vision]
2. …treats staff as individuals, supports and encourages their development [staff development]
3. …gives encouragement and recognition to staff [supportive leadership]
4. …fosters trust, involvement and cooperation among team members [empowerment]
5. …encourages thinking about problems in new ways and questions assumptions [innovative thinking]
6. …is clear about his/her values and practises what he/she preaches [lead by example]

7. …instills pride and respect in others and inspires me by being highly competent [charisma]

Trust in government (cf. European Social Survey)

How much trust do you have in the following institutions/authorities?

1. The Minister and Secretary of State for <u>X</u>
2. The Ministry of <u>X</u>
3. Politics in general

Please note we used a four-point Likert scale to measure trust in government.

Control variables

Age
Year of birth [xxxx]

Gender
Are you…[male, female, other]

Tenure
How long have you been working as a teacher/school leader [xx]

SUMMARY

How Previous Policy Experiences Affect the Frontline: Understanding implementation success and failure through a general policy alienation lens

Introduction

Teachers, healthcare workers and police, as well as other public employees working at the frontline of public service delivery, are often confronted with new policy programs that usually lead to new rules and regulations that have to be implemented. The fact that these 'frontline workers' are often confronted with new policies is, of course, in itself, not problematic – democratically elected governments have the mandate to do so. However, it can influence the way in which frontline workers perform their tasks, as grown practices may be challenged – over and over again. In this study, we aimed to capture this process and its effect with the concept of general policy alienation, thereby drawing on the policy alienation work by Tummers, Bekkers and Steijn.

It is not bold to state that policymakers are highly dependent on frontline workers for the successful implementation and - perhaps a bit more controversially - the formulation of their policies. Repeated research has shown that actual behavior during policy implementation does not necessarily align with policymakers' ambitions in as much as that a frontline worker might 'shirk or sabotage'. In this study, we provide several examples of frontline workers' responding just like this, for example, by starting a strike or by simply ignoring new policies - 'bend over, here it comes again'. This is problematic because such actions, ultimately, might result in a diminished legitimacy of the government. It can cause tension and conflicts and result in suboptimal circumstances for society at large. This is particularly the case if public funds are invested in the formulation and implementation of government policies that, apparently, are not supported by frontline workers. Ultimately, this impedes the improvement of public service provision, as this is more likely to be achieved if actors operating at different levels of the system collaborate.

Therefore, it is important to understand how frontline workers perceive and implement these policies. Surprisingly, the experiences of frontline workers with

new policies are often studied in isolation and ignore the fact that these policies are never developed in a vacuum. These experiences always have a history because they build on earlier experiences with other related policies. We describe this process as policy accumulation, i.e. the continuous aggregation of policies that follow each other. This suggests that frontline workers have a certain policy predisposition and a general attitude towards government policies, which we refer to as 'general policy alienation'. Drawing on change management studies – where terms such as change fatigue and change cynicism are used – we argue that frontline workers, when confronted with policies they perceive as being introduced too frequently and too inconsistently, could become indifferent to whatever new policy is introduced and result in them viewing new policies as just the new 'political flavor of the month'.

The policy alienation framework was developed to analyze frontline workers' experiences with specific government policies systematically and coherently. Policy alienation is defined as a "cognitive state of psychological disconnection from the policy…". However, as we argue above, it is not only relevant to investigate how frontline workers experience specific policies, but also how they identify with government policies in general. The current policy alienation framework does not take into account the effect of the accumulation of previous experiences and it does not allow for the evaluation of general government experiences. Therefore, we investigated how we could further develop the framework to take this effect into account. This allowed us to investigate what factors influenced general policy alienation, as well as the influence of general policy alienation on implementation willingness. In doing so, we introduced the term general policy alienation.

In summary, the main goal of this study was to analyze whether and to what extent frontline workers experienced general policy alienation, but also why this was the case and what the implications might be for policy implementation. Therefore, the central research question of this thesis is:

How can the general policy alienation of frontline workers be conceptualized and measured, what are its causes and what is its influence on implementation willingness?

Case study: Dutch secondary education

Our study investigated the education sector. Research has indicated that school leaders and teachers in this sector have experienced many problems with national education policies. This is particularly relevant because they play a crucial role in delivering services. The specific case we studied was the Dutch secondary education sector. This case was relevant for three reasons. First, the sector had experienced many problems in recent decades as a result of the reshuffling of authority and responsibilities between the ministerial and the school levels. Second, the sector was characterized by numerous policy changes. Third, research has shown that many school leaders and teachers are critical of government-initiated reform. This made it a suitable case to investigate policy implementation challenges, the consequences of policy accumulation and antecedents and effects of general policy alienation in-depth. This is illustrated by the following quote from a school leader we interviewed: "*Annoying are the continuous change and additional tasks. A perfect example is the introduction of an obligatory social internship for all high school students. We embraced this policy, invested many of our funds in it, and really saw its added value. Then, the obligation was withdrawn, as well as the government funding. This, in my opinion, rewards schools that act negligent. As a result of this, when new policies are introduced by the government and you do want to implement them loyally you eventually start thinking: Why would we?*"

Main research findings

In **chapter 2**, we report how we adapted the policy alienation framework to allow for the assessment of frontline workers' general experiences with government policies. So far, the policy alienation framework has been used mainly to analyze frontline workers' experiences with single policies. However, a complete picture is only provided if we take the effect of general policy experiences into account and if we allow for the fact that policies are not developed in a vacuum, but rather during a process of policy accumulation. Therefore, we focus on general policy alienation. Do frontline workers have the impression that, in general, they can influence the shaping of government policies? Furthermore, do they have the impression that government policies, overall, are meaningful and add value for society as a whole and their own clients? We define general policy alienation as an overall disconnect from government policies; in other words, a lack of com-

mitment, enthusiasm and identification with policies. As with specific policy alienation, general policy alienation can be conceptualized using five dimensions. This is summarized in Table S.1.

Table S.1 Definition of general policy alienation: five dimensions

Dimension	Definition
Strategic powerlessness	The influence that frontline workers usually perceive themselves as having on decisions concerning the content of government policies as captured in rules and regulations.
Tactical powerlessness	The influence that frontline workers usually perceive themselves as having on decisions concerning the way (new) government policies are implemented within their organization.
Operational powerlessness	The influence that frontline workers usually perceive themselves as having during the actual implementation of government policies.
Societal meaninglessness	The perception of frontline workers concerning the added value of government policy to socially relevant goals.
Client meaninglessness	The perception of frontline workers concerning the added value of government policy for their own clients.

Furthermore, using a sample of 1.096 Dutch teachers, we developed a valid and reliable five dimension, 26-item measure of general policy alienation. An overview of this scale is provided in Appendix I. Our analyses show that the average score on general policy alienation is 3.46 on a 1 to 5 scale and that teachers score particularly high on the strategic powerlessness and the two meaninglessness dimensions. The results show that general policy alienation was positively related to the alienation towards a specific policy program and negatively related to policy consistency, implementation willingness, and transformational leadership. Furthermore, we found that policy consistency was strongly related with strategic powerlessness and meaninglessness. Finally, we found that the meaningfulness of policy (for both society and clients), apparently, had more influence on implementation willingness than perceived powerfulness.

In **chapter 3**, we report how we developed a short measure of general policy alienation, based on the original 26-item measure. This approach produces important benefits. First, the reduced data requirement for a short measure saves survey time that a researcher can now use to measure additional variables. Second, many items that tap into the same concept may introduce fatigue or induce

boredom among respondents. Third, a short measure is more likely to be applied in other fields of public administration, where it could form a relevant contextual or explanatory variable. In order to establish a valid and reliable measure, we developed a 10-step approach for short-scale development that may be useful to researchers wishing to develop short measures themselves. An overview of this procedure is provided in Appendix II. Using three samples of, in total, 2.470 school leaders and teachers, we developed a short, five-item measure of general policy alienation modeled as a second-order construct. An overview of this scale is provided in Appendix I. Our analyses indicate the short version of the scale is a good substitute for the long scale, as we show that, as with the long version of the scale, general policy alienation was negatively related to the perceived meaningfulness of specific policy programs, policy consistency and implementation willingness. Furthermore, we found a negative relationship with trust in government and – perhaps a bit surprising – that general policy alienation appears to be unrelated to school characteristics, such as size and type of school.

In **chapter 4**, we report on a survey experiment we conducted to evaluate the effect of policy consistency on frontline workers perceptions of government and government policies. We argue that policy accumulation is, in itself, not problematic, but becomes problematic when policies are perceived as inconsistent. This assumption is based on literature that emphasizes the relevance of rational policymaking and the status quo biases of frontline workers. Specifically, we investigated how policy consistency – the degree to which policies are stable and constant over time – may improve policy meaningfulness (the added perceived added value of policies) and government legitimacy (how justified and appropriate government action is). From the literature, it is known that discretion during policy implementation is important for frontline workers. Therefore, we also investigated how the effect of policy consistency was influenced by perceived discretion. Using a sample of 779 teachers, we showed that policy consistency had, as expected, a positive effect on meaningfulness and, particularly, legitimacy. Furthermore, we found that this effect was moderated by the degree of autonomy frontline workers experience. Finally, we show that policy consistency was not a one-size-fits-all-solution, as the relationship between consistency,

meaningfulness and legitimacy is influenced by the type of policy respondents are confronted with.

In **chapter 5**, we report how we investigated the motivating effect of powerfulness (i.e. 'perceived discretion') for implementation willingness. We argue that, despite the fact that the topic of discretion continues to be hotly debated in policy design and policy implementation, there has been little systematic research into how the experience of having discretion motivates frontline workers to implement a policy. In this specific study, in contrast to the other studies in this thesis, we relied on an asymmetric explanation of frontline workers motivation. We hypothesized the existence of two complementary interpretations of the motivational role of powerfulness. The first assumed that powerfulness is quasi-necessary, although, on its own, not sufficient to motivate employees. The second interpretation was that powerfulness is only motivating when frontline workers perceive the policy to be implemented as meaningful. Using two samples of 1.317 healthcare workers and 1.096 teachers and large-N set-theoretic analysis, we show that powerfulness is – as expected – a quasi-necessary condition for high implementation willingness. Furthermore, we found mixed evidence for the assumption that a lack of powerfulness was quasi-sufficient for low implementation willingness. Finally, we show that, in combination with policy meaningfulness, powerfulness was quasi-sufficient for high implementation willingness. In other words: feeling powerful is necessary for high implementation willingness, but it is only sufficient in explaining implementation willingness when in combination with meaningfulness. These results underscore the potential added value of studying the relation between policy alienation and implementation willingness, both from an asymmetric and a symmetric perspective.

Why is it relevant to investigate this?

Our studies have contributed to the theoretical knowledge on policy implementation and street-level bureaucracy. Although the literature recognizes the important role of frontline workers for policy implementation, public administration and management research has and still tends to marginalize the perspectives and experiences of those who enact the policy in practice and, particularly, the micro-level (psychological) underpinnings of this. This is peculiar, since policy

implementation is sometimes complex, contradictory, and still one of the main challenges for civil servants worldwide.

Our first contribution was that we introduced the new concept of general policy alienation and, thereby, acknowledge that frontline workers bring with them a history of government policy (changes), and, hence, general ideas about their effectiveness, legitimacy, and meaningfulness. We emphasized that frontline workers were and are not neutral implementers. By studying policy experiences in relation to their historical context, we extend the theoretical work on policy accumulation and related concepts, such as policy succession and institutional layering.

Our second contribution is that we provided a straightforward suggestion on how to improve frontline workers policy perceptions. It has been noted "the cataloguing of failures when putting policies in place has been the hallmark of implementation studies since the 1970s". We, on the other hand, proposed and showed that policy consistency may contribute to improved policy perceptions of frontline workers by relying on work emphasizing the benefits of a rational policymaking perspective, as well as mostly political research on status quo bias of civil servants.

We aimed to contribute methodologically, by adopting relatively new and innovative methodological approaches. First, we conducted quantitative street-level bureaucracy, implementation research. Traditionally, this type of research has been quite qualitatively oriented. In this regard, it is not surprising that it has been noted "making the study of street-level bureaucracy both generalizable and comparative is an issue in its own right". Our quantitative approach allowed for the large-scale testing of relevant theories and assumptions. For instance, we adopted an experimental approach in investigating the effects of policy consistency on frontline workers' policy perceptions. Although experiments, by definition, manipulate situations (i.e. situations are not 'real', which limits ecological validity), they also allow us to isolate and explore causal effects of interest in ways that other methods cannot. By doing so, we contributed to the emerging tradition of a ´behavioral public administration´, which operates at the cross-point of public administration and psychology. It is relevant that street-level bureaucracy research is part of this development.

Second, we developed two valid and reliable measures: a measure of general policy alienation (26 items) and a short measure of general policy alienation (5 items). It has been noted that the field of public administration lags behind other social sciences in the development of measurement scales. We proposed and used systematic procedures that we hope can help other researchers develop scales themselves. This can help street-level bureaucracy and implementation research in making inferences that are also comparable across studies and contexts.

From a more practical point of view, our research may help national and local policymakers – basically, all (government) actors involved in policy implementation – better understand why the implementation of their policies succeeds or fails. This will apply particularly to those sectors when governments are heavily dependent on frontline workers to achieve their intended policy changes. This includes the healthcare sector (where governments rely on medical doctors and nurses), safety sector (where government rely on police and military), and the education sector (where governments rely on school board governors, school leaders and teachers).

Second, we developed measurement scales for use by policymakers or applied policy researchers to comprehensively (long, 26-item measure) or efficiently (short, 5-item measure) analyze how frontline workers experience government policies, also over time. If changes occur, or frontline workers indicate they feel extremely alienated, this may call for the introduction of appropriate interventions. In this way, this monitoring may help to improve the policy implementation process. By taking the experiences of frontline workers seriously, this may be a helpful tool to improve the relationship between policymakers and policy implementers.

Third, the practical recommendations we postulated – based on our research results – provide straightforward suggestions for civil servants and public managers who aim to strengthen their policy implementation. For instance, based on experimental evidence, we underscored the importance of policy consistency for successful policy implementation. Furthermore, we showed that the overwhelming majority of those frontline workers with high implementation willingness also experienced high levels of discretion. This should encourage practitioners interested in implementation, policy and organization design to move beyond

the question as to whether frontline workers should be granted discretion. Based on this study, the more salient question seems to be how to make best use of frontline workers' discretion to encourage behavior that eventually contributes to the achievement of policy goals.

Conclusions

In concluding this study, we can see we investigated the relationship between general policy alienation and policy implementation. Our research findings strongly suggest that governments should pay attention to the fact that frontline workers are not neutral implementers. They bring with them a history of government (policy) changes that affects how they perceive and behave in relation to new policies.

Therefore, we believe it is relevant to distinguish between specific and general policy experiences, as this will help to better understand policy implementation failure – or, more ideally, success. Our results underscored the potential added value of policy consistency for frontline workers. We showed that policy consistency increases perceived policy meaningfulness, as well as government legitimacy. Therefore, it might be a useful strategy for government to improve their policy implementation, via more committed implementers. Finally, we shed light on the ongoing discussion on discretion in policy design and implementation research. Our results showed that the majority of frontline workers who felt powerful had high implementation willingness. We urge scholars and practitioners to move from the question whether frontline workers should be granted discretion as to how to best make use of frontline workers' discretion instead.

Our findings have contributed a new angle to the continuing debate within policy implementation and street-level bureaucracy research on how to account for the complex, messy and, sometimes, contradictory implementation of public policies.

SAMENVATTING

Hoe eerdere beleidservaringen de frontlinie beïnvloeden: Beleidsimplementatie en -mislukking vanuit een algemeen beleidsvervreemdingsperspectief

Introductie

Leraren, verpleegkundigen en de politie, die in de 'frontlinie' verantwoordelijk zijn voor het leveren van publieke diensten, worden regelmatig geconfronteerd met nieuw overheidsbeleid. Dit beleid heeft over het algemeen als gevolg dat ze nieuwe regels, voorschriften en processen moeten implementeren. Dit is op zichzelf staand natuurlijk niet problematisch: democratisch verkozen overheden hebben het volste recht nieuw beleid te introduceren – mits zij hiervoor natuurlijk voldoende steun in het parlement vergaren. Echter, de continue introductie van nieuw beleid beïnvloedt wel de wijze waarop 'frontliniemedewerkers' hun werk doen, aangezien ingesleten gedachtepatronen en gedrag worden uitgedaagd of betwist en nieuw beleid over het algemeen om proactieve inzet van hun kant vraagt om de uitvoering tot een succes te maken.

Het is namelijk keer op keer aangetoond dat politici en beleidsmakers voor de implementatie van hun beleid sterk afhankelijk zijn frontliniemedewerkers. Maar het gedrag van frontliniemedewerkers bij beleidsimplementatie sluit niet per definitie aan bij de ambities van politici en beleidsmakers. In dit proefschrift zijn meerdere voorbeelden beschreven waarin frontliniemedewerkers beleid tegenwerken, door stakingen, door niet in de geest van beleid te handelen of beleid simpelweg niet uit voeren ('*bend over, here it comes again*'). Op deze manier kunnen ze grote uitdagingen creëren voor overheden die beleid willen veranderen en introduceren. Dit is problematisch, omdat deze acties uiteindelijk de legitimiteit van beleid kunnen aantasten, of resulteren in conflicten of onwenselijke maatschappelijke uitkomsten. Dit beïnvloedt de publieke dienstverlening negatief, aangezien bewezen is dat deze verbetert naarmate diverse actoren actief op verschillende niveaus beter samenwerken. Daarom is het belangrijk om nog beter te begrijpen hoe frontliniemedewerkers overheidsbeleid ervaren en implementeren.

Opvallend is dat eerder onderzoek naar de ervaringen van frontliniemedewerkers met overheidsbeleid vaak negeert dat overheidsbeleid niet ontwikkeld wordt in een vacuüm. Eerder is het zo dat de beleidservaringen van frontliniemedewerkers een 'verleden' hebben, omdat ze afhankelijk zijn van hun eerdere ervaringen met ander overheidsbeleid. We refereren aan dit proces als beleidsaccumulatie: de continue opeenstapeling van overheidsbeleid. Dit suggereert dat frontliniemedewerkers een bepaalde beleidspredispositie hebben: een meer positieve of negatieve basishouding jegens beleid. Geïnspireerd door managementstudies die 'verandermoeheid' en 'verandercynisme' onderzoeken, beargumenteren we dat frontliniemedewerkers die te vaak geconfronteerd worden met nieuw beleid dat ze bovendien als inconsequent en zwalkend ervaren, onverschillig en cynisch zijn richting nieuw beleid. Dat beleid beschouwen ze, in hun eigen woorden, simpelweg als het nieuwe politieke of ambtelijke stokpaardje.

Om de ervaringen van frontliniemedewerkers met specifiek overheidsbeleid te analyseren, ontwikkelden Tummers, Bekkers en Steijn het model van beleidsvervreemding. Beleidsvervreemding definiëren ze als "een cognitieve staat van psychologische ontkoppeling met het beleid...". Echter is het, zoals we hierboven beargumenteren, niet alleen relevant de specifieke ervaringen, maar ook de algemene ervaringen van frontliniemedewerkers met overheidsbeleid te analyseren. Bij het overslaan van dit laatste, is het lastiger te begrijpen waarom de implementatie van nieuw beleid mislukt. Het door Tummers *et al.* ontwikkelde model richt zich echter niet op die algemene ervaringen. Daarom onderzoeken we in dit proefschrift of en hoe het mogelijk is dit model zo door te ontwikkelen dat dit wel kan. We introduceren hierbij de term en het model algemene beleidsvervreemding om te refereren aan de algemene ervaringen met beleid die centraal staan. Dit maakt het mogelijk te onderzoeken welke factoren algemene beleidsvervreemding beïnvloeden en welk effect algemene beleidsvervreemding heeft op de implementatiebereidheid van frontliniemedewerkers en hun ervaringen met specifiek beleid. Samenvattend is de hoofdvraag van dit onderzoek:

Hoe kunnen we de algemene beleidsvervreemding van frontliniemedewerkers conceptualiseren en meten, wat zijn de oorzaken van algemene beleidsvervreemding en welk effect heeft het op hun implementatiebereidheid?

Onderzoekscasus: Nederlandse voortgezet onderwijs

De casus die in dit onderzoek centraal staat is het Nederlandse onderwijs. Onderzoek toont aan dat schoolleiders en leraren in deze sector in het heden en verleden diverse problemen met nationaal overheidsbeleid hebben ervaren. Dit is bijzonder relevant, omdat schoolleiders en leraren een cruciale rol spelen bij het leveren van kwalitatief hoogstaand onderwijs. De specifieke onderzoekscasus is het Nederlandse voortgezet onderwijs, die om drie redenen relevant is. Ten eerste hebben actoren in deze sector diverse problemen ervaren als gevolg van het regelmatig herschikken van verantwoordelijkheden tussen het ministerie en het niveau van de school en/of het schoolbestuur. Ten tweede kenmerkt de sector zich door een vrij grote hoeveelheid aan beleidswijzigingen. Ten derde heeft onderzoek laten zien dat veel schoolleiders en leraren in het voortgezet onderwijs kritisch zijn op door de overheid geïnitieerde onderwijshervormingen. Dit maakt het een interessante casus om uitdagingen omtrent beleidsimplementatie, de consequenties van beleidsaccumulatie en oorzaken en effecten van algemene beleidsvervreemding te onderzoeken. De volgende quote afkomstig uit een interview met een schoolleider illustreert dit: *"Wat irritant is, zijn de continue wijzigingen en extra taken. Een perfect voorbeeld is de introductie van de maatschappelijke stage voor alle leerlingen in het voortgezet onderwijs. We omarmden dit beleid, investeerden er veel tijd en geld in en zagen echt de toegevoegde waarde. Vervolgens werd besloten dat de maatschappelijke stage niet verplicht was en waren er ook geen financiële middelen meer beschikbaar. Dit beloont in mijn optiek scholen die laks handelen. De consequentie hiervan is wel dat ik bij nieuw beleid van de overheid dat je in principe loyaal wil implementeren toch begin te denken: Waarom zou ik?"*

Kern van de onderzoeksbevindingen

In **hoofdstuk 2** beschrijven we hoe we het model van beleidsvervreemding zo aanpassen dat we ook de algemene ervaringen van frontliniemedewerkers ermee kunnen analyseren: het model van algemene beleidsvervreemding. Bij het originele model is het namelijk alleen mogelijk de specifieke ervaringen met overheidsbeleid te analyseren, terwijl we in dit proefschrift nu juist aantonen dat een compleet beeld alleen verkregen wordt indien we ook naar meer algemene ervaringen kijken. Kortom, we onderzoeken in hoeverre frontliniemedewerkers,

in ons geval schoolleiders en leraren, het idee hebben dat ze overheidsbeleid kunnen beïnvloeden en of ze over het algemeen de indruk hebben dat beleid van toegevoegde waarde is voor hun eigen cliënten (in dit geval: leerlingen) en de samenleving als geheel. We definiëren algemene beleidsvervreemding als "een algemene staat van psychologische ontkoppeling met overheidsbeleid", oftewel: een gebrek aan betrokkenheid, enthousiasme en identificatie met overheids-beleid. Ook laten we zien dat algemene beleidsvervreemding uit vijf dimensies bestaat. Een beknopte definitie van de vijf dimensies is weergegeven in Tabel S.2. Daarnaast ontwikkelen we op basis van data van 1.096 leraren een gevalideerd meetinstrument van 26 items. Het meetinstrument is weergegeven in Appendix I. Uit de analyses blijkt dat de gemiddelde score op algemene beleidsvervreem-ding 3.46 is op een schaal van 1 tot 5 en dat leraren met name hoog scoren op de dimensies strategische machteloosheid en zinloosheid voor de samenleving en eigen cliënten. We laten zien dat algemene beleidsvervreemding – in lijn met onze hypotheses – positief samenhangt met specifieke ervaringen met beleid en negatief samenhangt met beleidsconsistentie, implementatiebereidheid en transformationeel leiderschap. Daarnaast valt op dat strategische machteloos-heid en zinloosheid sterk samenhangen met beleidsconsistentie. En, tot slot, dat de zinvolheid van beleid (voor samenleving én cliënt) meer invloed heeft op implementatiebereidheid dan gepercipieerde invloed op beleid.

Tabel S.2 Definities van de vijf dimensies van algemene beleidsvervreemding

Dimensie	Definitie
Strategische machteloosheid	De mate van ervaren invloed van frontliniemedewerkers op de inhoud van het beleid, zoals vastgesteld in wet- en regelgeving
Tactische machteloosheid	De mate van ervaren invloed van frontliniemedewerkers op de manier waarop hun organisatie het beleid implementeert
Operationele machteloosheid	De mate van ervaren invloed van frontliniemedewerkers op de manier waarop zij zelf het beleid uitvoeren
Zinloosheid voor de samenleving	De ervaring van de frontliniemedewerkers over de toegevoegde waarde van het beleid aan belangrijke doelen voor de samenleving
Zinloosheid voor de eigen cliënten	De ervaring van de frontliniemedewerkers over de toegevoegde waarde van het beleid voor hun eigen cliënten

In **hoofdstuk 3** ontwikkelen we een korte schaal van algemene beleidsvervreemding, die vijf items telt in plaats van 26 items. Dit is om drie redenen relevant. Ten eerste creëert dit ruimte in surveyonderzoek om andere relevante variabelen te meten. Ten tweede vermoeien veel vragen die min of meer over hetzelfde onderwerp of concept gaan respondenten snel. Ten derde is de kans groter dat onderzoekers deze schaal meenemen in hun eigen onderzoek, bijvoorbeeld als onafhankelijke, contextuele of controlevariabele. We ontwerpen een systematische procedure van 10 stappen om een korte schaal te ontwikkelen, die hopelijk ook voor andere onderzoekers die korte schalen willen ontwikkelen een instrument van toegevoegde waarde is. Deze procedure is weergeven in Appendix II. Voor het ontwikkelen van de schaal maken we gebruik van drie datasets van in totaal 2.470 schoolleiders en leraren. De analyses tonen aan dat een 5-item schaal, gemodelleerd als een tweede-orde-construct het meest geschikt is. Onze analyses laten zien dat de korte schaal een goede vervanger is van de lange schaal, aangezien we ook bij deze analyses vinden dat algemene beleidsvervreemding negatief samenhangt met de zinvolheid van specifiek overheidsbeleid, beleidsconsistentie en implementatiebereidheid. Daarnaast laten we ook de negatieve samenhang met vertrouwen in de overheid zien en – in zekere zin verrassend – dat algemene beleidsvervreemding niet lijkt samen te hangen met schoolspecifieke kenmerken zoals als aantal leerlingen en het type school (openbaar versus bijzonder onderwijs).

In **hoofdstuk 4** beschrijven we het ontwerp en de resultaten van een surveyexperiment dat we uitvoerden om de effecten van beleidsconsistentie nader te onderzoeken. Hier beargumenteren we dat beleidsaccumulatie op zichzelf niet problematisch is, maar dit pas wordt indien beleid wordt gezien als inconsistent. Specifiek onderzoeken we hoe beleidsconsistentie – de mate waarin beleid stabiel en constant is – kan bijdragen aan meer zinvolheid van beleid (beleid met toegevoegde waarde) en meer legitimiteit van de overheid (hoe gerechtvaardigd en passend overheidsoptreden is). Dit doen we op basis van een experiment waarbij we het type beleid waarmee respondenten geconfronteerd worden manipuleren (beleid dat zich richt op professionele ontwikkeling van leraren versus beleid dat zich richt op gelijke onderwijskansen), en ook de mate van consistentie van het handelen van de bewindspersoon in casu (consistent of

inconsistent, oftewel: stopt een nieuw bewindspersoon beleid van zijn voorganger of niet). Op basis van onderzoeksdata van 779 leraren tonen we aan dat beleidsconsistentie, conform verwachting, een positief effect heeft op zinvolheid en, in het bijzonder, legitimiteit. We laten ook zien dat dit effect afhankelijk is van de mate waarin ze autonomie in hun werkzaamheden ervaren. Tot slot laten we zien dat beleidsconsistentie geen 'one-size-fits-all'-oplossing is, aangezien de relatie tussen consistentie en zinvolheid en legitimiteit beïnvloed wordt door het type beleid waarmee respondenten geconfronteerd worden.

In **hoofdstuk 5** onderzoeken we het motiverende effect van ervaren invloed op beleid ('powerfulness'; discretie) op implementatiebereidheid. We doen dit omdat ondanks het continue debat over het belang van discretie, er weinig systematisch onderzoek is naar het motiverende effect hiervan. We adopteren hierbij, in tegenstelling tot de andere studies in dit proefschrift, een asymmetrische aanpak. In deze specifieke casus bedoelen we hiermee dat we een ander effect verwachten van de aanwezig- en afwezigheid van ervaren invloed op implementatiebereidheid. Op basis hiervan formuleren we twee complementaire interpretaties. Ten eerste dat ervaren invloed een quasi-noodzakelijke (rand)voorwaarde voor implementatiebereidheid is, maar op zichzelf staand niet voldoende om frontliniemedewerkers te motiveren. Ten tweede dat ervaren invloed alleen een motiverend effect heeft indien frontliniemedewerkers het te implementeren beleid als zinvol ervaren. Op basis van datasets van zowel leraren als medewerkers in het gezondheidsdomein (waaronder psychologen en psychiaters) tonen we aan dat ervaren invloed inderdaad een quasi-noodzakelijke (rand)voorwaarde is voor implementatiebereidheid. Voor de assumptie dat een gebrek aan invloed leidt tot lage implementatiebereidheid vinden we slechts deels bewijs. Tot slot laten we zien dat ervaren invloed in combinatie met zinvol beleid een quasi-voldoende voorwaarde is voor hoge implementatiebereidheid. Oftewel: ervaren invloed is een randvoorwaarde voor implementatiebereidheid, maar uiteindelijk is de zinvolheid van het beleid daarbij doorslaggevend. Deze bevindingen tonen aan dat het van toegevoegde waarde kan zijn de relatie tussen (algemene) beleidsvervreemding zowel vanuit asymmetrisch als symmetrisch perspectief te bestuderen.

Waarom is het relevant dit te onderzoeken?

Met dit onderzoek dragen we bij aan de kennis over beleidsimplementatie en de 'street-level bureaucratie'. Ondanks het feit dat de literatuur erkent dat frontlinie-medewerkers een belangrijke rol spelen bij beleidsimplementatie, marginaliseert management- en beleidsonderzoek hun percepties en ervaringen. Vaak staat bovendien de zogenaamde 'top-down' aanpak centraal. Dit is opvallend, omdat beleidsimplementatie complex, tegenstrijdig en een belangrijke – misschien wel de belangrijkste – uitdaging is voor politici en beleidsmedewerkers. In onze studie staat daarom het perspectief van schoolleiders en leraren centraal. Onze eerste bijdrage is dat we het nieuwe concept van algemene beleidsvervreemding introduceren en daarmee erkennen dat frontliniemedewerkers een historie van beleidswijzigingen en daarmee ideeën over de effectiviteit, legitimiteit en zinvol-heid van overheidsbeleid met zich meebrengen. We benadrukken hiermee dat frontliniemedewerkers geen neutrale uitvoerders van overheidsbeleid zijn. Dit is wellicht niet opzienbarend, maar wel iets wat in de praktijk van beleid maken nog een ondergeschoven kind lijkt. Door deze historie van beleidservaringen te bestuderen, breiden we het theoretische werk over beleidsaccumulatie en gerelateerde concepten als institutionele gelaagdheid uit. Onze tweede bijdrage is dat we niet alleen analyseren welke factoren bijdragen aan implementatie-mislukking, maar ook welke factoren bijdragen aan implementatiesucces. Dit in tegenstelling tot de meerderheid van implementatiestudies, die zich sinds de jaren '70 vooral lijkt te kenmerken door een complete focus op alles wat er mis gaat. Ons onderzoek daarentegen laat zien dat meer beleidsconsistentie bijdraagt aan positieve beleidspercepties, daarbij leunend op studies die een rationeel perspectief op beleid maken adopteren, waaronder politicologisch onderzoek naar de status quo bias van ambtenaren.

Methodologisch dragen we bij door relatief nieuwe onderzoekstechnieken te gebruiken. Ten eerste voeren we kwantitatief onderzoek uit naar de street-level bureaucratie. Traditioneel is dit een vrij kwalitatief georiënteerd onderzoeksveld, wat onder andere het trekken van generaliseerbare conclusies lastig maakt. Onze kwantitatieve aanpak maakt het mogelijk implementatietheorieën en -assump-ties op grotere schaal te testen. Dit doen we bijvoorbeeld door het uitvoeren van een surveyexperiment. Dit draagt bij aan de ontwikkeling van de 'gedragsbe-stuurskunde', die opereert op het snijvlak van bestuurskunde en psychologie. Het

is noodzakelijk dat ook onderzoek naar de street-level bureaucratie onderdeel is van deze ontwikkeling. Daarnaast ontwikkelen we twee schalen om ons concept te meten. Ondanks het feit dat het aantal meetinstrumenten ontwikkelt door bestuurskundigen stijgt, wordt er helaas nog te vaak 'ad-hoc' een meetinstrument ontwikkeld. Dit maakt het lastig bevindingen van verschillende studies met elkaar te vergelijken. Onze meetinstrumenten zijn daarnaast ook vrij beschikbaar voor politici, beleidsmakers en praktijkonderzoekers die de ervaringen van frontliniemedewerkers met algemeen en specifiek overheidsbeleid onderzoeken. Meer begrip van deze ervaringen lijkt op basis van dit onderzoek een randvoorwaarde voor beter begrip van implementatiesucces en –mislukking. Dit is met name relevant voor die publieke sectoren waarbij overheden voor het bereiken van maatschappelijke verandering afhankelijk zijn van frontliniemedewerkers, zoals op het gebied van zorg, onderwijs en veiligheid.

Conclusie

We hebben in dit onderzoek de relatie tussen algemene beleidsvervreemding en beleidsimplementatie onderzocht. Op basis van onze onderzoeksbevindingen, stellen we dat het relevant is als overheden meer aandacht besteden aan het feit dat frontliniemedewerkers alles behalve neutrale uitvoerders van overheidsbeleid zijn. Daarom stellen we dat het relevant is onderscheid te maken tussen specifieke en algemene ervaringen met beleid, omdat dit zorgt voor een beter begrip van implementatiesuccessen of –mislukkingen. Dit doen we door het verfijnen van het bestaande beleidsvervreemdingsraamwerk. Daarnaast onderstrepen onze resultaten het belang van beleidsconsistentie. We laten zien dat beleidsconsistentie gemiddeld genomen bijdraagt aan zinvoller beleid en meer legitimiteit van de overheid, maar dat dit wel afhankelijk is van autonomie en het type beleid wat de overheid continueert. Ons onderzoek suggereert daarmee dat oog hebben voor consistentie een zinvolle strategie kan zijn voor overheden om hun beleidsimplementatie te versterken doordat de uitvoerders meer toegewijd en betrokken zijn. Tot slot dragen we bij aan de voortdurende discussie over het belang van discretie voor frontliniemedewerkers bij het ontwerpen en uitvoeren van beleid. Onze bevindingen laten zien dat de meerderheid van de frontliniemedewerkers die discretie en invloed ervaart een hoge implementatiebereid heeft. We roepen academici en de praktijk daarom op om na te denken hoe deze

inherente discretie slimmer in te zetten bij de ontwikkeling en implementatie van beleid. Concluderend dragen we met dit onderzoek bij aan het debat over beleidsimplementatie en street-level bureaucratie en de vraag hoe om te gaan met de complexe, rommelige en soms tegenstrijdige implementatie van overheidsbeleid.

ABOUT THE AUTHOR

Nadine van Engen (Utrecht, 1987) studied Sociology and International Business Administration at the Warsaw School of Economics and the University of Groningen, where she graduated in 2012 (cum laude). She did an internship at the department of Sociology, under supervision of prof.dr. René Veenstra, where she realized doing a PhD would be a great way to start her career.

Nadine's research focuses on public policy and public management, adopting a behavioral, quantitative approach. She specializes in policy implementation, street-level bureaucracy, and scale development. She published in, among others, Public Administration, Public Management Review and Journal of Public Administration Research and Theory. She also published non-academic articles about her research in, for instance, De Volkskrant, Idee, and Platform Overheid, as well as edited a volume on good, meaningful and effective education policy. She taught several courses for both Public Administration and Pedagogical and Educational Sciences students, including Economics, Organization & Management, and Multi-level Governance. From 2013 until 2015, she was a board member at PNN, the Dutch national PhD representative body. In 2016, she completed the Training Program of the Netherlands Institute of Government. In 2017, she won the Erasmus Graduate School Best Paper Award and in 2018 she, together with prof.dr. Lars Tummers, received a NRO/NWO grant to study teacher leadership.

Currently, from May 2017 onwards, Nadine is working at the Inspectorate of the Budget at the Dutch Ministry of Finance.

Within policy implementation and street-level bureaucracy research, there is a continuing debate on how to account for the complex, messy, and, sometimes, contradictory implementation of public policies. Frontline workers, such as teachers, doctors, and police, as a consequence of their discretion, play a crucial role in successful policy implementation and the efficient, effective, and responsive delivery of public services. However, in practice, frontline workers do not always act in line with the ambitions of politicians and policymakers. In doing so, they can create major difficulties for governments, democratically mandated to change policy, in making their (new) policies a success.

This book quantitatively investigates how previous policy experiences affect the frontline. Contrary to previous work that mostly has focused on the experiences of frontline workers with specific policies, this study takes into account that policies are not developed in a vacuum. Rather, they build upon each other; a process described as policy accumulation. Based on research on change cynicism, this book argues that frontline workers - as a consequence of continuous policy changes that are sometimes perceived as inconsistent and too frequent - could become indifferent to whatever new policy is introduced, viewing new policies as just the 'new political flavor of the month'.

As such, this study opens the way for a better understanding of policy implementation failure and success, by showing how frontline workers' previous policy experiences shape how they perceive and enact new policies. In doing so, this study underscores the importance of alignment between policy makers and policy implementers for successful implementation. The author illustrates the dynamics surrounding policy implementation in the Dutch education sector by combining implementation and street-level bureaucracy theory with detailed empirical analysis.