Developing a short measure of general policy alienation

This chapter has been published as 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.

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

	Mean		Standard deviation		Minimum		Maximum	
Item	**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 unrelated 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 capturing 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.