LEGITIMACY OF POLICY INTERVENTIONS

A scale for measurement

June 2020

Rick Westland, s2671816
Universiteit Leiden, Public Administration (MSc)
Specialization Economics and Governance
Supervised by Dr. Hendrik Vrijburg

Commissioned by Netherlands Environmental Assessment Agency (PBL)

Table of contents

Acknowledgements

1 Introduction

Nowadays, climate change and the following climate policy is one of the main political issues in the Netherlands, as well as in other countries. The Paris Agreement (United Nations, 2015) has set the mark high for environmental goals for 2030, 2050 and beyond. Support for the goal to fight climate change is predominantly present, but translating these objectives into actual policy interventions remains challenging. Vringer and Carabain (2020) show this in their recent research paper, where they conclude that support for policy goals does not directly lead to support for their related interventions.

However, policy interventions are required to reach the goals set. While pronounced support for policy interventions is hard to achieve, the acceptance due to the policy's value to common interest is more attainable. This is what we could perceive as legitimacy, following a fundamental definition of legitimacy by Weber (1978). He defines legitimacy as the acceptance of exercised power. Legitimacy is based on many dimensions (Bokhorst, 2014), as it covers different aspects, such as effectiveness and efficiency, as well as more abstract values such as fairness and transparency (Tyler, 2006; Curtin & Meijer, 2006).

When evaluating policy interventions, legitimacy can be of great importance. First of all, the valuation of legitimacy of a policy intervention by the public is important for its effectiveness and efficiency (Wallner, 2008). When interventions are not deemed legitimate, they might not be followed by the public. Making sure that policy interventions are aligned with stakeholders, as well as the public, and that they are procedurally well thought out, can make for high compliance rates. Second of all, the many aspects that make up legitimacy apart from effectiveness and efficiency can have a positive impact on societal welfare. Aspects such as transparency and fairness are included in welfare analysis by PBL, SCP and CPB (2017). According to their definition, legitimacy of policy interventions is therefore to some extent welfare enhancing.

Legitimacy is a commonly used term in research and scientific articles. However, in politics and policy evaluation the term is not used as frequently. This could be due to the lack of clarity of the concept. A clear operationalization is required to give purpose to the concept for policy evaluation. In the literature legitimacy is usually only defined in abstract terms. Vringer and Carabain (2020) made a first attempt to operationalize legitimacy by measuring perceptions on statements about interventions based on good governance criteria (UNESCAP, 2009). An actual measure of perceived legitimacy can improve the evaluation (ex-ante as well as ex-post) of policy interventions.

For the operationalization of perceived policy legitimacy of transition policy, Vringer and Carabain (2020) suggested using the perceptions of the public on statements related to "good governance criteria", as elaborated on by UNESCAP (2009). These good governance criteria are, for example, participation, inclusiveness, and effectiveness. This paper tries to use the good governance related criteria to create a scale that measures the perceived legitimacy of transition policy interventions. The research question this paper answers is: To what extent are good governance criteria appropriate to measure perceived legitimacy of policy interventions?

To answer this question, I chose the following sub-questions:

1. What is the importance of policy legitimacy?
2. Can good governance criteria be used to assess legitimacy of interventions as perceived by the public?
3. How can we weigh the good governance criteria in creating an overall score for perceived legitimacy of the intervention?

The first question is explored in Section 2 of this paper. Understanding the importance of policy legitimacy shows why the operationalization of – and a scale to measure perceived legitimacy – is relevant. Using different resources, I give a clear definition of policy legitimacy and present ideas about its importance.

The second question is important because I am trying to measure legitimacy through statements that relate to the good governance criteria. This implies that there is a relationship between good governance criteria and legitimacy (see also Vringer and Carabain, 2020). First, I explore literature about the relationship and similarities between legitimacy and good governance. Next, I make use of the same data set as used by Vringer and Carabain (2020), to explore how useful the good governance criteria are in quantifying legitimacy. Their survey was the Dutch energy transition, in which they queried multiple hypothetical policy interventions among citizens and business representatives.

The third question is also covered in the empirical part of the paper. As the goal is to create a scale that can measure legitimacy, the weighting of items has to become clear. Using a graded response model, I make an effort to create a scale to measure legitimacy. The outcomes of this model can identify the way to weight the different items on the scale.

This paper contributes in both the societal as well as the scientific field. First of all, the answer to the research question could be of contribution to policy evaluation. It might also help policy makers in constructing policy interventions which are deemed legitimate by the public. By making the factors that contribute to the legitimacy of an intervention explicit, this paper can shape the way policy interventions are made, as well as the way they are presented to the public. Scientifically, it can teach us more about how people perceive legitimacy of interventions. This research can add to the degree in which we can empirically establish legitimacy and therefore be of influence to the theoretical understanding of the concept. It is new in the respect that it attempts to quantify legitimacy using good governance criteria. Next to that, I also apply methodology that is not common within public administration research.

The following section includes the theoretical literature review. After that, I elaborate on the empirical research methods that has been used during the empirical analysis and on the data in the research design paragraph. Subsequently, I will present the results of the empirical analysis, and, lastly, I will discuss the results and draw conclusions.

2 Literature review

This section serves as a theoretical basis. First, I define legitimacy and explore the importance of that concept. Then, I explore what aspects of a policy can contribute to its legitimacy and look at the similarities between legitimacy and the good governance criteria. Lastly, the good governance criteria are discussed.

2.1 Legitimacy

2.1.1 Definition

Legitimacy is a widely discussed concept in political theory. Weber (1978) defines legitimacy as the acceptance of exercised power. Close to this idea of legitimacy are Hanberger (2003), Coicaud (2002) and Montenegro, de Wit and Iles (2016). The latter take a different approach, defining legitimacy even broader than the others. To them, legitimacy is people accepting 'something'. 'Something', in this case, is undefined and could be a policy or a regime. The Netherlands Institute for Social Research (SCP, 2017) more distinctly shows the difference between legitimacy and public support. The two concepts are closely related and often confused. Legitimacy is defined as 'the right to exercise power and the acknowledgement of that right by the public.' Public support is 'the direct agreement with certain policy options.'

The OECD takes a similar approach towards legitimacy as Weber (1978), Hanberger (2003) and Coicaud (2002). To them, legitimacy is the answer to "whether, how and why people accept a particular form of rule as being legitimate." (OECD, 2010, p.15). An addition they make is that people should regard the present policy satisfactory and believe it is the best option available.

According to Smoke (1994), most writers connect legitimacy directly to regimes or governments in total. The need to connect this to a specific policy is because of the democratic nature of our politics nowadays, and the implications this has. A policy in itself needs to be deemed legitimate in order to be effective, and does not implicitly become legitimate because the government that imposes it is deemed as such (Smoke, 1994). Matti (2009) argues that policy legitimacy has become more and more important, as modern states face a decrease in political trust (Hanberger, 2003). Therefore, policy needs to be legitimate by itself and can no longer be dependent of the legitimacy of the regime. He even goes as far as to argue that the process of legitimacy becomes less important, as the actors that present it are of less value. The content of the policy has to stand by itself.

For this paper, I use the following definition of policy legitimacy: the acceptance of a policy intervention out of common interest. The definitions mentioned before are used to the extent that I mostly focus on acceptance. To make a clear distinction with public support, I added the part about common interest. People who do not prefer the policy over other options, might still accept it, due to its value for common interest. In that case it would still be deemed a legitimate policy.

2.1.2 Importance of legitimacy

From a narrow perspective on welfare, policy legitimacy can have a positive impact on welfare due to its contributions to voluntary compliance (Scholz, 1984). High levels of legitimacy of a policy, can make for high numbers of voluntary compliance with a low amount of compliance costs. Wallner (2008) makes a similar argument. She stresses that a lack of legitimacy of implemented policies, can result in policy failure. Policy legitimacy is important because of the moral obligation people feel to follow a policy that is not enforced (Bokhorst 2014). From a social benefit analysis perspective, this has social gains in comparison to a policy that needs strong enforcement and therefore has high compliance costs (Nas, 2016; Bokhorst, 2014; Gribnau, 2009). Net benefit will be an efficiency gain, due to lower implementation costs, while the policy benefits remain unchanged.

From a social welfare analysis perspective, as described by PBL, SCP and CPB (2017), welfare includes additional values apart from efficiency. They argue that depending on context, many personal preferences can have a positive impact on welfare. Values such as transparency, equity, participation, and so on, can have considerable impact on overall welfare in society. So even when we do not consider the monetary, effectiveness, or efficiency gain that legitimacy can have for the policy, legitimate policy can still be welfare enhancing for society.

2.1.3 Aspects that enhance policy legitimacy

The previous paragraphs explored what legitimacy entails and why it is important. Important to know as well is what aspects of a policy intervention enhance its legitimacy. Tyler (2010; 2006) clearly links legitimacy to fairness, as he argues that people are most influenced by the process through which authority is exercised on them. Their values and beliefs about what is a fair procedure are majorly important in determining legitimacy. Grimes (2006) follows the same line of reasoning. She argues that collective decision can never be made with full agreement of every stakeholder or citizen, and we therefore need a different method to assure legitimacy. Creating consensus about the procedures to make collective decisions is easier than creating consensus for every collective decision. Therefore, institutions and procedures need to be in place in order to come to legitimate decisions. This emphasizes the role of a fair procedure for legitimacy.

Bokhorst (2014) acknowledges a legal dimension of legitimacy. In order for a policy to be legitimate, it has to be legal, that is, align with law and order (Jagers & Hammar, 2009). Even though this is more of a prerequisite, it is still important to take into account.

Curtin and Meijer (2006) point out transparency as an enhancing aspect of policy for legitimacy. The relationship between transparency and legitimacy is more complicated. Even though transparency can help the spread of legitimacy, by making information about the decision making process available to more citizens, it cannot by itself enhance legitimacy (Curtin & Meijer, 2006). That is, the content and procedure need to be good in order for transparency to be effective in creating more legitimacy (De Fine Licht, Naurin, Esaiasson, & Gilljam, 2014). Therefore, with policy legitimacy, it does not seem like transparency can have a big impact, as it can only strengthen the effects of the content and procedure of the policy on legitimacy.

2.2 Connecting legitimacy and good governance

The aspects I described that enhance the legitimacy of a policy are closely related to the good governance criteria that were described by UNESCAP (2009). In fact, Gribnau (2009) explicitly connects good governance and legitimacy. He argues that the relationship between state and citizen or state and company is two-sided. They are both reliant of each other. Therefore, legitimate policy takes into account the common values and expectations of the public. These common values and expectations of the public are represented by the values that good governance includes.

Keping (2011) makes a similar argument. He argues that good governance is such a broad and desirable political state that it includes all the conditions for legitimacy. For example, he uses good government (which is more focussed on values of bureaucracy and an effective administration) to show that there are other concepts that include important aspects of legitimacy, but that they are not sufficient. He even states that if legitimacy is perceived as the voluntary recognition and obeying of social order and authority, good governance can be seen as an equivalent of legitimacy.

Lastly, Vringer and Carabain (2020) connect legitimacy to good governance. They expect good governance to be a good operationalization of legitimacy, which can be useful when trying to measure perceived legitimacy. The following paragraph will serve as an explanation of what exactly makes up the good governance criteria.

2.2.1 Good governance criteria

Good governance criteria originate from financial aid to third-world countries. To decide whether a government is eligible to receive aid, they have to show their will and capacity to act according to the good governance criteria. Although no government fulfils all eight criteria (Graham, Amos & Plumptre, 2003), aid countries have to show some level of good governance. The criteria assure 'that corruption is minimized, the views of minorities are taken into account, and the voices of the most vulnerable in society are heard in decision-making' (UNESCAP, 2009).

As mentioned, there are eight criteria that make up good governance according to UNESCAP (2009):

- Participation. Both men and women have to be involved, either direct or represented in society. Next to that, it also includes freedom of speech and a well-organized civil society.
- Rule of law. A fair legal framework is required, which holds every citizen accountable equally. Human rights are specifically important.
- Transparency. Policy and implementation are according to rules and regulations. Information has to be accessible freely to those affected and is made comprehensive for the public.
- Responsiveness. A government has to respond to stakeholders within reasonable time.
- Consensus oriented. Mediation is needed in a society with many viewpoints, in order to create a society broad consensus on the best interest for society as a whole.
- Equity and inclusiveness. Everybody in society should feel they have a part to play and opportunity to improve their wellbeing.
- Effectiveness and efficiency. Good governance has to produce results that are needed, making good use of their resources. It also considers sustainability and the protection of the environment.
- Accountability. The government, civil society organizations and the private sector should be accountable for their actions towards the people that are affected by those actions. This is a key requirement for good governance.

There are different definitions of good governance, but in the end the main idea is the same (Grindle, 2012). Clear is that the good governance criteria should contribute to a stable political and economic system, while "bad governance is a problem that countries need to overcome" (Grindle, 2012, p. 2).

Gribnau (2009) approaches good governance through horizontal supervision. He claims that good governance is a recent development in democracies that has a way of managing society that is more fluid and solution based than its predecessors. The direct vertical control such as law should decrease and be replaced by more normative rules, based on trust and a solution-based approach.

3 Methods

I start with the research design. Then I explain which data I use for this paper. Subsequentially, the operationalization of the concept of legitimacy is presented. Lastly, I explain which statistical techniques have been utilized and why.

3.1 Research design

This paper uses a cross-sectional research design (Bryman, 2016). That means multiple cases are examined and a large number of quantifiable data is used. The big advantage of this type of design is the possibility to explore patterns of association through survey data (Bryman, 2016). These association patterns are exactly what I want to analyze, as I am looking for a way to operationalize and measure legitimacy, which is explicitly a question about the correlation and covariation between variables.

The main research question states that I am looking to measure perceived legitimacy through the good governance criteria. Perceived legitimacy cannot be measured directly by asking people how legitimate they think an intervention is, because they will likely answer whether they support the intervention. Therefore, perceived legitimacy has to be measured indirectly through other variables. Such an indirectly measured variable is what is called a latent variable. Vringer and Carabain (2020) propose the good governance criteria as a collection of variables through which I can measure legitimacy. The good governance criteria are handled as aspects that make up the legitimacy of the policy intervention. Therefore, a statistical model can help to calculate a score for the latent variable that is measured through the good governance criteria. Whether that score actually measures perceived legitimacy of the policy interventions is not assured by the empirical methods. Therefore, the latent variable will not be labelled 'perceived legitimacy' in the methods and most of the results section, because the exact meaning of the latent variable is still unclear. The items included in the analysis are based on the eight good governance criteria and have been queried in the survey by Vringer and Carabain (2020). The following paragraph will introduce the data that will be used.

3.2 Data

The data used is a survey conducted by Vringer and Carabain (2020). The survey was distributed among Dutch "citizens and company representatives to measure policy legitimacy" (Vringer & Carabain, 2020, p. 4). The focus of the survey was the Dutch energy transition, aiming to reduce CO_2 emissions by 49% in 2030. Four hypothetical policy interventions aimed at fighting climate change were presented to the respondents.

For every intervention the respondents were asked to value them on eight different statements that are related to the good governance criteria. These statements will be presented in the operationalization paragraph (3.3). For every intervention the respondent had to respond to these eight statements on a 5-point Likert scale, ranging from (1) Strongly disagree to (5) Strongly agree. Next to these main items of analysis, for every intervention the respondent was asked whether they thought the intervention should be implemented in the described manner, also on a 5-point scale ranging from 'yes, definitely' to 'no, definitely not'. Other interesting data included in the survey are background variables, such as gender, age, household income, education and voting behavior in the last national elections. Important to note is that these background variables, were not queried for the business representatives.

While Vringer and Carabain (2020) analyze only two of the four interventions due to practical considerations, such as the comparability and time scheme, the survey contains two more interventions. As this paper aims to create a scale for the legitimacy of interventions, it is interesting to include all different interventions available. The heterogeneity in interventions can deepen the understanding of the functions of items for the different interventions and whether it measures the same concept for the different interventions. If that would be the case, it would be more generalizable to other interventions.

The two interventions discussed by Vringer and Carabain (2020) are two very similar interventions. In both interventions, all households and companies will get an in-home display to monitor their electricity usage. In the first intervention households and companies will have to pay for this collectively, while in the second intervention energy suppliers have to pay for the in-home displays. The third and fourth intervention included in this paper are also quite similar and differ in roughly the same way as the first and second do. In both interventions the intention of the policy is to stimulate the circular economy. The third intervention uses a payment per kilogram for waste of households or companies. Households and companies that already have low amounts of waste and separate it well, will pay less, while households and companies with high amounts of waste that is not separated will pay more. In the fourth intervention companies are obliged to use recycled material in both their products and packaging, which may result in higher product prices. Table 3.1 shows the four interventions with short descriptions.

Table 3.1 Intervention description

Intervention number	Description
1	All households and companies get an in-home display. Paid for collectively.
2	All households and companies get an in-home display. Pair for by energy supplier.
3	Payment per kilogram of waste for households and companies.
4	Companies obliged to use recycled material in products and packaging

The survey was sent to citizens aged 18 and over, and representatives of companies. These representatives only include owners, CEO's, financial directors, financial managers and general managers. The survey was filled in by 2111 respondents. For a more extensive description about the data collection see Vringer and Carabain (2020).

3.2.1 Missing data

The eight main items I use for analysis were queried on a Likert scale, which is common in questionnaire research. Respondents were required to provide a response to each item, which means there is no missing data in the classical sense. However, an extra answer option 'don't know' was included, which brought some complications. This category cannot be placed in an ordinal position as the other answer options can. A 'don't know' response can be picked by somebody who strongly agrees but is unsure about that, while it can also be picked by somebody who strongly disagrees, but is unsure. Therefore, it offers us no information about the position of the respondent. The statistical method chosen in this paper, can exclude single variable observations. Therefore, I chose to regard 'don't know' responses as missing values, which led to a small loss of data. It could also lead to a bias in which only people that have a clear opinion are included in the sample.

3.3 Operationalization of legitimacy

As I am using the dataset from Vringer and Carabain (2020), I have to utilize part of their operationalization of legitimacy. The following operationalization has been drafted by Vringer and Carabain (2020):

"Judgement on several underlying aspects concerning the public support of the intervention to get more grip on the overall level of support. In this case, we used the good governance criteria and additional conditions for legitimacy as mentioned by The Netherlands Institute for Social Research (SCP, 2017). We asked respondents: To what extent do you agree or disagree with the following statements concerning the intervention?

1. This intervention helps to decrease climate change and pollution (**effective**)
2. The costs are limited (**efficient**)
3. The implementation of this intervention is in good hands (**consensus-oriented, responsive**)
4. The set-up of the intervention is in good hands (**consensus-oriented, responsive**)
5. This intervention is feasible (**effective**)
6. This intervention is straightforward (**transparent**)
7. This intervention is fair (**equitable and inclusive, following the rule of law**)
8. This intervention takes the situation of everybody into account (**participatory**)"

Some comments can be made regarding this operationalization. First of all, the accountability criterium that was included in the description of good governance criteria in the theoretical paragraph is not included in this operationalization. This could mean that the data lacks important information about policy legitimation. At the same time, some criteria are more emphasized and queued than others. For example, the good governance criterium 'effective and efficient' is queued in three different questions, while 'equitable and inclusive' and 'rule of law' are merged into one question. Especially the questions that should measure the same conduct could be dependent of each other through a different latent variable than the one hypothesized, which is legitimacy of transition policy. Therefore, not all items might be suitable to include in the model, because of their interdependence. Statistical analysis can determine whether this is the case or not. The influence of these issues are discussed in paragraph 6.2.

3.4 Statistical analysis

The analysis in this paper can be divided into two parts: the exploratory factor analysis and the graded response model. In order to keep the extensive analysis readable, the two methods are not discussed in this chapter, but are discussed in their own chapters. These chapters include a description of the method, as well as the results of that method. This paragraph includes the descriptive statistics.

As the labels assigned to the eight statements related to good governance are quite long and need to be included in tables and figures, Table 3.2 introduces the abbreviations of the eight statements for the following chapters.

Table 3.2 Assigned abbreviations to the eight included items

Item	Abbreviation
Effective (instrument helps)	Helps
Efficienct	Efficient
Consensus oriented, responsive (implementation)	C. R. Implementation
Consensus oriented, responsive (set-up)	C. R. Set-up
Effective (feasible)	Feasible
Transparent	Transparent
Equitable and inclusive, following the rule of law	Fair
Participatory	Participatory

3.4.1 Descriptive statistics and correlations

To get a better sense of the data, I made some descriptive statistics. Table 3.3 shows the means with standard deviation for the eight items, separated for the different interventions. The table also shows the number of observations, which is the total number of observations (2111) minus the 'don't know' answers, because they are treated as missing.

Table 3.3 Means for the eight most important variables for all four interventions

Items	Intervention 1 Mean (SD)	Obs.	Intervention 2 Mean (SD)	Obs.	Intervention 3 Mean (SD)	Obs.	Intervention 4 Mean (SD)	Obs.
Helps	3.166 (1.241)	1,988	3.342 (1.202)	2,000	3.499 (1.194)	2,016	3.853 (.995)	2,030
Efficient	3.066 (1.318)	1,976	3.260 (1.286)	1,894	2.922 (1.221)	1,873	2.880 (1.139)	1,860
C. R. Implementation	3.208 (1.295)	1,996	3.675 (1.171)	1,992	3.148 (1.323)	1,989	3.284 (1.151)	1,957
C. R. Set-up	3.401 (1.235)	2,029	3.725 (1.139)	2,038	3.614 (1.218)	2,060	3.543 (1.084)	2,023
Feasible	3.098 (1.265)	1,994	3.440 (1.210)	2,005	3.439 (1.347)	2,045	3.351 (1.149)	1,999
Transparent	2.781 (1.309)	1,994	3.285 (1.267)	1,976	2.909 (1.321)	2,028	2.923 (1.175)	1,990
Fair	3.592 (1.150)	2,002	3.448 (1.193)	2,003	3.138 (1.251)	1,995	3.012 (1.101)	1,971
Participatory	3.046 (1.233)	1,983	3.283 (1.208)	1,978	3.163 (1.230)	1,995	3.210 (1.147)	1,983

Note: For full item names see Table 3.2.

The other descriptive statistics are about the sample, taken by Vringer and Carabain (2020). Table 3.4 shows the frequencies of some social-economic variables of the respondents. Note that the background variables were only queried for the citizens, which is why N does not add up to 2111. Voting behavior was not included in this table, due to the many answer options.

Table 3.4 Social-economic aspects of the sample

	Number of respondents	Percentage of sample
Business representatives	833	39,46%
Citizens	1278	60,54%
Gender		
Male	652	51,02%
Female	626	48.98%
Age		
18 – 39 years old	389	30,44%
40 – 64 years old	576	45,07%
65 years old and above	313	24,49%
Education		
Low	314	24,57%
Middle	511	39,98%
High	453	35,45%
Income		
<€28,600	238	18,62%
€28,600 - €71,000	569	44,52%
>€71,000	186	14,55%
Don't know/Don't want to say	285	22,30%

4   Exploratory factor analysis

For the creation of a scale, dimensionality of the included items is important. Dimensionality is the number of latent variables that can be found to explain the covariance between the proposed items. In the case of unidimensionality (so one latent variable for the items), a detailed model for scale creation can more easily be applied. In case of multidimensionality, the creation of a scale is more complicated. Before moving to the creation of the scale, I therefore first need to explore the dimensionality of the good governance criteria.

As there is no information available on how to measure legitimacy, I start with an exploratory factor analysis (EFA). An EFA can measure whether and how many latent variables can be found that explain the covariance within the proposed items (in my case the good governance criteria presented in the operationalization). Within factor analysis, this latent variable is called a factor.

Two main goals are achieved using this method. First of all, as mentioned before, the analysis can identify the dimensionality of the good governance criteria (Carpenter, 2018). Unidimensionality would, for the creation of a scale, be ideal, because a more detailed method can be used. Second of all, the EFA indicates factor loadings for all the different items. These factor loadings indicate for every item individually to what extent the latent variable explains their variation. This offers a first idea about the relationship between each individual item and the latent variable(s).

4.1.1   Pre-tests

Before performing a factor analysis, a short look at Spearman's correlation matrixes, the Bartlett's test of sphericity, and Kaiser-Meyer-Olkin (KMO) is required (Carpenter, 2018). The first two tests determine whether the correlation between the items is sufficient to perform a factor analysis. In the Spearman's correlation matrixes, correlations between the items in question of at least 0.3 are required to perform factor analysis (Carpenter, 2018). Otherwise, the relationship might not be strong enough to be able to point out a factor. As can be seen in the correlation matrixes on the next pages (Tables 4.2, 4.3, 4.4 & 4.5), none of the correlations is lower than 0.3, with 0.31 being the lowest value between 'effectiveness (instrument helps)' and 'equitable, inclusiveness and following the rule of law' for intervention 4. The highest correlation can be found between 'effectiveness (feasible)' and 'transparent' for intervention 1 with 0.74. The high values in this correlation matrix show that the covariance between the items is high.

The Bartlett's test is a second test to analyze the correlation between the items. The H_0 of a Bartlett's test is that the items are uncorrelated. For all four interventions the H_0 can be rejected (see Table 4.1). Therefore, it's likely that the factor analysis can find factors that explain the covariance. It also means that no items have to be excluded from the analysis.

The KMO tests whether the sample is adequate for factor analysis. High values (close to 1) are regarded as suitable for factor analysis. In this case, all four interventions have a very high value for KMO, which means the data is very suitable for analysis (see Table 4.1). According to Kaiser (1974), a score between 0.9 and 1 can be regarded as 'marvelous'.

Table 4.1 Bartlett's test of sphericity and Kaiser-Meyer-Olkin (KMO)

	Bartlett's test of sphericity		KMO
	X^2	P	
Intervention 1	9603.271	0.000	0.945
Intervention 2	8963.849	0.000	0.948
Intervention 3	9141.785	0.000	0.934
Intervention 4	6912.889	0.000	0.934

Table 4.2 Spearman's rank correlation matrix for intervention 1

	Helps	Efficient	C. R. Implement	C. R. Set-up	Feasible	Transparent	Fair	Participatory
Helps	1.00							
Efficient	0.57	1.00						
C. R. Implement	0.62	0.64	1.00					
C. R. Set-up	0.59	0.62	0.72	1.00				
Feasible	0.61	0.69	0.66	0.65	1.00			
Transparent	0.59	0.65	0.67	0.61	0.74	1.00		
Fair	0.53	0.55	0.62	0.61	0.57	0.52	1.00	
Participatory	0.61	0.58	0.65	0.63	0.63	0.61	0.57	1.00

Note: For full item names see Table 3.2.

Table 4.3 Spearman's rank correlation matrix for intervention 2

	Helps	Efficient	C. R. Implement	C. R. Set-up	Feasible	Transparent	Fair	Participatory
Helps	1.00							
Efficient	0.54	1.00						
C. R. Implement	0.56	0.60	1.00					
C. R. Set-up	0.58	0.57	0.71	1.00				
Feasible	0.61	0.66	0.66	0.67	1.00			
Transparent	0.61	0.62	0.66	0.66	0.70	1.00		
Fair	0.48	0.52	0.61	0.59	0.59	0.58	1.00	
Participatory	0.56	0.56	0.62	0.61	0.61	0.61	0.59	1.00

Note: For full item names see Table 3.2.

Table 4.4 Spearman's rank correlation matrix for intervention 3

	Helps	Efficient	C. R. Implement	C. R. Set-up	Feasible	Transparent	Fair	Participatory
Helps	1.00							
Efficient	0.51	1.00						
C. R. Implement	0.56	0.63	1.00					
C. R. Set-up	0.57	0.55	0.65	1.00				
Feasible	0.63	0.60	0.63	0.65	1.00			
Transparent	0.57	0.66	0.66	0.60	0.72	1.00		
Fair	0.56	0.58	0.62	0.58	0.59	0.60	1.00	
Participatory	0.55	0.58	0.61	0.56	0.58	0.58	0.72	1.00

Note: For full item names see Table 3.2.

Table 4.5 Spearman's rank correlation matrix for intervention 4

	Helps	Efficient	C. R. Implement	C. R. Set-up	Feasible	Transparent	Fair	Participatory
Helps	1.00							
Efficient	0.38	1.00						
C. R. Implement	0.54	0.56	1.00					
C. R. Set-up	0.56	0.46	0.61	1.00				
Feasible	0.57	0.58	0.64	0.61	1.00			
Transparent	0.47	0.57	0.61	0.55	0.65	1.00		
Fair	0.31	0.45	0.47	0.42	0.44	0.48	1.00	
Participatory	0.49	0.47	0.55	0.52	0.56	0.56	0.44	1.00

Note: For full item names see Table 3.2.

4.1.2 Extraction method

To conduct a factor analysis, the extraction method has to be set. This determines how factors and factor loadings are calculated. Even though little information exists about what method should be used (Costello & Osborne, 2005), Carpenter (2018) advises to use maximum likelihood or principal axis factoring. I chose to use principal axis factoring, as the data does not meet the requirement of maximum likelihood to be normally distributed (Carpenter, 2018). A Shapiro-Francia test is included in the appendix (see Table 9.1, 9.2, 9.3 & 9.4) that shows the violation of the normality hypothesis for most of the items. Because this test is quite sensitive, I also examined the bar graphs of the different items and came to the same conclusion.

4.1.3 Outcome

In order to determine the number of factors that should be retained (and thus to determine the dimensionality), I use scree tests. Carpenter (2018) mentions that scree tests are considered an accurate manner to determine the number of retained factors. To understand what scree tests are, a minor understanding of eigenvalues is needed. Eigenvalues are characteristics of the factors that are found that indicate how much total variation in the items can be explained by that factor. A high eigenvalue therefore means the factor is able to explain a large part of the variation in the items. In many papers an eigenvalue of more than 1 is used as a cut-off line, as this indicates the factor is able to explain more than a single item can, but scree plots are considered more accurate. Scree tests use a visual representation of eigenvalues in a plot. The cut-off line is where the plot elbows and a straight, almost horizontal line is formed.

Following this line of reasoning, only one factor can be retained for each of the four interventions. The scree plots (Figure 4.1) show clear elbowing after the first factor and apart from the first factor all eigenvalues are far below 1. This means that all seven other factors do not explain enough variance to be retained. The first factor explains a large amount of the covariance between the items and therefore has a high eigenvalue.
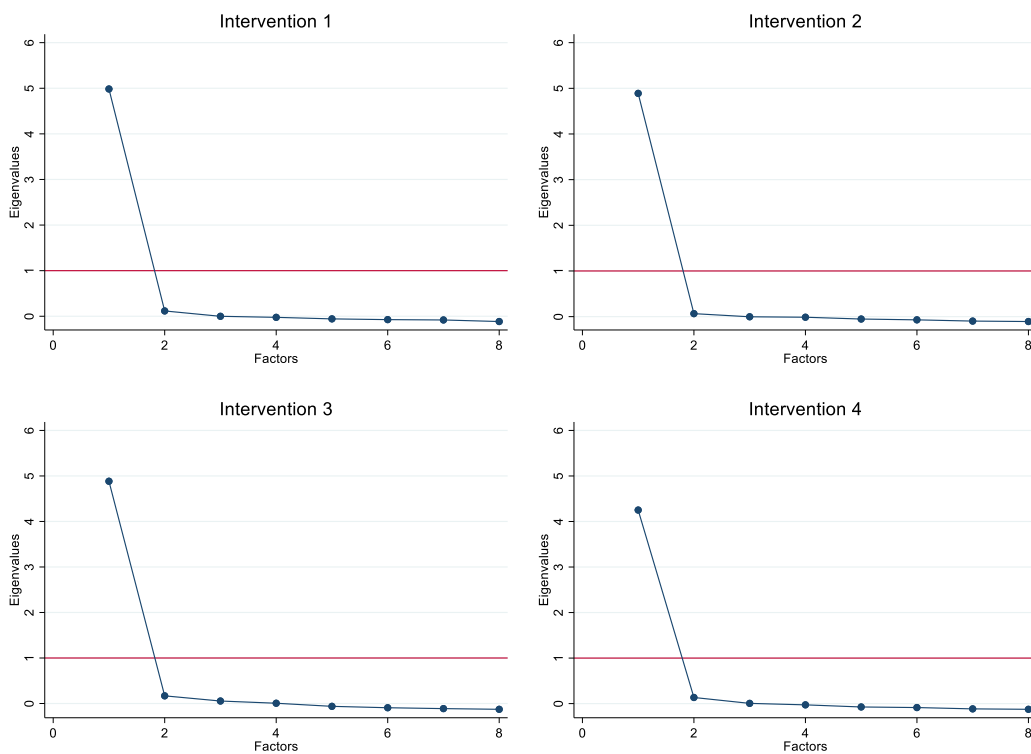

Figure 4.1 Scree plots for all four interventions

As unidimensionality for each intervention is confirmed, I can use more advanced techniques to produce a scale from the eight items. It allows me to create a scale and calculate overall scores for every respondent on the latent variable. I continue to explore this latent variable by examining the functioning and importance of the different items for it.

The exploratory factor analysis was reproduced with only one factor included. A graphical representation of the results can be found in the Figures 4.2, 4.3, 4.4 and 4.5. The figures show the factor loadings of the latent variable on the items and the unique variance for every item. The factor loadings are quite high overall. With the high covariance in the correlation matrixes, this was to be expected. Still, it means that the maintained factor can explain a big part of covariance between the eight items. This strong unidimensional factor is a good starting point for the item response model, because it confirms two of the three important requirements for utilizing such a model, which is explained in the next chapter.

Furthermore, the factor loadings show a quite similar model for at least the first three interventions. The fourth intervention, focused on obliging companies to use recycled material in their products and packaging, has some factor loadings that differ from the others. For example, the factor loading on 'equitable, inclusiveness and following the rule of law' is lower than for the other three interventions. This lower value means that in this intervention, the item 'equitable and inclusive, following the rule of law' has more unique variance. A bigger part of its variance is therefore not related to the latent factor. If we would assume that the factor found here is the factor I am looking for, namely perceived legitimacy of transition policy, it could mean that for this intervention that item is of less importance. The highest factor loadings are found for 'effective (feasible)'. Argued in the same way, it is to be expected that this item is of importance for the latent variable.
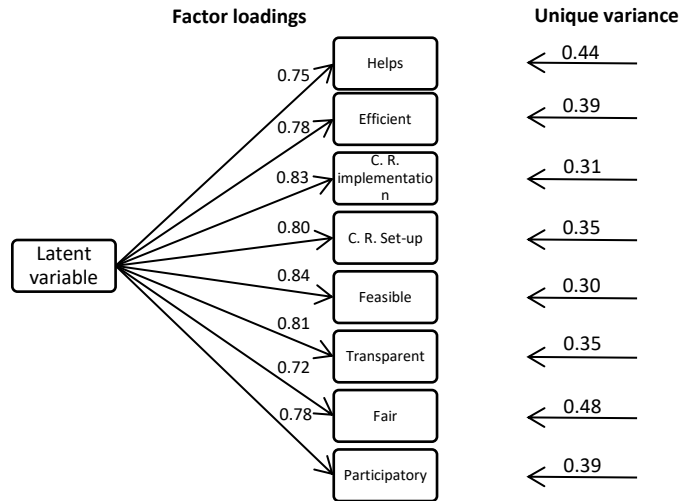
Factor loadings

0.75 — Helps
0.78 — Efficient
0.83 — C. R. implementation
0.80 — C. R. Set-up
0.84 — Feasible
0.81 — Transparent
0.72 — Fair
0.78 — Participatory

Latent variable

Unique variance

0.44
0.39
0.31
0.35
0.30
0.35
0.48
0.39

Figure 4.2 Graphic representation of EFA for intervention 1
Note: For full item names see Table 3.2.

Factor loadings

0.73 — Helps
0.75 — Efficient
0.81 — C. R. Implementation
0.81 — C. R. Set-up
0.83 — Feasible
0.82 — Transparent
0.73 — Fair
0.76 — Participatory

Latent variable

Unique variance

0.47
0.44
0.34
0.34
0.30
0.32
0.47
0.42
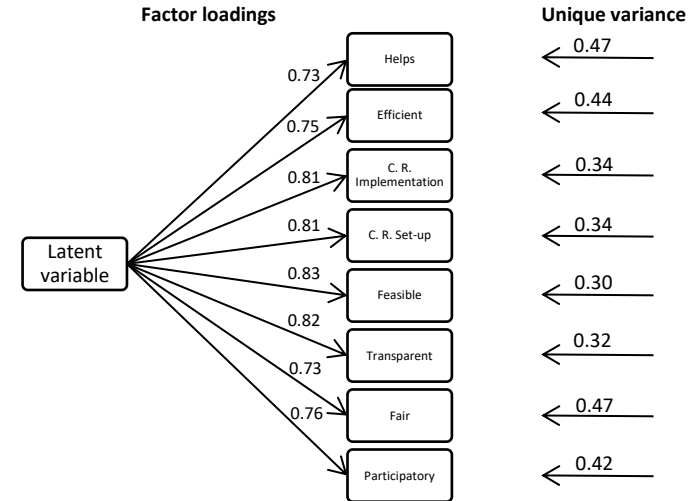
Figure 4.3 Graphic representation of EFA for intervention 2
Note: For full item names see Table 3.2.

Factor loadings

0.72 — Helps
0.76 — Efficient
0.80 — C. R. Implementation
0.76 — C. R. Set-up
0.81 — Feasible
0.82 — Transparent
0.80 — Fair
0.78 — Participatory

Latent variable

Unique variance

0.47
0.43
0.36
0.42
0.34
0.33
0.37
0.39
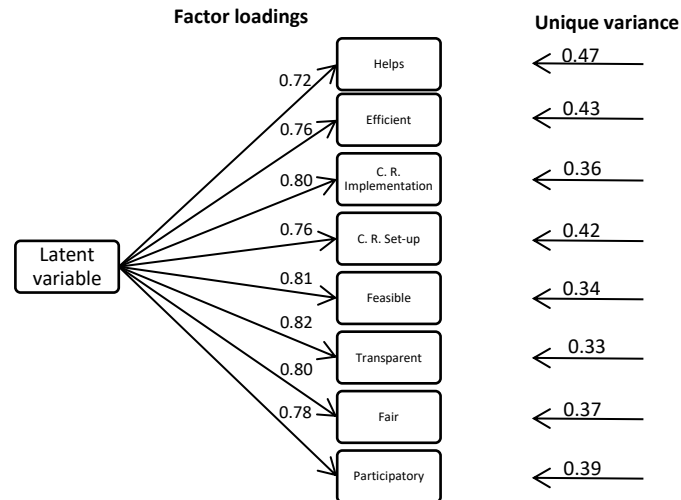
Figure 4.4 Graphic representation of EFA for intervention 3
Note: For full item names see Table 3.2.

Factor loadings

0.66 — Helps
0.70 — Efficient
0.79 — C. R. Implementation
0.74 — C. R. Set-up
0.82 — Feasible
0.78 — Transparent
0.60 — Fair
0.71 — Participatory

Latent variable

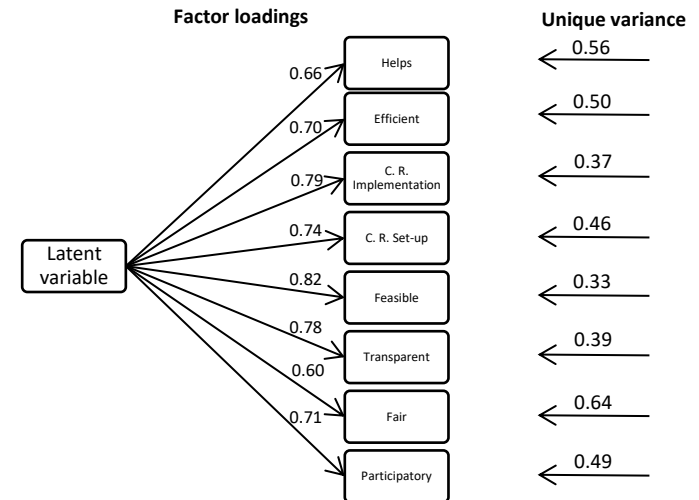**Unique variance**

0.56
0.50
0.37
0.46
0.33
0.39
0.64
0.49

Figure 4.5 Graphic representation of EFA for intervention 4
Note: For full item names see Table 3.2.

5 Graded response model

Because I now know that the responses to the eight items are unidimensional, I can choose an advanced model to calculate scores for the latent variable. A simple model would be to assign weightings to the different answers for all items. However, determining these weightings is hard and may not be precise. Because a simple model that weighs every item equally has multiple disadvantages, I will utilize item response theory (IRT). Item response models are a good method to create and evaluate multi-item scales (Toland, 2014). This model calculates a score for every respondent on the latent variable. There are several models, but in this case, a graded response model is appropriate, as they are specifically meant for models that include items with ordinal values, such as Likert scales. These models allow to evaluate, or at least estimate, the probability that a particular test subject will receive a specific grade or score for each item. It also allows to estimate how well the test questions measure that latent trait or ability. The explanation of this model can therefore further the understanding of the relationship between the items and the latent variable.

To be able to perform the graded response model, the data has to meet the pre-requisites of such an analysis. Lameijer et al. (2020) distinguish unidimensionality, local independence and monotonicity. Unidimensionality and local independence are covered by the exploratory factor analysis, as mentioned before. If it is the case that we find only one factor and the unique variances of the items is not big, unidimensionality and local independence can be assumed. Monotonicity requires the researcher to make clear that it can be stated confidently that higher levels of the items require higher levels of the latent variable. With the theoretical exploration that was performed in the theoretical section, I assume that all the good governance criteria have a positive impact on the latent variable.

The latent variable in a graded response model is denoted by θ. Note that the graded response model calculates a relative score. That is, the mean θ of the total sample on the basis of which you calculate the model will always be 0 with a standard deviation of 1. A negative score therefore does not necessarily mean that the level of the latent trait is 'negative' or 'bad'. The only conclusion that can be drawn is that a negative score represents a respondent that has a lower score on the latent trait than the mean of the sample. Additionally, a θ-score usually ranges between -4 and 4.

The graded response model calculates two types of defining parameters for every item in the scale: difficulty and discrimination. The term difficulty originates in educational testing, where the items are questions that can be classified as hard or easy. The difficulty parameter shows the threshold of the level of the latent trait (θ) above which a respondent has at least a 50% chance to choose that response category or a higher category. This is an expression of the 'difficulty' of the item, as the item threshold parameter (difficulty) can clarify which item categories can distinguish between people with low levels of θ and people with high levels of θ. The second parameter is discrimination. This parameter can be seen as the slope of an item, which shows 'the strength of the relationship each item on a multi-item scale has with the latent trait variable being measured' (Toland, 2014, p.122). The higher this parameter, the better this item can discriminate between people above and beneath one of the thresholds (difficulty) of that item.

To clarify this even better, I make use of an example before moving towards the model I constructed using the good governance criteria. Assume I would like to test people's ability in mathematics. Using different questions, I would like to get an overall score for the latent variable θ that is mathematical ability. Assume that one of the questions would be: 'What is $2 + 3 * (4 - 1)$?' Another question would be: 'Calculate the value of x in $\log(x) + x^4 = 60$'. Obviously, the first question is a lot easier than the second one. Therefore, the difficulty parameter of the first question will be at a lot lower level of θ

than that of the second question. Figure 5.1 is a made-up graph that depicts the difference in difficulty between the items. On the y-axis the probability to give the right answer is shown. Notice how for the easy question, the point where there is a 50% chance to give the right answer is at a level of θ = -1.22. For the harder question this level of θ is a lot higher at θ = 1.59. People at approximately θ = 0 (the red dotted line) already have almost 100% chance to get the easy question right, while they also have a close to 0% chance to get the hard question right.
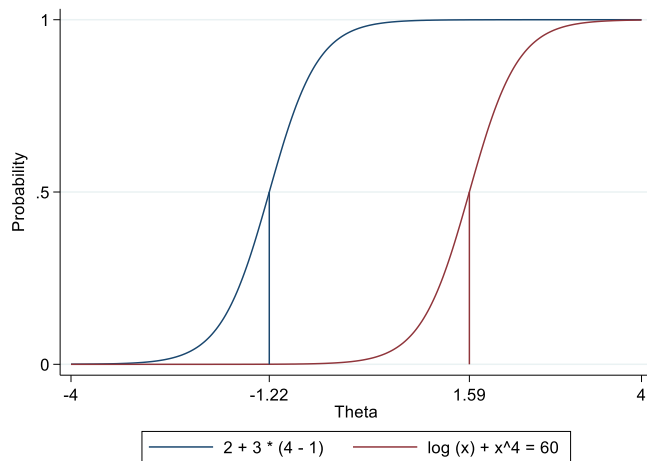
Figure 5.1 Threshold example Boundary characteristics curves

To explain discrimination, I will use a different question as examples. Still, let us assume I want to measure mathematical ability and the question of $\log(x) + x^4 = 60$ remains. The other question would be: 'What are the first two decimals of π?' Notice how in Figure 5.2 the slopes of the lines differ. Asking about the first decimals of π is not really a measure of mathematical ability, but more about knowledge. Therefore, the question about π is answered correctly by some people who do not do as well on all of the other questions, while it is answered incorrectly by some people that do very well on all the other questions. This is what is called discrimination. Apparently, the question about π does not do well at discriminating between people who have a high level of mathematical ability and people who have a low level of mathematical ability. Therefore, its discrimination parameter will be lower and the slope of the line is less steep.
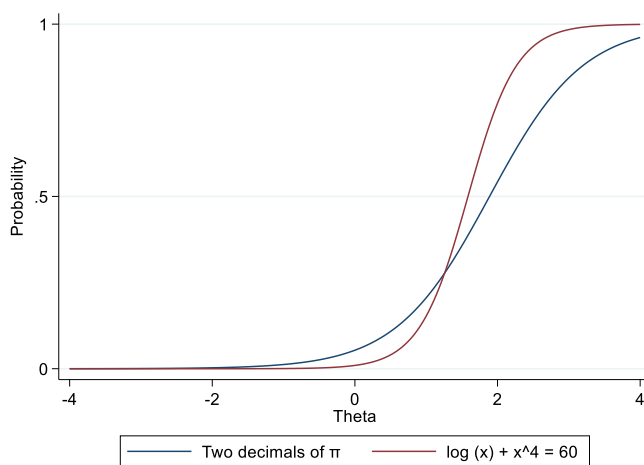

Figure 5.2 Discriminatory example Characteristic curves

A third important concept of an item response model is information and is a combination between the difficulty and discrimination values. Information is an expression of how precise a model or item can estimate the score of θ for that specific point on the range. It is therefore negatively correlated with the standard error of the estimation of θ. Information is high around the difficulty threshold and higher for items with high values of discrimination.

Because this model is quite complicated, this chapter covers only one of the four interventions. That way, the results are still understandable and readable without assuming too much knowledge from the reader. The raw tables of results of the other three models can be found in the appendix (see Table 9.5, 9.6 & 9.7). The intervention that is analyzed here is intervention two, which offers an in-home display to all households and companies, paid for by the energy company (see paragraph 3.2).

5.1 Outcome

5.1.1 Discrimination and thresholds

First, I explain what the model found for discriminatory values for the eight different items (see Table 5.1). These values indicate the strength of the relationship between the item and the latent variable (θ) (Toland, 2014). A higher value therefore indicates more precise measurement of the latent variable through that item. The discrimination values range from 2.2 for 'equitable and inclusive, following the rule of law' to 3.4 for 'effective (feasible)'. To understand what this means, remember the example about mathematical ability. A higher discriminatory value means the item measures the latent variable θ more precisely and is thus more capable of distinguishing between different levels of θ.

Threshold values indicate at which level of θ there is a 50% chance that a respondent will answer with that response. Higher thresholds therefore indicate items that measure higher levels of θ. Note the difference in items between this model and the example about mathematical ability. In the example about mathematical ability, the items were binary. The respondent either gets the question right or wrong. The eight items I included in this model, are scored on a Likert scale, therefore having five response options (1 being 'Strongly disagree' and 5 being 'Strongly agree'). Where for the mathematical ability example I would get one value for the difficulty threshold, in this case I get four thresholds. Every threshold depicts the value of θ for which the respondent has a 50% chance to pick that response category or higher.

The thresholds for the items seem similar, indicating that one question does not directly measure higher values of θ and others measure lower values of θ. It shows for instance that 'effective (instrument helps)' has a high score for the highest threshold with 1.242 while for example 'consensus oriented, responsive (set-up)' has a lower value with 0.637. This indicates that the 'effective (instrument helps)' question has more value when predicting high scores of θ around that 1.242 threshold. Translating this in more practical terms, 'effective (instrument helps)' can be seen as a fundamental quality for legitimate policy, as a high score for this is associated with a very high score of θ, while, for other items, the associated θ-score is lower. However, the differences are small and the outer ranges of θ are not covered.

Table 5.1 Final GR Model for intervention 2

Items	Discrimination	Thresholds			
	a	b1	b2	b3	b4
Helps	2.213 (.085)	-1.517 (.058)	-0.890 (.042)	-0.084 (.034)	1.242 (.050)
Efficient	2.420 (.094)	-1.374 (.053)	-0.662 (.038)	0.064 (.033)	1.012 (.045)
Cons., resp. (implementation)	3.219 (.127)	-1.627 (.056)	-1.007 (.040)	-0.357 (.031)	0.664 (.035)
Cons., resp. (set-up)	3.089 (.120)	-1.723 (.059)	-1.121 (.042)	-0.404 (.031)	0.637 (.035)
Effect. (feasible)	3.355 (.130)	-1.447 (.050)	-0.836 (.036)	-0.068 (.030)	0.834 (0.37)
Transparent	3.233 (.125)	-1.267 (.046)	-0.659 (.034)	0.024 (.030)	0.959 (.039)
Equi. & incl., rule of law	2.186 (.084)	-1.677 (.063)	-0.991 (.044)	-0.144 (.034)	1.041 (.046)
Participatory	2.512 (.095)	-1.464 (.054)	-0.798 (.038)	0.048 (.032)	1.167 (.046)

Note: a = the slope of the characteristic curve and therefore discrimination (SE). b1 = threshold for an answer of >= 2 (SE). b2 = threshold for an answer of >= 3 (SE). b3 = threshold for an answer of >= 4 (SE). b4 = threshold for an answer of = 5 (SE). For full item names see Table 3.2.

To get a better understanding of what these values mean, I use characteristic curves. As we can see in Table 5.1, the item 'consensus oriented, responsive (implementation)' and the item 'equitable and inclusive, following the rule of law' have different levels of discrimination. Their value for b2 (a response of Neutral (3) or higher) is almost similar. This means that at approximately the same level of θ the chance of a respondent answering 3 or higher is exactly 50%. We can see this in Figure 5.3, as the two lines intersect the 0.5 probability at almost the same value of θ.
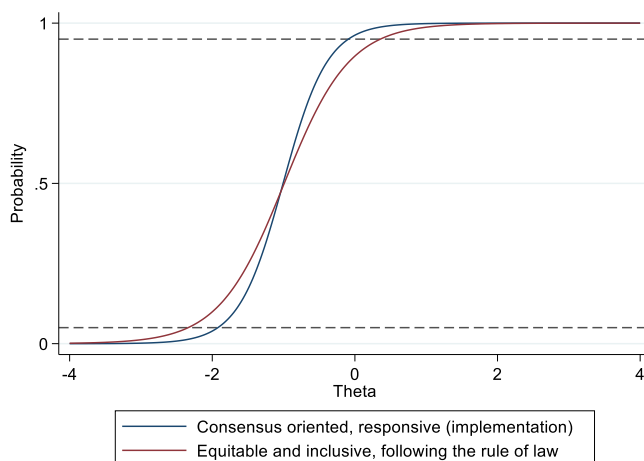

Figure 5.3 Response of neutral (3) or higher for the item 'consensus oriented, responsive (implementation)' and the item 'equitable and inclusive, following the rule of law' compared

Figure 5.3 shows the difference in the steepness of the curves for the two items. The two dashed lines show a probability of 0.05. The graph shows that the range of θ for which the model can confidently expect an answer above or beneath the threshold, is bigger for the item 'consensus oriented,

responsive (implementation)' than for the item 'equitable and inclusive, following the rule of law'. That means that for a bigger portion of the sample, an answer to the item 'equitable and inclusive, following the rule of law' is not giving clear information about where they belong on the θ-scale.

It is also possible to show the difference in discrimination and thresholds by looking at all the response options for two different items (see Figure 5.4 and 5.5). I now compare the item 'effective (feasible)' and the item 'equitable and inclusive, following the rule of law'. The figures show the five response categories and the probability they are picked by a respondent with a certain level of θ. The discrimination again is seen in the slope of the curves. The higher the curve, the more precise a certain response indicates a θ-score. Notice how, in the graph, the probabilities of the five response categories for every point on the x-axis add up to 1, because a respondent always picks one of the responses (missing values are not included in the analysis, as mentioned before).

The thresholds can be found in Figure 5.4 and 5.5 as well. The range of thresholds for the item 'effective (feasible)' is smaller than the range for the item 'equitable and inclusive, following the rule of law'. This can be seen by comparing the 0.5 probability intersection of the two outer responses for the two figures. This means that even though less precise overall, the item 'equitable and inclusive, following the rule of law' can give us more information for the high and low values of θ.
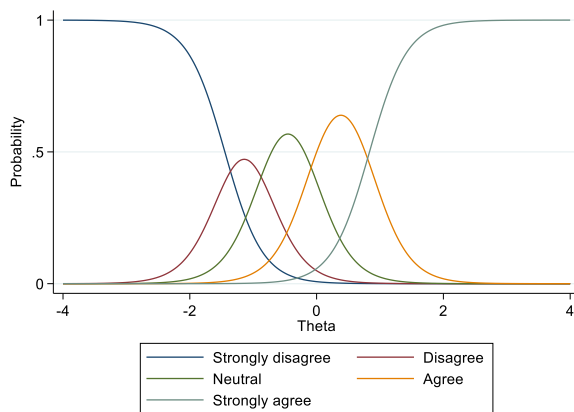

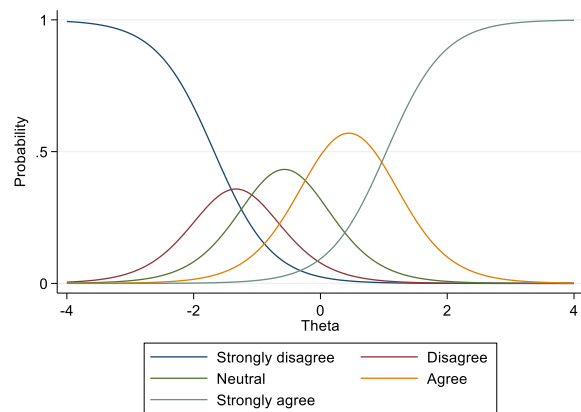*Figure 5.4 Category characteristic curves 'Effective (feasible)'*

Figure 5.5 Category characteristic curves 'equitable and inclusive, following the rule of law'

5.1.2   Model information

A third important value in IRT is the value for information. In the previous analysis I already mentioned how we can see the precision of items for the estimation of θ in the graphs. It is also possible to give a value to the information an item gives. Every item has its own information function. For this, I again compare the items 'effective (feasible)' and 'equitable and inclusive, following the rule of law'. Note that I said earlier that the item 'effective (feasible)' is more precise and offers greater information, but the item 'equitable and inclusive, following the rule of law' provides information over a wider range.

Figure 5.6 shows the difference in information for the two items. The item 'effective (feasible)' has a higher value of information for a particular range, but the range of information on the item 'equitable and inclusive, following the rule of law' is broader. The same can be done for all items at once, resulting in Figure 5.7.
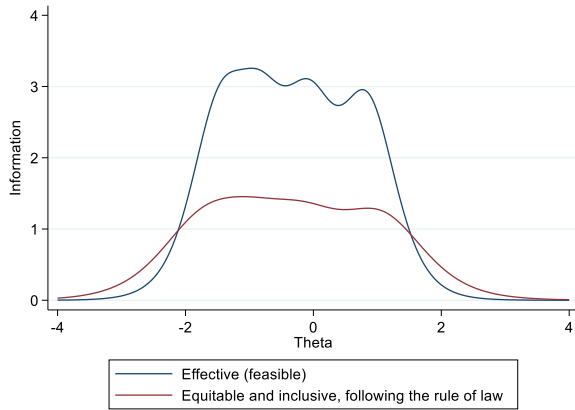
Figure 5.6 Item information functions of item 'Effective (feasible)' and item 'Equitable and inclusive, following the rule of law'
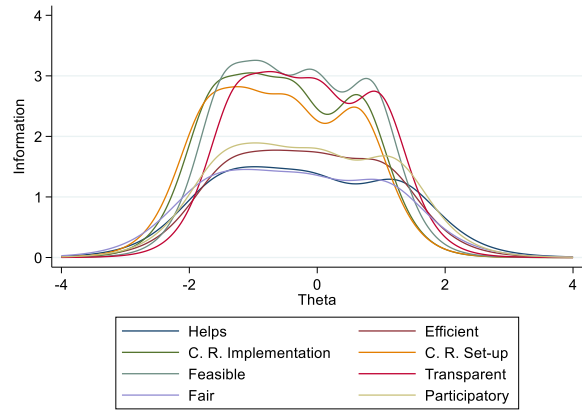
Figure 5.7 Item information functions of all items
Note: for full item names see Table 3.2

The information curves are clustered in different groups. For example, 'equitable and inclusive, following the rule of law' has almost the same information function as 'effective (instrument helps)'. This can be a reason to exclude a certain item from the model, as it might not offer additional information (Toland, 2014). This can only be the case if it is deemed plausible that the two items measure exactly the same thing, thus not offering additional information. In our model, this is not the case. The two items that are compared measure wholly different concepts of good governance. Therefore, exclusion of items out of the model is not necessary.

An information curve can also be generated for the complete model (Figure 5.8). This shows especially at which range the model is useful and what values of θ can confidently be predicted using this model. The relation to standard error is also depicted: higher values for information mean lower values for standard error.
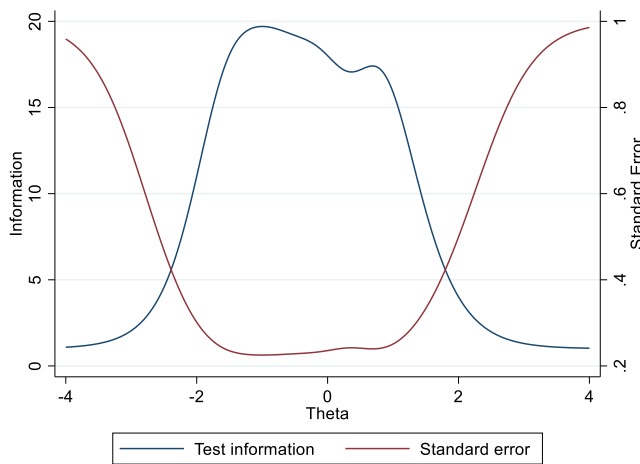

Figure 5.8 Test information function

The GR model is effective at a range from θ ≈ -2.1 till θ ≈ 1.8. To improve this range, questions that more specifically give information about low levels of θ or high levels of θ are needed. For example, the questionnaire could include more extreme statements, such as: 'this policy treats people equally'. This is a very fundamental way of formulating the 'equitable and inclusive' criterium of good

governance. This question would therefore be expected to be associated with low levels of θ and can therefore give us a wider range on low levels of θ.

Important to note is that values outside this confident range have a higher standard error. I found some values of θ that are outside this range. Another important note here is that the data shows a normal distribution of θ-scores for the confident part, but there is a clear increase for the highest possible and the lowest possible score (see Figure 5.9). The exact reason why is unclear, but it could be a biased response to the questionnaire. For instance, somebody could have rushed to fill in the questionnaire and therefore filled in all the most positive or all the most negative answers. It might also be a legit answer, as they might have had a positive association with transition policy in general and therefore fill in all the positive answers. It could also be that part of them should be on an even higher level of θ which cannot be measured on this scale.
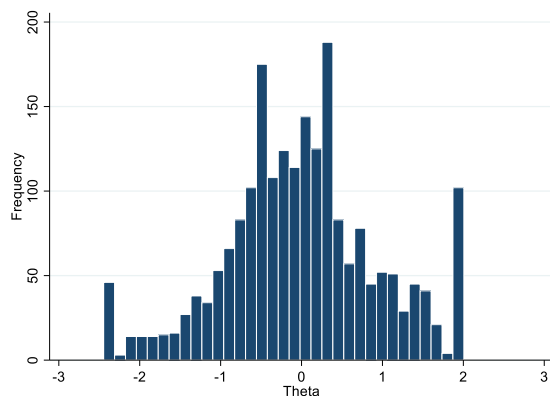

Figure 5.9 Frequency diagram of θ-scores for intervention two

5.1.3   Interpretation of the graded response model

The graded response model offers additional information about the importance of different items for different values of θ. The discrimination values of the items show a similar pattern to the factor loadings from the factor analysis. The same items have strong factor loadings as high values for discrimination. When it comes to the thresholds, there are no big differences between the items. Some items have a broader range than others, but effectively they all measure in the same range of θ. It could therefore be interesting to find additional items that can more effectively distinguish between high values of θ and low values of θ.

'Effectiveness (feasible)', 'transparency', 'Consensus oriented and responsiveness (implementation)' and 'consensus oriented and responsiveness (set-up)' are the most important items for the model generated for intervention 2. They offer the most information and can therefore best determine the scores of the respondents. Out of the four, 'transparency' has the biggest range regarding high scores of θ while 'consensus oriented and responsiveness (set-up)' has the biggest range regarding low scores of θ. It has to be noted that the differences are minor. In the end, it means that with a low score for the latent variable θ, a respondent is likely to score 'consensus oriented and responsiveness (set-up) higher than 'transparency', as his θ level might be above the first threshold for 'consensus oriented and responsiveness (set-up)', while he is still below the lowest threshold for 'transparency'. For example, a respondent with a θ-score of -1.5 would most likely answer 'Disagree' (2) to 'consensus oriented and responsiveness (set-up), while he is most likely to answer 'Strongly disagree' (1) to 'transparency'. In order to get a better sense about the generalizability of these statements, additional effort is required in comparing graded response models for different interventions.

5.2 How can theta scores say something about legitimacy of an intervention?

Now that the graded response model calculated a θ-score for every respondent, validation of that score is required. Up to this point, I discussed the latent variable without attaching a label to it. Using some background characteristic of the respondents, the exact meaning of θ can be explored. Did the graded response model actually measure perceived legitimacy of these interventions? To validate the model, an exploration of the outcome variable (θ) is interesting.

The θ-scores will be tested on a Spearman correlation with a question in the survey about direct support for the policy intervention. Even though support and legitimacy are not equivalent, they are closely related. A close relationship between the two is therefore a validation of the model. The θ-scores will also be correlated with each other. This offers additional data about whether the model measures a general perception of legitimacy of transition policy, or a score that is specific to the proposed intervention.

Furthermore, the θ scores will be compared to background variables. This part of the analysis includes putting the θ-scores of my interventions against a paper by McCright (2008). In his article, he analyzes support for climate policy and uses almost the same background variables as I do. Again, even though support is not equivalent with legitimacy, the main tendencies of distribution among different background variables are expected to be similar. Therefore, I utilize his claims as a basis for my comparison. They will be analyzed using multiple graphics.

5.2.1 Validation with support and across θ-scores

As mentioned in the methods section, the respondents were not only asked about their opinion on the eight items included in the graded response model, but also whether or not they would support the implementation of that intervention. To do so, the survey queried: 'Should this intervention be implemented as described?' In order to compare the results of the graded response model and public support for the intervention, a Spearman's rank correlation was performed between the θ-score and their answer to the question about support for the intervention.

Table 5.2 Spearman's test comparing θ-score to support for the policy intervention

	Observations (n)	Spearman's ρ	Significance (p)
Intervention 1	2088	0.75	0.00**
Intervention 2	2096	0.67	0.00**
Intervention 3	2086	0.71	0.00**
Intervention 4	2098	0.60	0.00**

*Note: * = significant at level of α=0.05. **= significant at level of α=0.01*

The Spearman's rank correlations show that the θ-scores are strongly related to the support for the proposed intervention (see Table 5.2). According to Leclezio, Jansen, Whittemore and De Vries (2015), Spearman scores of above .7 can be regarded as very strong relationships, while scores varying from 0.4 till 0.69 are regarded as strong relationships. These high values could indicate two things. On one hand, it could indicate that the model measures perceived legitimacy, as public support and legitimacy are closely related concepts. On the other hand, it could also indicate that it measures some expression of public support. At the same time, it is not a perfect correlation, meaning that it doesn't have to be an expression of public support for the intervention. This discussion is elaborated on in the discussion segment of this paper. The distribution of θ-scores among the different answer categories to the question of support is presented in Figure 5.10, using a violin plot. The violin plots show the frequencies

of the theta scores in the width of the 'violin'. The white dot marks the mean, while the blue bar represents the standard error for that category.
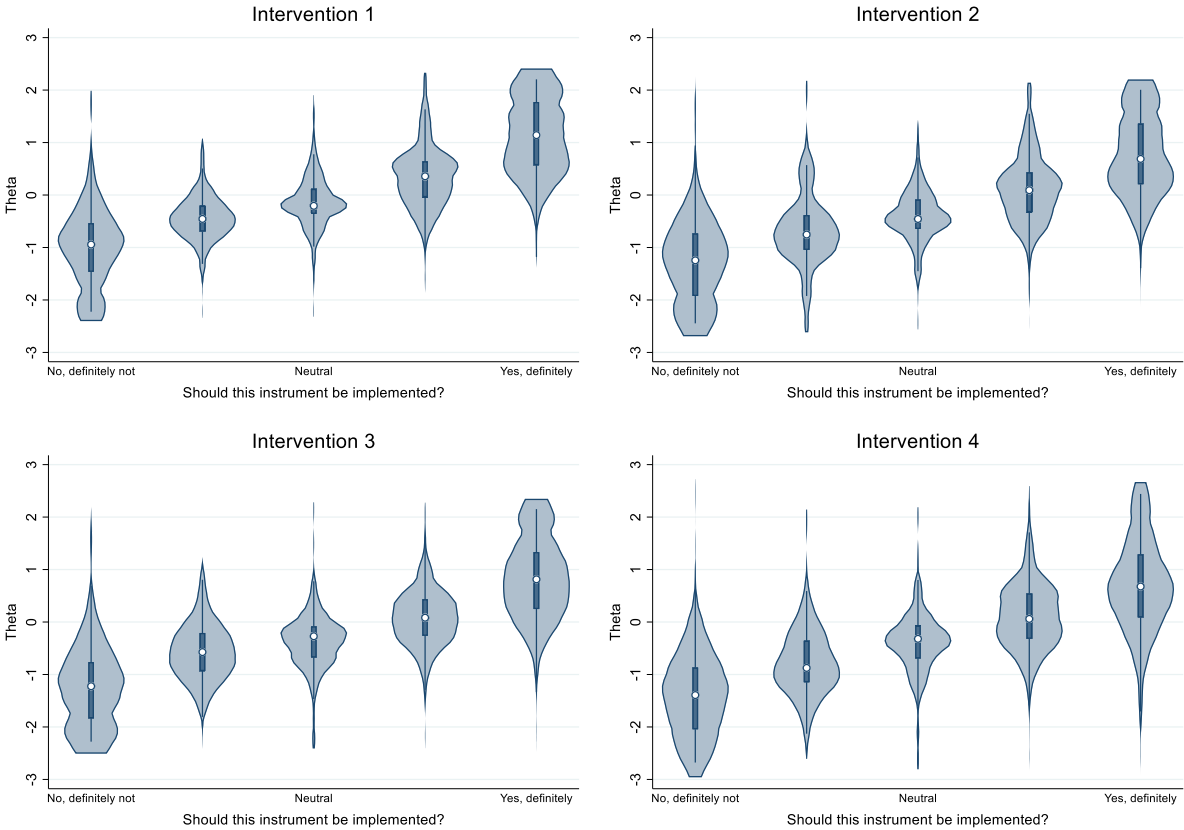

Figure 5.10 Violin plot comparing θ-scores and support

Another way to validate the θ-scores is to correlate them to each other. If what θ measures has to do with a perceived legitimacy of transition policy in general, it is expected that the scores for the four interventions correlate. If they are uncorrelated, it is likely that θ is an expression that is specific to the proposed intervention. Using a Pearson correlation, the four scores are compared. Table 5.3 shows that all correlations are significant. There is a high correlation found between the θ-scores for intervention 1 and 2. From all four interventions, the proposed policies are most similar as well. For the other interventions, the correlation is smaller. This indicates that the four graded response models all measure a personal perception of the respondent that is very specific to the chosen intervention. This corresponds with the approach of my research. The items were all specifically meant to measure the perceived legitimacy of a policy intervention, and not transition policy in general.

Table 5.3 Pearson correlation between the four different θ-scores

	θ-score intervention 1	θ-score intervention 2	θ-score intervention 3
θ-score intervention 2	0.5734**		
θ-score intervention 3	0.3393**	0.3095**	
θ-score intervention 4	0.3589**	0.3838**	0.3983**

*Note: * = significant at level of α=0.05. **= significant at level of α=0.01*

5.2.2   Descriptive statistics with background variables

Using density plots, averages and scatterplots, I describe the theta scores for six background variables. Density plots depict the frequency of a value relative to the total size of that group or category. In other words, it depicts a percentage, which makes groups of different sizes more easily comparable. This section does not include significance tests for these background variables, because of the explorative approach of this section. More in depth analysis including significance could be of great value, but goes beyond the scope of this paper. Most of the background variables are only queried for citizens, so there is a lower amount of observations of n=1278.

Citizens and business representatives

The first one is the distinction between citizens and business representatives. In Figure 5.11 below we can see the density plots for both groups for all 4 interventions. There is no specific trend to be found. For intervention 2 and 4, citizens seem to be slightly more positive, while for intervention 1 and 3 business representatives seem to be slightly more positive. The interventions themselves do not give clear indications which of the two groups was expected to be more positive.
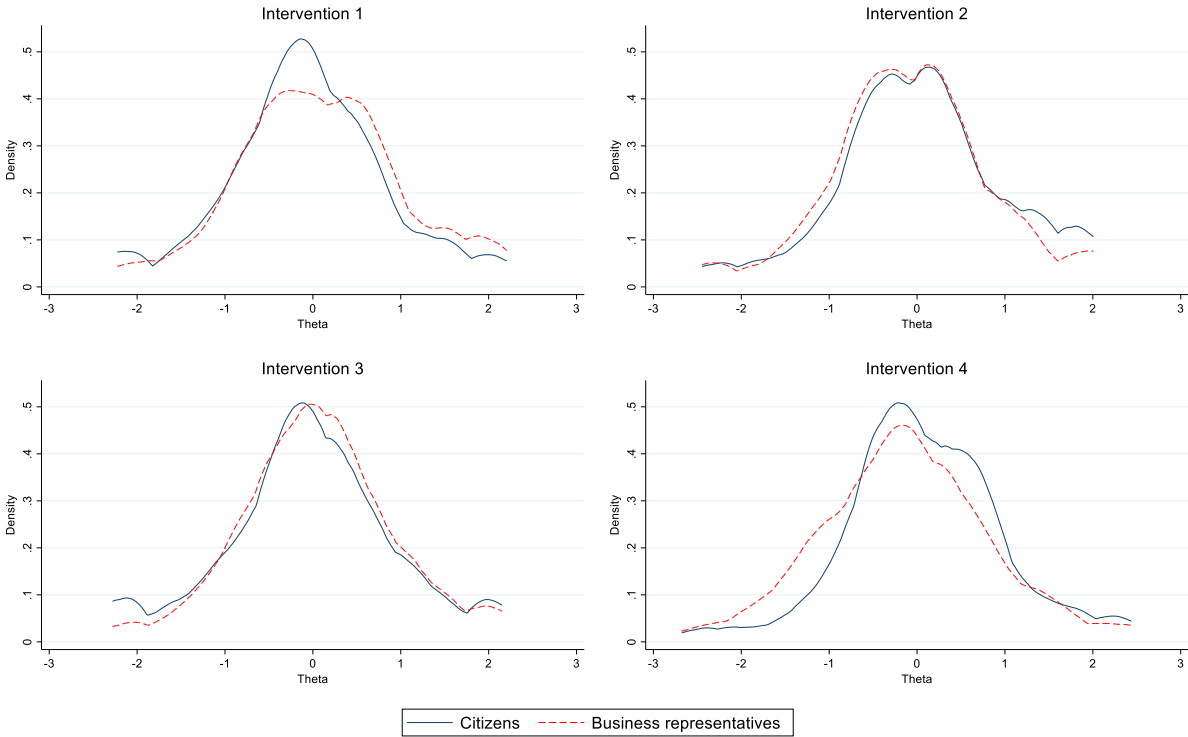

Figure 5.11 Density plots comparing θ-scores for citizens and business representatives

<u>*Gender*</u>

For the other background variables, only the data about citizens will be used, as the other five background variables are exclusively queried for citizens. The second background variable is gender. McCright (2008) shows that for his cases women are usually more supportive of transition policy than men. My results support that conclusion. Figure 5.12 shows that the two groups on all four interventions are similar. It seems that women are slightly more positive on three of the four interventions, except for intervention 4, where men seem to be more positive.

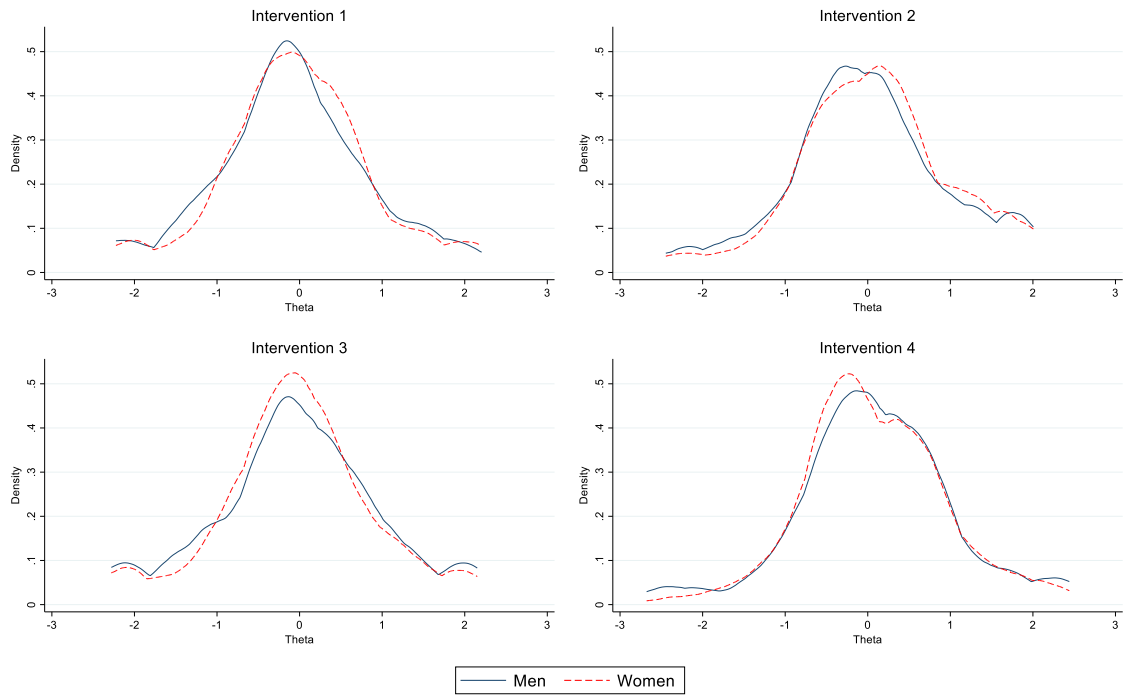*Figure 5.12 Density plots comparing θ-scores for men and women*

Age

The third background variable is age. McCright (2008) found that young adults are more supportive of climate change policy than other groups. In my samples this cannot be found. As we can see in Figure 5.13, the violin plots show no sign of a relationship between age and the θ-score.
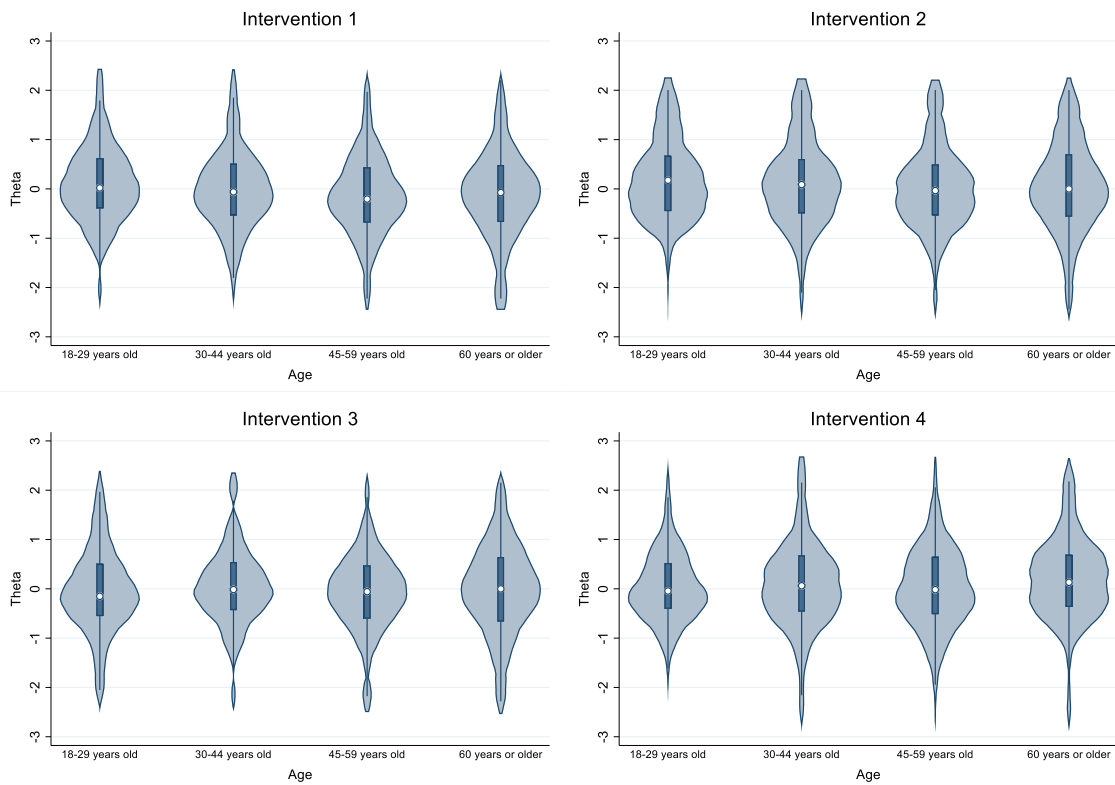

Figure 5.13 Violin plots comparing θ-scores to age

Education

The fourth background variable used is education level. The education level was divided in three levels: low, middle and high. McCright (2008) found a positive correlation for overall socioeconomic status. That is, the higher the education and income, the higher the support for climate policy. In our interventions, this is divided (see Figure 5.14). While intervention 1 and 4 seem to support that claim, the other graphs suggest something else. Intervention 2 seems to suggest that the highly educated respondents were less positive about the policy intervention. Intervention 3 on the other hand, shows no big differences between highly and averagely educated respondents, but shows a tendency towards more extreme scores for the lowly educated respondents.
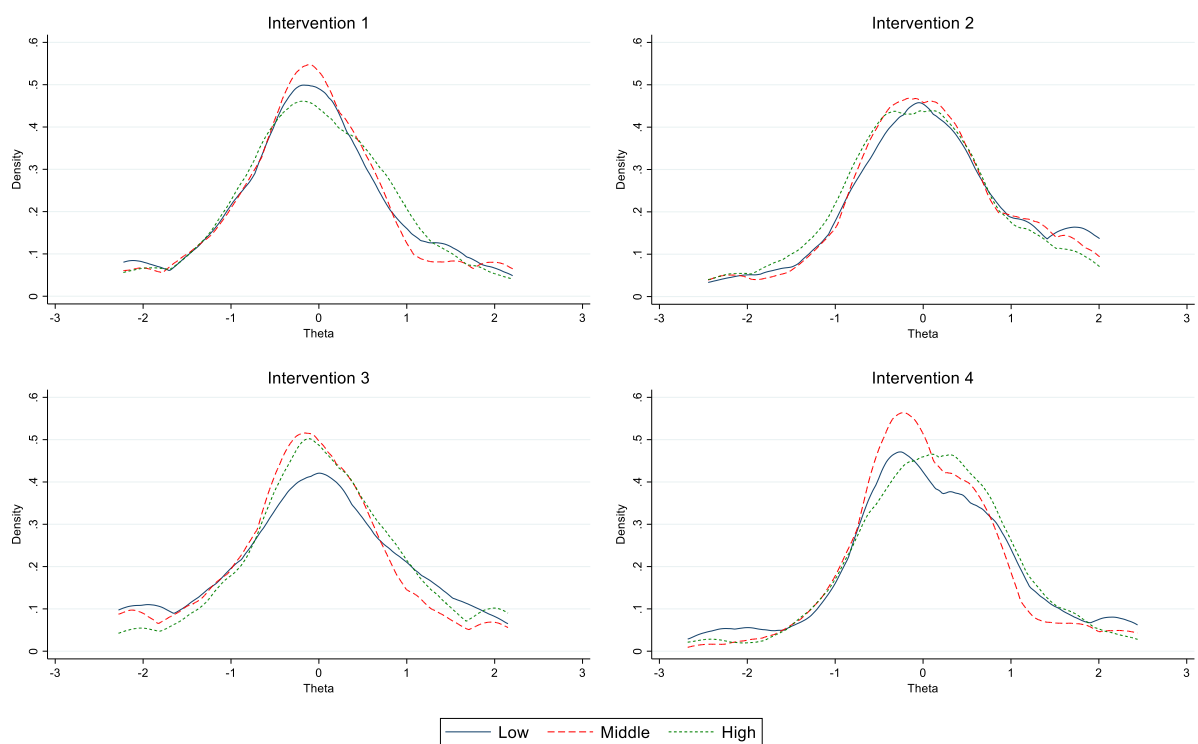
Figure 5.14 Density plots comparing θ-scores for different levels of education

Income

For socioeconomic status, McCright (2008) took education and income level into account. Income level will also be used as a background variable. With this background variable, the claim made by McCright (2008) is not very likely for my sample. For intervention 1, intervention 2 and intervention 3 the respondents with the highest income are the least positive about the policy intervention (see Figure 5.15). For intervention 4 this is not the case, as they seem to be the most positive of all groups. Interesting to note is that the group with the lowest earners relatively, has the most people with an extremely positive score for all 4 interventions.

Figure 5.15 Density plots comparing ϑ-scores for different levels of income

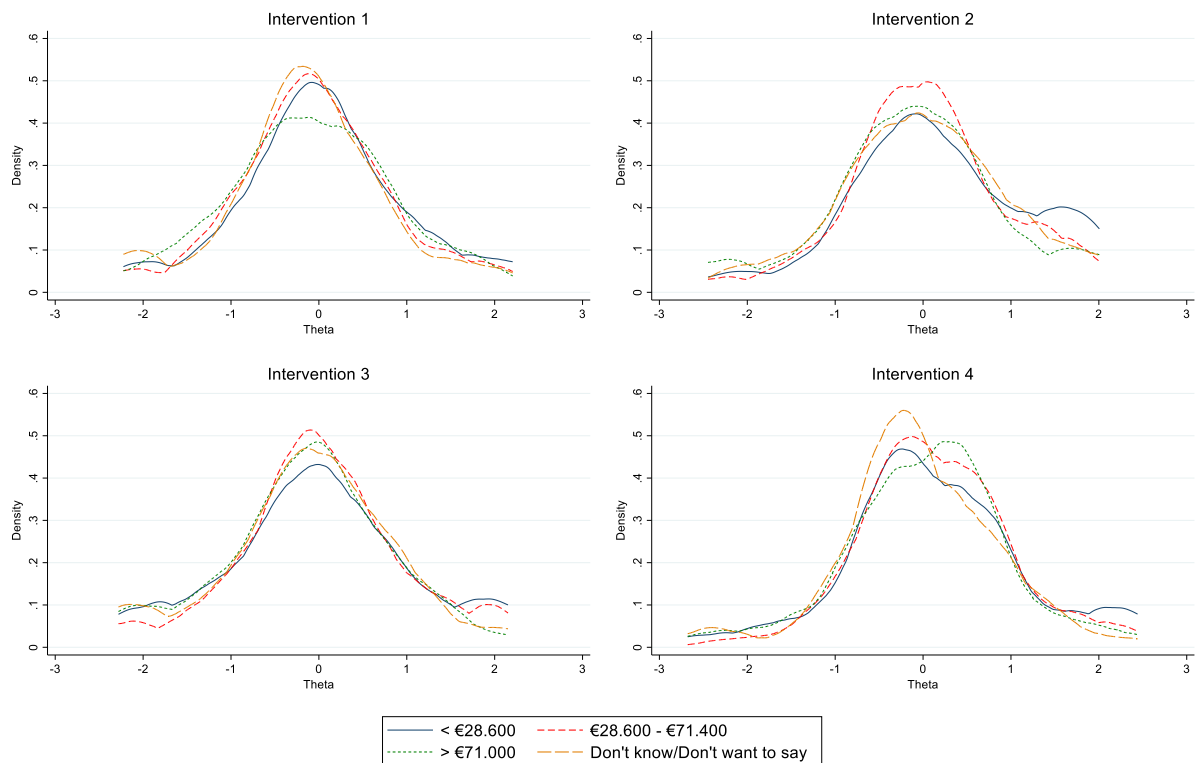### *Voting behaviour*

The final background variable that is considered is voting behavior. Not all parties that were mentioned are included, as their sample sizes are too small to calculate means and standard deviation. Therefore, only groups with at least 30 observations will be included in Figure 5.16.

Of course, there are some easy assumptions to make for voting behavior. Green left-wing parties like Groenlinks and D66 are in favor of climate policy, while more right-wing conservative parties like PVV and VVD are known to be less in favor of climate policy. The results shown in Table 4.10 below support those claims. Parties known for being in favor of climate policy (D66, GroenLinks, PVDA, Partij voor de Dieren) all have positive means of θ for the four interventions, which means they are above the average mean for the whole sample (which, as explained before, is 0). Parties that are known for opposing climate policy (PVV, 50 Plus, VVD to a lesser extent), have more negative values for the means of θ. VVD, the biggest party, is around the mean of the whole sample with means of θ close to 0. PVV on the other hand has the lowest mean score for θ of all parties on all interventions.
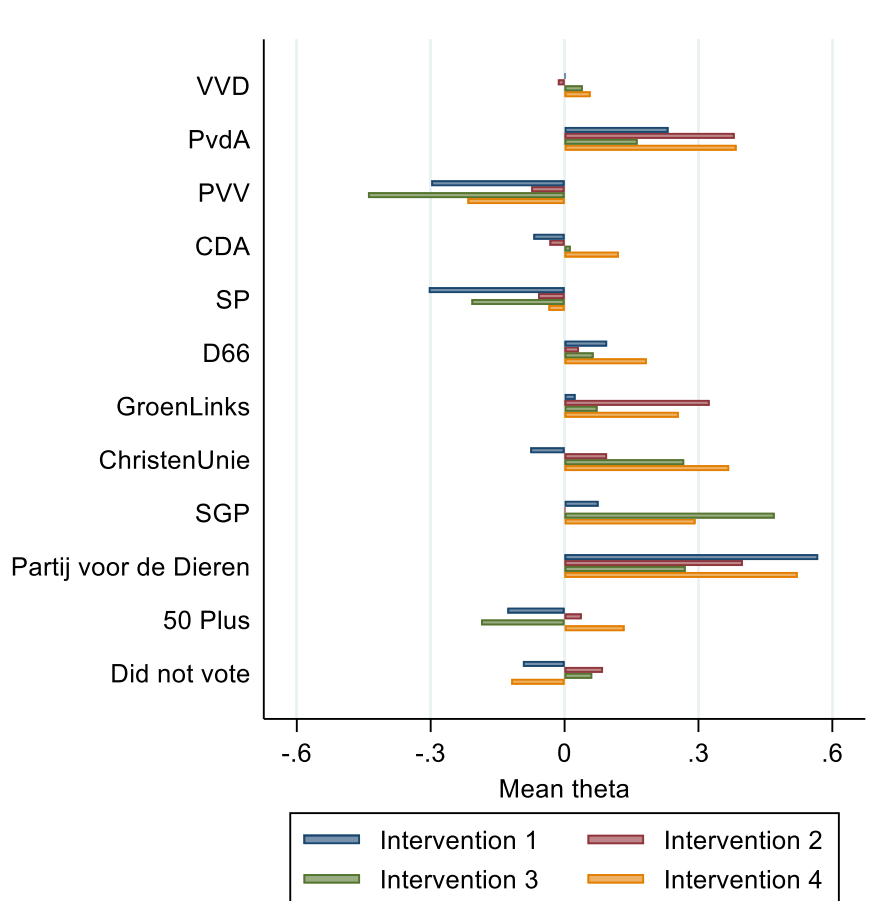
Figure 5.16 Means and standard deviation (SD) for voters of different political parties

Final remarks

Not all background values align with the expectations when comparing them to claims made by McCright (2008). Overall, the background variables do not seem to explain a lot of variance. The only variable that gives clear indications is voting behavior. The values here are as would be expected when measuring perceived legitimacy of transition policy interventions, with the voters of climate policy opposing parties clearly scoring lower on θ than those of climate policy supporting parties. At the same time, this could also indicate that the model measures some expression of public support. This will be discussed more in depth in the discussion segment.

6 Discussion

In this segment I discuss the results of the thesis. The main points of discussion are mentioned, which includes ideas for future research. Finally, the addition of this paper to existing literature is made clear.

6.1   What is the right label for the latent variable measured?

In this paper, I made an effort to develop a scale of a latent variable through the good governance criteria. The question remains what exactly is measured by this scale. In line with the aim of the study, the intention is to measure perceived legitimacy. There are different indications that what is measured by the scale, is actually perceived legitimacy. First of all, the literature presented at the start of this paper suggests that it might be a measurement of perceived legitimacy. Gribnau (2009), Keping (2011) and Vringer and Carabain (2020) suggest that legitimacy could be connected to good governance. Therefore, when finding a single latent variable from the good governance criteria, there is a suggestion this is perceived legitimacy. Second of all, the comparison to support (directly, as well as indirectly through the background variables) suggests that the scale measures a variable that is close to support. As mentioned before, this was to be expected, as the concepts are closely related. This can be interpreted in two ways. It can be seen as a validation that what is measured is legitimacy, as it is a concept that is close to public support. At the same time, it could also be interpreted as an indication that what we are measuring is an expression of public support, instead of perceived legitimacy.

This discussion is of great importance to my conclusions, as it is fundamental in answering the research question. Overall, the indications that what is measured by the scale is actually perceived legitimacy are present, but not conclusive. Therefore, it is not possible to conclude what is measured exactly, only that it might be perceived legitimacy. To determine what exactly is measured by θ, more validation of this outcome variable is needed. The third part of the results section of this paper could be a starting point for that type of research. For example, an option would be to use the θ-score as a dependent variable in a regression analysis. Using multiple background variables could serve two purposes. It could work as a validation or falsification of θ as an expression of legitimacy and, on the other hand, it could give more information about which groups in society are most positive towards different transition policy interventions.

Even though I cannot conclude what is the right label to attach to the outcome of the scale, its importance to people their perceptions of the policy is proven. The strong correlation to public support indicates so. Therefore, the way the good governance criteria interact within the scale is still relevant, even if what is measured might not exactly be perceived legitimacy of the intervention. The conclusions can still indicate the relevance of certain included good governance criteria for a policy intervention.

6.2   Completeness of the model

It could be questioned to what extent this model is complete for measuring perceived legitimacy of a policy intervention. As mentioned before in the results section, the range of the model is quite narrow, meaning that more extreme perceptions on legitimacy of the proposed interventions cannot be measured confidently. This could be because of the absence of some important aspects of legitimacy.

The conducted survey queried hypothetical policy interventions, which were then rated on different aspects of that proposal by the respondents. Some authors (see Grimes, 2006; Tyler, 2006) suggest that the fair process of decision making in policy writing is very important for the legitimacy of that policy. This process legitimacy is beyond the scope of this paper, but might be of great importance for

overall legitimacy of interventions. The conclusions of this paper solely focus on the content of the policy.

While introducing the operationalization as proposed by Vringer and Carabain (2020), some remarks were made on the completeness of that operationalization and the fact that some good governance criteria were queried in two items (see paragraph 3.3). The factor analysis showed that the double mentioned criteria did not offer any additional problems, therefore, the influence of this is minor and can be neglected. However, accountability was not queried in the operationalization. This could well be because this is process focused and is therefore related to the exclusion of the process legitimacy mentioned before. For the conclusions of this paper, that means the model might not be complete yet.

A suggestion for further research would therefore be to look at actual implemented or proposed policy interventions and include additional items to measure process legitimacy. The good governance criteria could still be of great use here, as many of them can also be operationalized for evaluating the policy process. For example, 'transparency', 'inclusiveness', 'participation' and 'accountability' can all be connected to the policy process with relative ease.

6.3   Generalizability

The included interventions in this paper could all be classified as 'mild' interventions, that have not yet been heated by political debate. For 'mild' interventions it is to be expected that a similar model as presented in this paper is found, while there is no indication what would be the results for more controversial topics. For example, windmills are more of a controversial topic, as the local population of the targeted area is usually quite opposed to the intervention (Poumadère, Bertoldo and Samadi, 2011). Therefore, it could be interesting to conduct a similar research design to a wider variety of interventions. The results of such a design could broaden the generalizability of this paper or find a different mechanism for controversial policy interventions.

6.4   The importance of the different good governance criteria for the scale

The theory presented at the start of this paper, gave some indications on what criteria could be of major importance. An argument was made by Tyler (2010; 2006) and Grimes (2006) that fairness was of great importance to legitimacy. The findings of this paper do not support that hypothesis. Even though all proposed statements could be included in the scale, because they all seem to be relevant, 'equity, inclusiveness and following the rule of law' (the equivalent of fairness) does not have a high explanatory value in the model. An explanation for this difference in perception, is the way fairness was measured in the survey I used. Tyler (2010; 2006) and Grimes (2006) both argue that the process of choosing an intervention and the fairness of that process is of great importance. In this case, there was no information given about the process that led to the choice of policy intervention. Therefore, it was solely based on the fairness of the intervention itself. This could explain the difference in results.

Another point made in the theoretical section of this paper was the role of transparency in legitimacy. Curtin and Meijer (2006) argued that transparency only helps the spread of legitimacy, but does not legitimize policy intervention by itself. Of course, the statistic model presented does not take these types of relationship into account, but transparency does seem to have a big effect in the scale for legitimacy. It could still be the case that the findings of Curtin and Meijer (2006) are correct, but the spread of legitimacy by transparency in that case is of importance for overall legitimacy of the intervention.

Apart from the theory, some more indications on important good governance criteria can be found in the scales. Feasibility was indicated as the most important factor in measuring perceived legitimacy

and might therefore be particularly interesting. Other important aspects were consensus orientation and responsiveness.

6.5 Scientific addition to existing literature

The strength of this paper lies in its contribution in methodological regard. There have been many papers about policy legitimacy and legitimacy of certain interventions, but they have mostly been theoretical and conceptual. The contribution this paper makes is an effort to make perceived legitimacy quantifiable and measurable. The use of an item response theory in public administration research is also a new and unexplored field of research. Typically, item response theory is utilized in psychometrics research, education research and test development (Baker, 2001).

7 Conclusion

This segment will answer the research question: To what extent are good governance criteria appropriate to measure perceived legitimacy of policy interventions? The article by Vringer and Carabain (2020) suggested using the good governance criteria as described by UNESCAP (2009) to operationalize legitimacy. They operationalized this through eight statements that were all related to the good governance criteria. To answer the research question, this operationalization was taken to create a scale to score perceived legitimacy of the interventions on.

I conclude that the operationalization as presented by Vringer and Carabain (2020) is appropriate to measure a single latent concept. This implies that it is possible to create a scale based on these items.

The created scale gives insight in the importance of the different good governance criteria for the latent concept. However, this paper does not give a conclusive answer to the question what is measured by this scale. There are indications in the literature, as well as some in the results of this paper, that point towards perceived legitimacy of the interventions, but no conclusive answer can be given based on this paper. Additional research is required in order to do so.

Furthermore, I conclude that the proposed scale is not yet complete. There are indications in the literature that some aspects of legitimacy are not included in the operationalization, such as process legitimacy or accountability. The results of this paper also show that the range of the scale can still be increased, improving its completeness.

The created scale does show some clear indications for important good governance criteria for policy interventions. Even though it is not certain what is measured by the scale, the strong relationship to public support for the intervention does prove its relevance. Feasibility is one of the main indicators for the scale and therefore important for interventions, and so are transparency, consensus orientation and responsiveness.

This paper stresses the importance of legitimacy for policy evaluation and found clear theoretical reasons to do so. Therefore, I recommend policy evaluators to include legitimacy in their analysis. This paper makes a first attempt at quantifying perceived legitimacy, aiding policy evaluators in a concrete analysis of the concept. Refinement of that scale is needed, but this paper shows that quantifying a concept that is close to legitimacy is plausible.

8 References

Baker, F. B. (2001). *The basics of item response theory* (2nd ed.). ERIC Publications.

Bokhorst, A. M. (2014). *Bronnen van legitimiteit* (Dissertation). Retrieved from: https://research.tilburguniversity.edu/en/publications/bronnen-van-legitimiteit-over-de-zoektocht-van-de-wetgever-naar-z

Bryman, A. (2016). *Social research methods* (5th ed.). Oxford, United Kingdom: Oxford university press.

Carpenter, S. (2018). Ten steps in scale development and reporting: A guide for researchers. *Communication Methods and Measures, 12*(1), 25-44. https://doi.org/10.1080/19312458.2017.1396583

Coicaud, J. M. (2002). *Legitimacy and politics: A contribution to the study of political right and political responsibility*. Cambridge, United Kingdom: Cambridge University Press.

Costello, A. B., & Osborne, J. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical assessment, Research, and Evaluation, 10*(1), 1-9. https://doi.org/10.7275/jyj1-4868

Curtin, D., & Meijer, A. J. (2006). Does transparency strengthen legitimacy? *Information polity, 11*(2), 109-122. https://doi.org/10.3233/IP-2006-0091

De Fine Licht, J., Naurin, D., Esaiasson, P., & Gilljam, M. (2014). When does transparency generate legitimacy? Experimenting on a context-bound relationship. *Governance, 27*(1), 111-134. https://doi.org/10.1111/gove.12021

Graham, J., Amos, B., & Plumptre, T. (2003). *Principles for good governance in the 21st century*. Ottawa, Ontario: Institute On Governance.

Gribnau, J. L. M. (2009). Legitimiteit, recht en governance. In P. Essers, G. Raaijmakers, G. van der Sangen, A. Verdam, & E. Vermeulen (Ed.), *Met recht: Liber Amicorum* (pp. 153-164). Deventer, The Netherlands: Kluwer.

Grimes, M. (2006). Organizing consent: The role of procedural fairness in political trust and compliance. *European Journal of Political Research, 45*(2), 285-315. https://doi.org/10.1111/j.1475-6765.2006.00299.x

Grindle, M. (2012). Good governance: The inflation of an idea. In B. Sanyal, L. Vale, & C. Rosan (Ed.), *Planning ideas that matter* (pp. 259-282). Cambridge, United Kingdom: MIT Press.

Hanberger, A. (2003). Public policy and legitimacy: A historical policy analysis of the interplay of public policy and legitimacy. *Policy Sciences, 36*, 257-278. https://doi.org/10.1023/B:OLIC.0000017471.88620.9a

Jagers, S. C., & Hammar, H. (2009). Environmental taxation for good and for bad: the efficiency and legitimacy of Sweden's carbon tax. *Environmental politics, 18*(2), 218-237. https://doi.org/10.1080/09644010802682601

Kaiser, H. F. (1974). An index of factorial simplicity. *Psychometrika, 39*, 31-36. https://doi.org/10.1007/BF02291575

Keping, Y. (2011). Good governance and legitimacy. In D. Zhenglai, & S. Guo (Ed.), *China's search for good governance* (pp. 15-21). New York, New York: Palgrave Macmillan.

Lameijer, C. M., van Bruggen, S. G. J., Haan, E. J. A., Van Deurzen, D. F. P., van der Elst, K., Stouten, V., … Terwee, C. B. (2020). Graded response model fit, measurement invariance and (comparative) precision of the Dutch-Flemish PROMIS® Upper Extremity V2. 0 item bank in patients with upper extremity disorders. *BMC Musculoskeletal Disorders, 21*(1), 1-17. https://doi.org/10.1186/s12891-020-3178-8

Leclezio, L., Jansen, A., Whittemore, V. H., & de Vries, P. J. (2015). Pilot validation of the tuberous sclerosis-associated neuropsychiatric disorders (TAND) checklist. *Pediatric neurology, 52*(1), 16-24. https://doi.org/10.1016/j.pediatrneurol.2014.10.006

Matti, S. (2009). *Exploring public policy legitimacy: a study of belief-system correspondence in Swedish environmental policy* (Dissertation). Retrieved from http://urn.kb.se/resolve?urn=urn:nbn:se:ltu:diva-17742

McCright, A. M. (2008). The social bases of climate change knowledge, concern, and policy support in the US general public. *Hofstra Law Review, 37*(4), 1017-1047.

Montenegro de Wit, M., & Iles, A. (2016). Toward thick legitimacy: Creating a web of legitimacy for agroecology. *Elementa: Science of the Anthropocene, 4*, 1-24. https://doi.org/10.12952/journal.elementa.000115

Nas, T. F. (2016). *Cost-Benefit Analysis: Theory and Application* (2nd ed.). Lanham, Maryland: Lexington books.

OECD. (2010). *The State's Legitimacy in Fragile Situations*. Paris, France: OECD Publishing. https://doi.org/10.1787/9789264083882-en

PBL, SCP & CPB. (2017). *Naar een Verkenning Brede welvaart*. Den Haag: Planbureau voor de Leefomgeving, Sociaal en Cultureel Planbureau en Centraal Planbureau.

Poumadère, M., Bertoldo, R. and Samadi, J. (2011). Public perceptions and governance of controversial technologies to tackle climate change: nuclear power, carbon capture and storage, wind, and geoengineering. *WIREs Climate Change, 2*(5), 712-727. https://doi.org/10.1002/wcc.134

Scholz, J. T. (1984). Voluntary compliance and regulatory enforcement. *Law & Policy, 6*(4), 385-404. https://doi.org/10.1111/j.1467-9930.1984.tb00334.x

SCP. (2017). *Kwesties voor het kiezen*. Den Haag, The Netherlands: Sociaal en Cultureel Planbureau. Retrieved from: https://www.scp.nl/publicaties/publicaties/2017/02/13/kwesties-voor-het-kiezen

Smoke, R. (1994). On the importance of policy legitimacy. *Political Psychology, 15*(1), 97-110. https://doi.org/10.2307/3791441

Toland, M. D. (2014). Practical guide to conducting an item response theory analysis. *The Journal of Early Adolescence, 34*(1), 120-151. https://doi.org/10.1177/0272431613511332

Tyler, T. R. (2006). Psychological perspectives on legitimacy and legitimation. *Annual Reviews Psychology, 57*, 375-400. https://doi.org/10.1146/annurev.psych.57.102904.190038

Tyler, T. R. (2010). Legitimacy in corrections: Policy implications. *Criminology & Public Policy, 9*(1), 127-134.

UNESCAP (2009) *What is Good Governance?* Retrieved from:
https://www.unescap.org/resources/what-good-governance

United Nations. (2015). *Paris Agreement.* Retrieved from: https://unfccc.int/process-and-meetings/the-paris-agreement/the-paris-agreement

Vringer, K., & Carabain, C. L. (2020). Measuring the legitimacy of energy transition policy in the Netherlands. *Energy Policy, 138*, 1-8. https://doi.org/10.1016/j.enpol.2019.111229

Wallner, J. (2008). Legitimacy and public policy: Seeing beyond effectiveness, efficiency, and performance. *Policy Studies Journal, 36*(3), 421-443. https://doi.org/10.1111/j.1541-0072.2008.00275.x

Weber, Max (1978) *Economy and Society* (G. Roth, & C. Wittich, Trans.). Los Angeles, California: University of California Press.

9 Appendix

Table 9.1 Shapiro-Francia W' test for normal data intervention 1

Variable	Obs.	W'	V'	z	Prob>z
Helps	1,988	0.99	17.26	6.82	0.00
Efficient	1,976	0.99	11.05	5.75	0.00
C. R. Implement	1,996	0.99	13.04	6.15	0.00
C. R. Set-up	2,029	0.99	13.09	6.16	0.00
Feasible	1,994	0.99	6.82	4.60	0.00
Transparent	1,994	1.00	5.53	4.09	0.00
Fair	2,002	0.98	25.22	7.73	0.00
Participatory	1,983	0.99	9.92	5.49	0.00

Table 9.2 Shapiro-Francia W' test for normal data intervention 2

Variable	Obs.	W'	V'	z	Prob>z
Helps	2,000	0.98	21.19	7.31	0.00
Efficient	1,894	0.99	6.40	4.43	0.00
C. R. Implement	1,992	0.98	22.61	7.47	0.00
C. R. Set-up	2,038	0.98	23.68	7.59	0.00
Feasible	2,005	0.99	8.08	5.00	0.00
Transparent	1,976	0.99	9.05	5.27	0.00
Fair	2,003	0.99	15.31	6.53	0.00
Participatory	1,978	0.99	9.98	5.50	0.00

Table 9.3 Shapiro-Francia W' test for normal data intervention 3

Variable	Obs.	W'	V'	z	Prob>z
Helps	2,016	0.98	28.92	8.06	0.00
Efficient	1,873	1.00	2.41	2.09	0.02
C. R. Implement	1,989	0.99	12.48	6.04	0.00
C. R. Set-up	2,060	0.98	19.52	7.13	0.00
Feasible	2,045	0.98	19.72	7.15	0.00
Transparent	2,028	0.99	6.61	4.53	0.00
Fair	1,995	0.99	11.64	5.88	0.00
Participatory	1,995	0.99	12.32	6.01	0.00

Table 9.4 Shapiro-Francia W' test for normal data intervention 4

Variable	Obs.	W'	V'	z	Prob>z
Helps	2,030	0.98	29.11	8.08	0.00
Efficient	1,860	1.00	1.50	0.97	0.17
C. R. Implement	1,957	0.99	7.28	4.75	0.00
C. R. Set-up	2,023	0.99	10.58	5.65	0.00
Feasible	1,999	1.00	6.11	4.34	0.00
Transparent	1,990	1.00	3.94	3.28	0.00
Fair	1,971	1.00	4.06	3.35	0.00
Participatory	1,983	0.99	10.39	5.60	0.00

Table 9.5 Graded response model Intervention 1

Items	Discrimination a	Thresholds b1	b2	b3	b4
1	2.293 (.087)	-1.292 (.051)	-0.671 (.038)	0.101 (.034)	1.416 (.053)
2	2.768 (.104)	-1.073 (.043)	-0.425 (.033)	0.171 (.032)	1.232 (.046)
3	3.380 (.130)	-1.140 (.042)	-0.546 (.032)	0.422 (.030)	1.051 (.040)
4	2.861 (.108)	-1.408 (.050)	-0.789 (.037)	-0.064 (.031)	0.940 (.040)
5	3.440 (.132)	-1.090 (.041)	-0.507 (.031)	0.201 (.030)	1.186 (.042)
6	3.086 (.118)	-0.835 (.037)	-0.205 (.030)	0.434 (.032)	1.423 (.049)
7	2.180 (.084)	-1.756 (.066)	-1.167 (.049)	-0.279 (.035)	0.951 (.044)
8	2.668 (.100)	-1.137 (.045)	-0.575 (.045)	0.291 (.033)	1.429 (.051)

Note: a = the slope of the characteristic curve and therefore discrimination (SE). b1 = threshold for an answer of >= 2 (SE). b2 = threshold for an answer of >= 3 (SE). b3 = threshold for an answer of >= 4 (SE). b4 = threshold for an answer of = 5 (SE).

Table 9.6 Graded response model Intervention 3

Items	Discrimination a	Thresholds b1	b2	b3	b4
1	2.247 (.087)	-1.632 (.061)	-0.961 (.044)	-0.266 (.034)	1.042 (.046)
2	2.536 (.097)	-1.223 (.049)	-0.358 (.034)	0.423 (.035)	1.469 (.054)
3	3.021 (.116)	-1.119 (.043)	-0.448 (.032)	0.077 (.030)	1.073 (.042)
4	2.616 (.100)	-1.634 (.058)	-0.980 (.042)	-0.272 (.032)	0.699 (.037)
5	3.182 (.126)	-1.241 (.045)	-0.646 (.034)	-0.145 (.030)	0.687 (.035)
6	3.153 (.122)	-0.932 (.039)	-0.280 (.030)	0.320 (.031)	1.250 (.045)
7	2.672 (.101)	-1.217 (.047)	-0.576 (.035)	0.153 (.032)	1.276 (.048)
8	2.621 (.099)	-1.275 (.048)	-0.638 (.036)	0.147 (.032)	1.300 (.049)

Note: a = the slope of the characteristic curve and therefore discrimination (SE). b1 = threshold for an answer of >= 2 (SE). b2 = threshold for an answer of >= 3 (SE). b3 = threshold for an answer of >= 4 (SE). b4 = threshold for an answer of = 5 (SE).

Table 9.7 Graded response model Intervention 4

Items	Discrimination a	Thresholds b1	b2	b3	b4
1	1.844 (.078)	-2.514 (.104)	-1.649 (.068)	-0.743 (.043)	0.849 (.046)
2	2.084 (.084)	-1.457 (.059)	-0.367 (.037)	0.594 (.040)	1.819 (.069)
3	2.922 (.117)	-1.586 (.055)	-0.709 (.036)	0.058 (.031)	1.220 (.046)
4	2.406 (.095)	-1.997 (.073)	-1.085 (.045)	-0.202 (0.33)	1.052 (.045)
5	3.219 (.131)	-1.564 (.053)	-0.780 (.036)	0.043 (.030)	1.076 (.041)
6	2.665 (.104)	-1.266 (.048)	-0.387 (.033)	0.422 (.034)	1.637 (.057)
7	1.520 (.066)	-1.888 (.082)	-0.684 (.047)	0.468 (.044)	2.183 (.093)
8	2.100 (.083)	-1.617 (.062)	-0.808 (.042)	0.151 (.035)	1.531 (.059)

Note: a = the slope of the characteristic curve and therefore discrimination (SE). b1 = threshold for an answer of >= 2 (SE). b2 = threshold for an answer of >= 3 (SE). b3 = threshold for an answer of >= 4 (SE). b4 = threshold for an answer of = 5 (SE).