Validating citizenship indicators

Marc Helbling
Social Science Research Center Berlin (WZB)
helbling@wzb.eu
www.marc-helbling.ch

Abstract: After a long period of case study research and small-N comparisons, over the recent past various scholars have started to compare a relatively large range of countries by constructing citizenship indicators. Since there are almost as many indices as there are large-N studies one might wonder whether it matters which indicators we use. Surprisingly, the emerging pioneer spirit has left little space to methodological questions that would enable us to answer this question. Various validity tests will show that most of the indicators consist of different components and are weakly related to each other even if they can be explained by similar factors. Differentiating different phases of a policy process it will further be shown that public debates are not directly translated into policies, which in turn do not always lead to the intended outcomes. It will thus appear that we need to be extremely careful when we choose our indicators.

(August 17, 2011)
INTRODUCTION

Over the recent period of the last half-decade an important development in the citizenship literature took place. After a long period of case study research and small-N comparisons, various scholars and projects have started to compare a relatively large range of countries by building citizenship indicators. In the second half of the 1990s Waldrauch and Hofinger (1997) presented a first “Index of Legal Obstacles to Integration” (LOI) that covers eight Western European countries for the year 1995.

A real boom of citizenship indicators occurred in the second half of the 2000s. Koopmans et al. (2005; 2010) presented two indicators measuring the restrictiveness of nationality acquisition (or individual equality (Koop_IE)) and cultural rights attribution (or cultural difference (Koop_CD)). In a first project data for five Western European countries and three time periods (1980, 1990, 2002) have been collected (Koopmans et al. 2005). More recently, this dataset has been expanded to ten Western European countries and a fourth time period (2008) (Koopmans et al. 2011).


1 By indicator I understand a measurement that operationalizes in a quantitative way a social phenomenon and represents an aggregate of data. While various authors use different terms like ‘indicator’, ‘index’ or ‘dimension’, I use these terms interchangeably to define the measurements at their most aggregate level and speak of ‘sub-indicators’ if I refer to specific components of an indicator. By ‘variable’ or ‘item’ I mean the most basic element of an indicator as defined by the authors.

2 For an overview of all indicators see in the Appendix Table A1.
dex” (CIVIX) measures language, country-knowledge and value commitment requirements across the EU-15 for the years 1997 and 2009.

Finally, the “Migration Integration Policy Index” (MIPEX) (MPG 2006; Huddleston and Niessen 2011) includes all member states of the European Union plus Norway, Switzerland, Canada and the USA for the years 2007 and 2010. Besides these studies that base their comparisons mostly on policy indicators, Janoski (2010) has focused on outcome variables and uses naturalization rates to test his arguments in 18 OECD countries for the period 1970 to 2005.

Surprisingly, the emerging pioneer spirit has left little space to methodological questions; especially validity questions have hardly been addressed. A discussion of whether the new indicators measure what they are supposed to measure is however all the more important as there are almost as many indices as there are large-N studies. In order to use these indicators in an efficient way we need to know how good they are, what they really measure and whether it matters which one we apply for a specific research question.

The main goal of this paper is therefore to encourage researchers to think more systematically about how to build and use indicators. As we know from other study fields there is no agreed upon criterion for determining when an indicator is valid (enough). There is no rule that tells us how strongly an indicator needs to be related to other measurements in order to be considered valid. Therefore there will be no concluding discussions about how valid the existing indicators really are and whether any of these indices are better than others.

By comparing indicators we however get aware of commonalities and differences; we

---

3 See however the EUDO Citizenship Forum debate on citizenship indicators on which this contribution builds: http://eudo-citizenship.eu/citizenship-forum.
learn what is specific about an indicator and what is unclear. In the end, it strongly depends on the specific research question whether or not one particular indicator is preferable to the others for the measurement of citizenship policies. Or as Janoski (2011) has put it: “What we need citizenship indicators for depends on who are “we””. Different research projects require different data. Are we interested in public debates on citizenship issues, policy outputs, policy outcomes or the effects of one element on the other?

As obvious as it might sound that indicators need to be built for specific research purposes, it is all the more astonishing that in most works it is hardly ever specified what exactly is under study and why certain items are retained to build an indicator and not others. It is often not explained why new indicators are built instead of using exiting ones or what the advantage of one index over another one is. Exceptions can be found in Koopmans et al. (2005: 7-16) and Goodman (2010: 756-757). Both present two-dimensional spaces of nationhood conceptions and policies and thereby distinguish crucial aspects that according to them should not be confounded. While they provide thorough theoretical discussions, they however do not show us how similar/different their dimensions are empirically to justify a two-dimensional space and validate their conceptions.

To encourage a more rigorous discussion on citizenship indicators I pursue two strategies in this paper. In the first part, I organize my discussion around classic validation tests. Discussing the content, convergent and construct validity of the existing indicators allows us to see to what extent the indicators are related to each other and measure the same underlying theoretical constructs. In the second part, I then differentiate between various phases of a policy process to make clear how important it is to be specific about the phases of a policy cycle we are interested in.
Besides comparing citizenship indicators and encouraging a more systematic discussion this paper also pursues a more analytical aim. Comparing, on the one hand, indicators that are based on different policy fields as we will see and different phases of a policy-making process the obvious question is whether we find the same dominant ideologies in different areas within a country. Might it be that a certain understanding of citizenship is inscribed in a country and affects integration and naturalization policies, public debates and implementation processes in the same way? This question relates to the literature on citizenship models and the idea that they constitute broader societal phenomena that can also be observed beyond formal and legal regulations (see Helbling 2010; Michalowski and Finotelli 2012).

Speaking of cultural idioms Brubaker’s (1992) argument on the difference between France and Germany implies that citizenship models constitute more than formal regulations and also reflect ways of thinking and talking about nationhood. In other words, the way cultural boundaries of a nation are drawn in national regulations reflects some deeper societal understanding of what the nation is. If this were the case, we would observe these cultural idioms in different arenas and high correlations between the various indicators. Even if we concede that models change over time we should still observe high inter-correlations as for example new ideas are discussed in public debates that are then translated into new policies and implemented accordingly. In other words, if there are such strong cultural idioms, we do not need to care that much about which indicators we use for our analyses as they would all measure the same theoretical constructs.

However, as we will see, this is not at all the case. It makes a difference, for example, whether an indicator includes naturalization laws or cultural rights that are attributed to immigrants and whether we investigate public debates on immigration issues or
concrete policy outputs. I thereby follow those researchers that started to question the idea of citizenship models and argue that models are too simplistic as they ignore variation within a country (Bader 2007: 875-877).

Before I start, I would like to clarify that I only address validity and not reliability questions in this paper. This is not to say that reliability that “concerns the extent to which an experiment, test, or any measuring procedure yields the results on repeated trials” (Carmines and Zeller 1979: 11) is of lesser importance. We certainly all agree that if we use indicators we need to make sure that they are both reliable and valid. Given the limited space of such a paper I however prefer to start with a discussion on validity that is about the interrelationship between a concept and an indicator and the question of whether or not an indicator measures what it is intended to measure (Carmines and Zeller 1979: 12).

Thinking about the validity of an indicator forces us to be careful about our intentions and to clarify what we seek to measure. Thus, we not only need to discuss how to build valid indicators. We also have to be clear what we want to measure. This is why for Cronbach (1971: 447) validations are not so much about the tests and measurements themselves as about the “interpretation of data arising from a specific procedure.”

For this reason I am interested here in which policy fields and which phases of a policy cycle are covered by an indicator. I do not discuss the equally important questions of data collection and indicator construction. A source of contestation is for example that the scoring process of the MIPEX data is rather subjective and opaque (Howard 2009: 34-35). Another problem concerns operationalization and the questions of which aspects are included in an indicator and how they are coded. It is often not clear why certain items have been chosen and not others. Howard (2006: 446-447), for ex-
ample, simply states, that his criteria constitute the most general elements of a country’s citizenship policy. Koopmans et al. (2005: 33) only provide some very general rules how to reasonably choose empirical indicators (only choose indictors that do vary across cases, exist at least in most of them and have the same meaning).

The number of variables that are covered by an indicator and which concrete values are attributed also affects the validity of an indicator – but again, this is of the domain of data collection. The discussion by Michalowski and Van Oers (2011) of the Civic Integration Index (CIVIX) (Goodman 2010), for example, shows how coding choices can influence the classification of citizenship policies. And while at first sight it seems clear what naturalization rates measure, Janoski (2010) and Reichel (2011) show that they can be conceptualized in very different ways. Finally, there are also scholars who prefer to focus on the sociology of citizenship policies, i.e. problematization of the processes of policy creation and policy discourses, instead of simply measuring policies (Duyvendak et al. 2011). As important as all these aspects are, this is material for another paper.

FIRST PART: VALIDITY TESTS

Criterion validity

To discuss the validity of citizenship indicators I follow Bollen’s (1989) lead who distinguishes four major approaches: Criterion validity, content validity, convergent validity and construct validity (see also Carmines and Zeller 1979: 17-27).

Let us start with criterion validity that we unfortunately cannot assess for different reasons. There are two slightly different understandings of criterion validity. First, criterion validity (sometimes also called concurrent validity) might involve the com-
parison of an indicator with some other generally accepted measure of the same concept (Carmines and Zeller 1979: 18; Ray 2007: 12). If there was already a citizenship indicator that was accepted by the research community as the most valid measurement, the validity of new indices could simply be assessed by comparing them. However, since researchers have just started to measure citizenship policies, it is not surprising that there is so far no such generally accepted indicator.

Second and according to Nunnally (1978: 87, quoted in Carmines and Zeller 1979: 17) criterion-related validity “is at issue when the purpose is to use an instrument to estimate some important form of behavior that is external to the measuring instrument itself, the latter being referred to as the criterion.” Typically applied in psychology and education criterion validity (sometimes also called predictive validity) is for example relevant to assess tests and selection procedures. Carmines and Zeller (1979: 18) give the example of a test to screen applicants for police work that could be validated by correlating test scores with future performance of the selected persons. Adapted to our field one could compare a policy with its outcome. Does an integration policy lead to a certain degree of integration and a naturalization policy to a certain number of naturalized citizens? However, as much as it is difficult to define the performance of policemen, as much is it often unclear what the performance of a policy is expected to be. Moreover and as we will see below there are crucial differences between policy outputs and outcomes, and it is therefore better to clearly distinguish the two aspects.

It appears that it becomes almost impossible to assess the criterion validity of citizenship indicators and that such tests need to be precluded at the current stage of citizenship research. It is also questionable whether it will ever be possible to investigate the criterion validity of citizenship indicators as this is particularly difficult for abstract
social phenomena especially in the social sciences (1979: 19-20). After all, citizenship indicators are also expected to reflect some more general citizenship models. In that regard Koopmans et al. (2005) and Goodman Wallace (2010) are most explicit when they present four ideal-typical citizenship regimes, which they intend to measure with their indicators. However, it is not only difficult to clearly define what, for example, a universalist or multicultural citizenship model is. It is all the more challenging to agree on concrete criteria that need to be measured to operationalize these concepts. As we will see in the next part, this problem can at least partly be solved by clarifying the domain that is supposed to be measured and thus by assessing the content validity of an indicator.

**Content validity**

Content validity involves a theoretical discussion to test the validity of an indicator by clarifying the domain of a concept and a judgment of whether the measure fully represents the domain (Bollen 1989: 185). Thus, we first need to clarify what the domain is we are investigating and how we define citizenship policies. As Coppedge and Gerring (2011: 249) point out: “What is clear is that methodological problems affecting contemporary indices begin at the level of definition.”

To better understand the content of the various existing indicators I first differentiate between policy outputs and policy outcomes (see Table 1). While some indicators focus mostly on the policy outputs and thus on the concrete rules and guidelines, others are interested in the consequences of these policies and thus their outcomes. In the literature on citizenship indicators this differentiation is however seldom made ex-
plicit, and it is often not clear why one indicator is preferred over another one and whether or not they are used to answer the same research questions.

Moreover, the two policy aspects are sometimes combined. Both Howard (2009) and Koopmans et al. (2005; 2010) use naturalization rates as part of their policy (output) indicators (see Table 1). Howard (2009: 24) uses naturalization rates as a “correction” that helps him “to account for the potential problem of a country appearing to have a very inclusive naturalization policy, but in reality – whether due to administrative “discretion” or other barriers or disincentives – being much more restrictive in practice.” For exactly the same reason Koopmans et al. (2005: 38; 2010: 8 note 5) have included naturalization rates in their indicator. They argue that there are many informal factors that might affect access of migrants to the nationality of the host country. (In their most recent analysis they have however excluded the naturalization rates (Koopmns et al. 2011)). By including naturalization rates both studies are at the same time pointing to and blurring the crucial differentiation between policy outputs and outcomes (see Janoski 2010: 36). It thus becomes unclear what they are actually measuring. While they are completely right to point out that policies do not necessarily lead to the intended outcomes due among others to diverging implementation processes, comparing the two phases of such a policy process would present an interesting research question in itself (see below).

Not only is the difference between policy outputs and outcomes often blurred. Among the two groups various aspects of citizenship policies and their consequences are retained to build an indicator. To better understand the content of the various indicators I first differentiate between immigration, integration and naturalization policies and in a second step between naturalization and rejection rates.
Policy outputs: Immigration, integration and naturalization policies

A differentiation is often made between immigration and immigrant policies that reflect the two modes of social closure of nation-states, one at the territorial border and one inside the territorial border (see Weber 1946: 78; Brubaker 1992: ch.1; 2010). While immigration policies turn around the question of who is granted access to the territory of a state, immigrant policies are concerned with those people that have already crossed the national borders.

Hammar (1990: 21) sub-divides immigrant policies into regulations of status of denizens and citizenship. He thereby differentiates between three gates at the entrance of states in Western Europe: entry, settlement and full membership. While ‘entry’ concerns immigration policies, ‘settlement’ refers to regulations of status of denizens, by which Hammar (1990: 13) understands foreign residents that have gained a secure residence status without however being naturalized citizens. The respective regulations can also be called integration policies that stipulate what is expected from immigrants, which conditions they have to meet to remain in a country and to become a part of the society, and the rights they are attributed. The main difference between ‘settlement’ and ‘full membership’ are that the latter means becoming a naturalized citizen and receiving all political rights. In other words, immigrant policies can be divided into integration and naturalization policies.

These different understandings of immigration and especially immigrant policies are also reflected in the various existing citizenship indicators as we see in Table 1. Howard’s CPI indicator only includes elements that are related to formal requirements for having access to citizenship (Howard 2009: 4). More particularly he measures jus soli and dual citizenship regulations as well as naturalization requirements (Howard 2009: 20-26).
While Howard’s CPI indicator focuses on naturalization policies, Banting’s and Kymlicka’s MCP-indicator refers exclusively to integration policies and more specifically on cultural rights that are accorded to immigrants. With their indicator Banting and Kymlicka (2011: 6) seek to measure the degree of “public recognition, support or accommodation for ethnocultural minorities to maintain and express their distinct identities and practices.” (see also Banting and Kymlicka 2006b: 1).

Koopmans et al. (2005; 2011) combine these two aspects and add some new elements. They develop a two-dimensional space of citizenship policies that permits to account for both how discriminatory the individual access to citizenship is and to what extent cultural group rights are granted. The first dimension (individual equality (Koop_IE)) measures how restrictive nationality acquisition regulations are, but also includes individual rights for foreign nationals (such as protection against expulsion and voting rights) and anti-discrimination rights. The second dimension mainly covers cultural rights (cultural rights outside and within public institutions, affirmative action etc.) (cultural difference (Koop_CD)). More recently they also included regulations concerning family reunification and access to public service employment (Koopmans et al. 2011).

Waldrauch and Hofinger’s (1997) indicator also combines various aspects of immigrant policies. Besides naturalization regulations their indicator includes elements that concern residence regulation, labor market access, family reunifications and the legal status of the 2nd generation immigrants. In a similar way, the seven MIPEX sub-indicators cover labor market mobility, family reunification, long-term residence and education. Moreover, the MIPEX indices also measure policies pertaining to political participation, anti-discrimination policies and access rights to nationality.
Goodman’s (2010) Civic Integration Index covers not only one or two but three policy fields. It consists of four and eight variables for the years 1997 and 2009 respectively covering especially language requirements at all three gates of entrance (immigration, integration, naturalization).

We see that with regard to policy fields some indicators are very parsimonious and cover specific aspects like naturalization rights (CPI) or cultural rights (MCP) whereas others include both naturalization and integration aspects (Koopmans et al., MIPEX, LOI) or even all three policy fields (CIVIX). Moreover and as we have seen above, the indicators by Howard (2009) and Koopmans et al. (2005; 2010) also include policy outcomes.

The fact that so many aspects are included in some of the indicators (but to different degrees) makes it important to discuss to what extent these indicators measure similar concepts. It is especially important to discuss whether we can combine immigration, integration and naturalization requirements in the same indicator or whether it might be better to treat them as separate indicators (see Coppedge and Gerring 2011: 250-251).

While both the CPI and MCP indicators are limited to one specific (output) sub-domain, the other indicators provide sub-indicators for several sub-domains and thus allow investigations at different levels of aggregation. Whether or not we aggregate depends in the end on our research question. Do we analyze immigrant policies, naturalization policies, integration policies or even a particular form of integration policies such as for example cultural rights that are accorded to immigrants? In the end, it makes only sense to assess the content validity of a measurement when the domain under study is clearly defined and the definition is accepted by at least a majority of the researchers working in this field (Carmines and Zeller 1979: 22).
Policy outcomes: Naturalization and rejection rates

While the studies we discussed in the previous part look (mostly) at policy outputs, another group prefers to take naturalization or rejection rates as a dependant variable and thus to focus on policy outcomes. To measure the outcome of citizenship politics, naturalization rates are most often used, which is the ratio between the yearly number of naturalisations and the number of foreign residents in a country. Surprisingly, so far hardly any studies have investigated naturalization rates across a large range of countries (see however Dancygier 2011). There are however some studies that compared and aimed at explaining naturalization rates at the regional or local level in federal states (Dornis 1999; Bultmann 1999; Hagedorn 2001a,b; Ludwig 2004, 2001; Bolliger 2004; Thränhardt 2008: 15-26).

In his recent work, Janoski (2010) brings up a new (what he calls a more comprehensive) way to measure naturalization rates. Instead of just looking at the number of people who have passed the official naturalization procedures, he also accounts for different naturalization regimes, some of which allow some people to "circumvent" the naturalization processes. For example, in countries with a jus soli policy citizenship is automatically conferred to children of resident immigrants. Janoski considers this an alternative form of naturalization and accordingly adjusts his indicator, among others, by adding up the number of naturalized people and the number of people who receive the nationality at birth.
The more general question is however whether naturalization rates reflect effects of citizenship policies at all. In the same paragraph where Howard (2009: 24) explains why he accounts for naturalization rates he states that they “should not be relied on too closely, since the rates may depend significantly on the provenance of the immigration population, as well as the demographic patterns of immigration […].” Howard is pointing here to the crucial content validity problem of naturalization rates that depend on both the demand and the supply side. For a low naturalization rate might be explained by a restrictive naturalization policy, and/or by the fact that there is low demand on the part of the foreign population. Moreover, since the ratio is measured within the overall foreign-resident population, changes of the naturalization rate might tell us more about immigration flows (the denominator in the calculation) than about the desire of foreign residents to become naturalized (see Ludvig 2004).

For these reasons, in my own study I preferred to measure rejection rates, that is, the ratio between the rejected and the submitted applications (Helbling 2008). Contrary to the naturalization rate the rejection rate almost exclusively depends on the supply side. There are however still two shortcomings: It does not consider that certain candidates interrupt their applications during the procedure and that some alien residents might not apply because they do not expect to be naturalised in settings with a restrictive naturalisation policy (deterrence effects).

Convergent validity

Given the large variety of aspects and policy domains that are covered by the different indicators one might wonder to what extent the indices that focus mainly on policy aspects are related to each other. Convergent validity involves the comparison of al-
ternative measures of the same concepts. By means of correlation tests it is evaluated whether or not two or more indicators are similar and thus converge. Convergent validity (for Carmines and Zeller (1979) this is no separate category) is tightly related to content validity as high correlations might be a sign that two or more indicators reflect the same content and thus measure the same underlying concept.

By means of correlation tests, both Howard (2009: 32-35) and Koopmans et al. (2011) compare their own indicators with each other and in addition make a comparison with the Migration Integration Policy Index (MIPEX) (MPG 2006) and, but only Koopmans et al., with the Legal Obstacles to Integration Index (LOI) (Waldrauch and Hofinger 1997). Since all but Howard’s indicator also include information on integration regulations, only the components that are directly related to naturalization regulations are accounted for in the two tests. Howard (2009) and Koopmans et al. (2010) run slightly different correlation tests, but both of them come to the conclusion that all indicators are all highly related to each other and thus seem to measure the same concept in a consistent way.

In the first part of Table 2 I have replicated these findings by comparing the indicators that measure more specifically aspects related to naturalization rights. Whereas I have taken Howard’s CPI indicator as it stands, for the MIPEX and Koopmans data I have retained the sub-indicators that measure more specifically naturalization rights (Koop_Nat, MIPEX_Nat). It appears again that these indicators measure pretty much the same (correlations between 0.86 and 0.91).

While it is not surprising that the naturalization indicators are highly correlated, one might wonder how the overall indicators are related to each other, as they have been conceptualized in different ways (see Table 1). In the second part of Table 2 I look at

---

4 For CIVIX and LOI no specific naturalization sub-indicators exist.
Goodman’s CIVIX, Howard’s CPI, the two Koopmans dimensions as well as at Banting’s/Kymlicka’s MCP, Waldrauch’s/Hofinger’s LOI and the MIPEX indicators. Contrary to the naturalization indicators the strength of correlation varies a lot when we compare the overall indicators. I present the correlation coefficients in four groups and consider correlations above 0.7 (in bold) as high. We see relatively high correlations among the first two groups (values between 0.69 and 0.85). The first group includes indicators of which both include naturalization policies (Koop_IE, CPI, LOI, MIPEX). These are the expanded indicators that we have already compared in the first part of Table 2. If we compare in the second group the two indicators that mainly consist of cultural rights elements (Koop_CD, MCP) the relationship is also relatively strong (0.72).

However, if we correlate indicators that mainly focus on naturalization policies with the cultural rights indicators (third group), correlations are clearly weaker (with one exception between 0.55 and 0.66). The only exception to this general pattern is the high correlation between the Koop_IE and MCP indices. Correlations get fully insignificant when we compare in the fourth group the Civic Integration Index, that also includes immigration requirements, with the other indicators (correlations between 0.01 and 0.37).

It thus appears that not all indicators measure the same underlying theoretical concept. First of all, there is a difference between indicators that focus at least to a certain extent on naturalization policies (CPI, Koop_IE, MIPEX, LOI) and those that mostly cover cultural rights aspects (Koop_CD, MCP). Second, the indicator that covers policies at all three entrance gates (CIVIX) is clearly different from the other indicators. Although the indicators do not always clearly differentiate between immigration, integration and naturalization policies as most indices cover more than one field, it
becomes quite clear that the three policy fields are weakly related to each other at best. In other words, a liberal integration policy does not necessarily lead to generous naturalization regulations and different cultural idioms are present in the various policy fields.

-- Table 2 about here --

**Construct validity**

The last way to evaluate whether or not an indicator measures what it is supposed to measure concerns construct validity that “assesses whether a measure relates to other observed variables in a way that is consistent with theoretically derived predictions.” (Bollen 1989: 188; see also Casper and Tufis 2003).

Among the indicators that are presented in this paper three have been used as dependent variables. One of the main goals of Howard (2009), Janoski (2010) and Koopmans et al. (2011) was to investigate the variance of their indicators and to explain why certain countries pursue more restrictive citizenship policies than others, respectively why naturalization rates are higher or lower. While Koopmans et al. (2010) analyze the levels of rights, Howard (2009) is interested in policy changes and Janoski (2010) investigates the naturalization rates at specific points in time.

The MIPEX and MCP data have so far only been used as independent variables (see Banting et al. 2006; Crepaz 2006; Hooghe et al. 2009; Kessler and Bloemraad 2010; Dinesen and Hooghe 2010; Fleischmann and Dronkers 2010; McLaren 2010). To my knowledge the LOI data have never been used in a quantitative multivariate analysis.
The three studies that have aimed at explaining the output or outcome of citizenship policies (or the changes thereof) have developed a very similar explanatory framework. Although they have used different data, applied different methods and a large variety of explanatory and control variables, crucial aspects of their research designs are very similar. First of all, they all look simultaneously at long and short-term explanatory factors (see Table 3). As of the long-term factors, both Howard (2009) and Janoski (2010) investigate the impact of colonialism on naturalization policies and rates, respectively. They argue and show that countries that have had a long colonization experience were early on confronted with the problem of how to incorporate natives/immigrants and thus developed liberal naturalization regimes. Howard (2009) develops a second long-term argument and shows that early democratizing countries developed a civic national identity tied to liberal values.\footnote{According to Janoski (2010: 21 n.19) democratization overlaps to a large extent with colonization and thus causes multicollinearity problems.}

Koopmans et al. (2011) also develop a long-term argument even if empirically they do not go back as far as Howard and Janoski. In a Brubakerian perspective and against a widely held opinion they show that no convergence of citizenship policies occurred in Western Europe over the last decades. On the contrary, it appears in their analysis that the policies at the end of the 2000s are best explained with the situation in 1980.

At the subnational level in Switzerland two studies have also revealed the impact of long-term factors on citizenship indicators. In my own work I show that the local understanding of citizenship measured by means of voting results over two decades has an important impact on rejection rates (Helbling 2008). In Manatschal’s (2011) study traditional understandings of citizenship and belonging play a crucial role at explaining integration policies at the regional level.
While the long-term variables differ across the studies, the short-term variables in all three cross-country studies are almost the same. They all aim at explaining short-term changes by looking at the influence of political parties. At the same time, all three studies show that socio-structural and socio-economic factors such as immigration and unemployment rates (or GDP) fail to explain the respective dependent variables. Koopmans et al. (2011) find a significant effect of the share of immigrant-origin voters.

Howard (2009) asks whether widely existing anti-immigrant sentiments become activated politically by the far right, which in turn might dampen liberalization processes. He indeed finds a quite strong relationship between the strength of far right parties and citizenship liberalization. Janoski (2010) prefers to focus on the actors promoting naturalization and reveals that increasing left party power in the post-World War II period leads to higher naturalization rates. Koopmans et al. (2010) tested a similar argument but did not find evidence to support the argument according to which more liberal citizenship policies occur under left-wing governments. They however observed that more restrictive policies are due to electoral pressures from right-wing populist parties.

In my own study I show that local political struggles among political actors and especially the strength of the right-wing populist forces lead to specific national self-understandings within a municipality, which in turn explains to a high degree local rejection rates (Helbling 2008). Contrary to all studies discussed here, Manatschal
(2011) did not find any evidence that the electoral success of the right-wing populist Swiss Peoples Party (SVP) led to more restrictive integration policies.

Given the fact that all studies used different data and applied different methods the finding with regard to the role political parties play seems to be quite robust. Such a result is not at all self-evident; especially in the light of the content and convergent validity tests that have revealed important differences between the indicators. For example the CPI and Koopmans et al. indicators that can both be explained by the strength of right-wing populists correlate at a relatively low level (see Table 2). Moreover, we will see below that the policy indicators are rather weakly related to policy outcomes measured with Janoski’s naturalization rates. And even if all these indicators were highly interrelated we would not necessarily expect such a result. As Casper and Tufis (2003) have shown, even highly correlated indicators can lead to different results. For three highly interrelated measures of democracy they come to the conclusion that most of the explanatory variables do not display the same effects.

SECOND PART: POLITICS AND POLICIES

The policy process

The discussion on policy output and outcome has already made clear that it is important to differentiate between different steps of a policy process. Most commonly the following phases are differentiated: The initiation and discussion of a problem, the formulation of a solution (decision-making and policy output), the implementation of a policy, the outcome of a policy and finally its evaluation that might lead back to a discussion of the problem (see Anderson 1975; Jones 1984; Kevenhörster 1997: ch.7). To what extent one phase follows the previous one or occurs simultaneously and how
a previous phase influences the next one is still heavily disputed in the literature. Moreover, it is certainly not always possible to clearly distinguish the different phases. Nonetheless, such a differentiation serves at least as a heuristical tool to distinguish different crucial aspects of a policy process.

Following Czaika and de Haas (2011: 20), who discuss the effectiveness of immigration policies, I make a difference between the four following phases (see Graph 1): (1) Public policy discourses (broadly stated objectives), (2) Citizenship policies (stated policy objectives and specific laws, regulations and measures), (3) Implementation (discretion and constraints) and (4) Naturalization and rejection rates (outcome).

Czaika and de Haas (2011) further differentiate three potential gaps between the four phases: (1) There might be a discursive gap between the public discourses on the one side and the effectively taken decisions (policies) on the other hand. (2) It might further be that there is a gap between the implementation of a policy and the policy objectives (implementation gap). (3) Finally, the objectives and implementation of a policy might not lead to the intended outcomes (efficacy gap). In the following part I will compare the different policy phases by discussing the discursive, implementation and efficacy gaps.

-- Graph 1 about here --

The discursive gap

To gain a better understanding of the discursive gap we need to compare our policy (output) indicators with public discourses. By “public discourse” I understand “public communication about topics and actors related to either some particular policy do-
main or to the broader interests and values that are engaged.” Ferree et al. (2002: 9). I use the concept of “public discourse” interchangeably with the concept of “public debate” that stands for “all communication related to a particular issue” (Helbling et al. 2011).

The question here is whether public discourse on citizenship issues lead to the intended policies. Basically this is about the relationship between politics and policies and the question of whether struggles around a policy leads to a specific policy content. The expectation is that in the course of a public debate either a compromise is made or a dominant position emerges that is then translated into a specific policy. To a certain extent this has already been shown by those studies I have discussed above that came among others to the conclusion that the dominance of certain political parties lead to certain policies (Helbling 2008; Howard 2009; Janoski 2010; Koopmans et al. 2011). It thus seems that the policy programs of the parties in power or of those playing an increasingly important role in public debates (such as the right-wing populists) are indeed reflected in citizenship policies.

Since in this paper I am rather concerned with the validity of indicators and the question of how they relate to each other, the question of causality (and thus whether public debates reflect policies or have an impact on them) is not so important here. I rather like to know whether there is any relationship at all. To do this we need an indicator that measures public discourse. A common way to investigate public discourse is to analyze the media. Naturally, critics of this approach are quick to point out the selection bias of the media—according to them what is reported in the media rather reflects the strategies of the media than what is really going on in the public. Others argue that only what is reported in the media is relevant in a public debate—only those aspects that are perceived by other actors, and aspects that are not reported are
simply judged irrelevant. Moreover, there are methods that take news as a source of reported positions and arguments by political actors—thereby filtering out journalists’ own arguments (see Helbling and Tresch 2011).

An alternative way to measure public discourse is to account for party manifestos, which reflect the parties’ political strategies; especially if we assume that political parties play a major role in such political struggles. For some people, however, the relevance of party manifestos is disputable, because voters do not read them (Kriesi et al. 2008: 66-67). On the other hand, Robertson (1976: 72) underlines that manifestos provide the basis for statements given by politicians during campaigns. Moreover, validity tests have shown that positions presented in manifestos and put forward in electoral campaigns correlate to a high degree (Helbling and Tresch 2011).

I know of only one project that provides quantitative data on public debates on immigration and other issues that allows measuring political actors’ positions. The comparative research project 'National Political Change in a Globalizing World' (Kriesi et al. 2011) has studied public debates on immigration and other issues between 1998 and 2006, on the one hand, and party competition in the context of national election campaigns in the 1970s, 1990s and 2000s, on the other hand. Data in six Western European countries have been collected (Austria, France, Germany, the Netherlands, Switzerland and the United Kingdom). The research team content analyzed newspaper articles related to the electoral contest or to politics in general and coded actor’s positions towards political issues. For the electoral campaigns all political parties were included and for the public debates all actors that took position toward political issues.

To measure political parties’ positions reflected in their manifestos we rely on the dataset of the Comparative Manifestos Project (CMP) (Budge et al. 2001; Klinge-
mann et al. 2006). The CMP dataset includes information that has been retrieved from party manifestos and published in the context of national election campaigns. Information on 56 issues has been collected for the period 1945 – 2003 for mostly Western countries and Eastern European countries since 1990. There is no category for citizenship issues but two related categories that capture positions on nationalism, patriotism and multiculturalism.

I have built aggregated indicators that reflect the average positions of all political parties and political actors per country for the time periods that are closest to the points in time for which the data for the policy indicators have been collected. Thus, the policy indicators are compared with the overall political climate in electoral campaigns, public debates and party manifestos in the same period. The major problem is that most datasets I use only include data for a limited number of countries and time periods (with the exception of the manifesto data). As we see especially for the data on public debates in the last column in Table 4 we compare only a small number of observations, which makes it quite difficult to observe strong interrelationships.

Another problem concerns the issue categorizations in both datasets. The NPW-dataset includes data on immigration issues in the large sense. Citizenship issues are only one aspect in this category and the number of observations is too small to build an indicator on citizenship issues only. In the CMP-dataset the respective categories include very general issues such as nationalism, patriotism and multiculturalism. We therefore certainly have a content validity problem. However, it is still a good idea to compare the different indicators because, first, we simply have no other data at hand. Second, it is not inconceivable that political actors take similar positions on citizenship and related issues and that thus debates on more general immigration issues are at least a proxy for more specific citizenship debates.
In Table 4 I have correlated the indicators from electoral campaigns and public debates from the project 'National Political Change in a Globalizing World' with the different policy indicators. In Table 5 the same indicators have been correlated with the manifesto data. In the second column in Table 5 the positions of political parties have been weighted by their electoral strength. As we immediately see, there is virtually no relationship between policy indicators on the one hand and the public debates indicators on the other hand. The only exception is the CPI data that seem to be tightly related to the overall positions in electoral campaigns.

In the light of these results we should not jump to any final conclusions given the low number of cases and the poor quality of the public debate data. It might very well be that we face here a type 2 error or false negative and that we would indeed observe a relationship if we disposed of more and better data. We can however clearly say that there is no direct and strong relationship between the two phases of a policy process. As a side note I would like to refer to Vligenthart’s (2007) work that came to similar conclusions with much better data and sophisticated time series models. Comparing media and parliamentary debates on immigration and integration issues in the Netherlands he investigated partly different arenas. It is however nonetheless revealing that the debates in both arenas hardly influence each other.

Public debates do not directly lead to specific policies, but are probably filtered at various stages. In other words, the null finding in Tables 4 and 5 show us that it is definitely too simplistic too assume that there is some kind of a citizenship or nationhood model if one understands by this a general idea of how a nation-states should be composed that is dominant in all phases of a policy process. If there were for example something like a German citizenship model that is independent of specific institutions and actors and deeply historically embedded then we would observe the same posi-
tions and discourses in various circumstances. The fact that this is not the case shows us that it makes indeed sense to take a closer look at policy processes as changes occur between the different phases of such a process.

-- Tables 4 and 5 about here --

The implementation gap

While there seems to be a discursive gap we know from previous research that there is also an implementation gap. There is however no such thing as an implementation indicator that would allow us to compare the two phases in the same way I did above with public debates and policies. So we have to make do it with case studies that have studied implementation process in the field of citizenship politics.

The implementation gap becomes most apparent in federal states when we observe that a national policy is implemented differently in different regions of a country. A case in point is Switzerland. Manatschal’s (2011) results show that there is a large variety of policies across Switzerland. Going even a level further down I have shown in my own work that there is a large variety of citizenship policies at the local level (Helbling 2008).

Diverging applications of citizenship laws and understandings of nationhood are not found only in Switzerland.6 With regard to the implementation of naturalisation laws, the one case that probably comes closest to the Swiss system is Germany, where the citizenship law is regulated at the national level but executed by the Länder (sub-national states) (see Dornis 1999; Bultmann 1999; Hagedorn 2001a,b; Ludwig 2004,

---

6 The following paragraphs are adapted from Helbling (2008: 18-20).
2001; Thränhardt 2008: 15-26). The Länder are accorded the competence to organise the naturalisation procedure and to decide who fulfills the criteria for being naturalised and who does not. Various studies have shown that naturalisation rates and citizenship politics differ significantly between the Länder (Hagedorn 2001a,b, Dornis 1999, 2001).

In other countries, divergent practices of citizenship can be observed, too. Waldrauch and Cinar (2003) and Cinar and Waldrauch (2006) present diverging implementation policies and integration requirements at the regional level in Austria. North (1985, 1987) examined the administrative structure of the American naturalisation program, and came to the conclusion that the formal procedures and the approaches of the examiners vary a great deal among district offices. There are differences with regard to how language tests are evaluated and how and when people are persuaded to withdraw their dossiers when they seem not to be qualified for naturalisation (North 1985: 38-39).

Differing implementations of the naturalisation laws can even be observed in highly centralised states such as France. Weil (2004: 377-387) observed applications that clearly contradict the idea of a voluntaristic citizenship model with which France is often associated. In some regional offices, candidates are even manifestly dissuaded from applying for French citizenship. While Costa-Lascoux (1996: 149) reports that the assimilation of candidates is tested very differently from one regional office to the other, Hagedorn (2001b: 43-44) found that candidates are often refused for lack of assimilation. Since the degree of assimilation is exclusively judged by regional officers, arbitrary decisions cannot be excluded (see also Fulchiron 1996).

Variations at the sub-national level can be observed not only in the context of naturalisations but also, more generally, when we analyse local integration policies and
how municipalities manage ethnic diversity. It is often argued that cities have always been more confronted by problems related to migration than nation-states are, and thus constitute more interesting cases to analyze (Rogers and Tillie 2001; Rogers 2000; Penninx et al. 2004b; Favell 2001). In those studies, it is argued that the local political structures shape the ways migrants are integrated. Although an influence of the national citizenship regime on local politics can sometimes be detected (e.g. Garbaye 2004), various empirical studies have demonstrated that forms of local integration and citizenship policies cannot be explained (at least not exclusively) by their embedment in a national system.

The essays collected in Penninx et al. (2004a) demonstrate that in various countries local policies towards immigrants vary a lot and that this heterogeneity is mainly attributable to differences between local political systems and prevailing citizenship paradigms. Garbaye (2000) has explored how two cities in France and Great Britain manage ethnic diversity and identifies three elements that explain the different approaches: the relations between central and local governments, the organisation of political party systems and the organisation of local government. It is particularly interesting to observe that even in a country such as France, cities make differential use of their existing discretionary power to integrate foreigners (Moore 2001, 2004).

**The efficacy gap**

Finally, I also like to know to what extent certain policy outputs lead to the intended policy outcomes. Does a restrictive naturalization policy lead to less naturalizations? To find an answer to this question in Table 6 I compare the sub-indicators of naturalization regulations (see Table 2) with the naturalization rates indicator provided by
Janoski (2010). We see that there is indeed a link from outcomes to outputs, but that this interrelationship is relatively weak. It thus seems that other factors besides policies also explain the number of naturalized immigrants.

-- Table 6 about here --

The results of other studies that have looked at the interrelations between policies and their outcomes are rather mixed. Janoski (2010) found that his “barrier to nationality index” that focuses on naturalization criteria and the naturalization process, has a strong impact on naturalization rates. In my own study I have also revealed that the dominant local understanding of citizenship that is measured by means of politicians’ attitudes towards naturalization criteria explains rejection rates (Helbling 2008: ch. 5). Comparing the MIPEX naturalization sub-indicator with averaged naturalization rates for the period 2002 to 2007 Reichel (2011) however found no interrelationship at all. In any case there seems to be an efficacy gap and that other factors than formal regulations have also an effect on the number of people who get naturalized.

Since many of the citizenship indicators discussed in this paper also include elements of integration policies we should not only look at the effects of these policies on naturalization rates but also on immigrant integration. There is already a series of studies that have investigated the effects of integration policies. Findings in this field have again been mixed or contradictory, however. On the one hand, some have argued that liberal or multi-cultural integration regimes are least successful on integrating immigrants. In other words, states with more assimilationist policies or those with less generous welfare states (and particularly a combination of the two) are more successful in
integrating migrants (Koopmans 2010; Joppke 2007; de Wit and Koopmans 2005). The idea here is that the lack of a social safety net forces immigrants to leave ethnic enclaves and “acquire the linguistic and cultural skills that are necessary to earn a living” (Koopmans 2010: 21).

Indeed, in recent years states that had formerly exemplified multiculturalist integration approaches, such as the Netherlands and Great Britain, have been found to be turning increasingly to assimilationist policies as a result of the perception that multiculturalism has failed to properly integrate migrants (Brubaker 2001; Joppke 2004; Givens 2007; Ireland 2004). On the other hand, Bloemraad’s (2000) comparison of the U.S. and Canada argued that the latter country’s more favorable approach to multiculturalism benefits the civic and political incorporation of new immigrants.

Most studies, however, that have already investigated the influence of policies have not found any effects (Tubergen et al. 2004; Fleischmann and Dronkers 2010; Dinesen and Hooghe 2010). The problem with some of these studies, however, is that they focus exclusively on socio-economic integration (Tubergen et al. 2004; Fleischmann and Dronkers 2010), use rather crude indicators (Van Tubergen et al. 2004), or are based on a very low number of cases (Koopmans 2010; Ersanilli and Koopmans 2010). Van Tubergen et al. (2004: 711) simply differentiate whether destination countries have a point system or not. Koopmans’ conclusions are based on studies that include between three and eight countries only.

Investigating social and political integration gaps between natives and immigrants with a cross-national data set and eighteen countries Helbling et al. (2011) show that countries with the most comprehensive and immigrant-friendly integration policies also exhibit the lowest political engagement gaps. They have shown that integration
policies played a role for lowering immigrant-native disparities in political engagement.

CONCLUSION

One of the main questions that guided us through this paper was whether or not it makes a difference which indicators we use to study citizenship. And the answer is clearly yes. It appeared that while some of the indicators are strongly related to each other, others are not and thus seem to measure different concepts. The content validity tests showed that the various indicators cover a large range of different aspects. Some focus mainly on naturalization policies whereas others include mostly cultural right aspects and other integration policies. Still others combined different aspects. We also saw that there are policy output and policy outcome indicators and a third group that combined both aspects.

The convergent validity tests confirmed this first picture of rather heterogeneous indicators. While indicators measuring mostly either naturalization or integration policies are highly correlated among each other, it appeared that both groups of indicators are only weakly related to each other. It thus makes a difference whether we use a naturalization or integration indicator.

This general picture was put in question with the construct validity tests that revealed that indicators that cover different aspects and that correlate only at a low level with each other are still explained by the same factors. Among others it was shown that political parties play a crucial role to explain policy outcomes as well as naturalization and rejection rates.
The differentiation between policy outputs and outcomes laid the basis for a more thorough discussion of policy processes in the citizenship field. I differentiated public debates, policies, their implementation and policy outcomes. While I showed that there is a weak link between outputs and outcomes, it became clear that there are important gaps between public debates and policies, on the one hand, and policies and their implementation on the other hand.

These findings make it clear that we need to be extremely careful when we choose our indicators. Depending on our research questions and the policies or phases of a policy process we like to investigate if some indicators are more valid than others. It will also become clear that disaggregating the citizenship field helps us make useful comparisons and to realize that we do not encounter the same political climate in all areas and phases of a policy process. Public debates are not directly translated into policies, which in turn do not always lead to the intended outcomes. This puts in question the concept of citizenship models and the idea that dominant understandings of citizenship are inscribed in a society (Helbling 2010; Michalowski and Finotelli 2012).

The last aspect I like to discuss concerns the question of how precise the indicators ought to be. Despite all the differences, we have seen that indicators that include at least to a certain extent naturalization policies are highly related to each other despite the fact that they have been constructed differently and include different amounts of information.

In their discussions of democracy indices Elkins (2000) and Coppedge and Gerring (2011: 249) argue that contrary to simple dichotomous indicators continuous and more complex ones are more precise as they are more sensitive to gradations (see also Bader 2007: 876). Elkins (2000) shows that graded measures have superior validity and reliability. Related to the problem of gradation is the question of how many sub-
indices need to be aggregated to construct valid indicators. Everyone might agree that the more aspects are covered the more valid and precise an indicator is as it contains more information (see Michalowski and Van Oers 2011). Moreover sources of error are also reduced when a larger variety of data are accounted for (Coppedge and Gerring 2011: 251).

As we have seen, the indicators, that include information on naturalization policies, have been built in very different ways and include different amounts of information. Howard’s Citizenship Policy Indicator (CPI) is certainly the most straightforward indicator as it is based on three aspects of legal regulations and six variables. Waldrauch and Hofinger (1997) included almost 80 items in their index. Koopmans et al.’s (2005; 2011) 19 sub-indicators of which the individual equality dimension (Koop_IE) is built, do not only involve legal but also cultural aspects that depend on jurisprudence, administrative decrees and local implementation practices – information that is more difficult to find and code. Finally, the MIPEX-indicator is based on a large range of over 140 sub-indicators that have been collected by means of expert surveys.

In the light of the high correlations between these indicators (see Table 2) one might wonder why we do not simply stick to indicators that require relatively few resources to be built. More generally, we also need to ask why so many resources are invested in large projects? Would it not be better to heavily reduce the number of items that are covered in such a project and instead collect data over time and for countries outside the Western world? Such questions not only concern naturalisation, but also other domains, such as the research projects that measure policy positions of political parties. Various studies have shown that different methods based on newspaper and party manifesto codings as well as expert and population surveys lead to similar results although the respective methods require different amounts of resources (Marks et al.
REFERENCES


Duyvendak, Jan Willem, Rogier van Reekum, Peter Scholten, Christophe Bertossi (2011): The of/for distinction. EUDO Citizenship Forum (European University Institute, Robert Schuman Centre for Advanced Studies, EUDO Citizenship Observatory).


parative Political Studies 42(2): 198-223.


Janoski, Thomas (2011): “What we need citizenship indicators for depends on who are “we””. EUDO Citizenship Forum (European University Institute, Robert Schuman Centre for Advanced Studies, EUDO Citizenship Observatory).


Ludvig, Alice (2004): “Why should Austria be different from Germany? The two recent Nationality Reforms in Contrast”, German Politics 13(3): 499-515.


Reichel, David (2011): We need different indicators for different research questions. EUDO Citizenship Forum (European University Institute, Robert Schuman Centre for Advanced Studies, EUDO Citizenship Observatory).


Table 1: Overview content citizenship indicators

<table>
<thead>
<tr>
<th>Cultural</th>
<th>Economic</th>
<th>Social</th>
<th>Political</th>
<th>Policy outputs</th>
<th>Policy outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Immigration (entry)</td>
<td>Integration (settlement)</td>
<td>Natural. membership</td>
<td>Natural. rates</td>
<td>Reject. rates</td>
<td></td>
</tr>
<tr>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>(X)</td>
</tr>
<tr>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>

Notes: For the abbreviations see appendix Table A1.
Table 2: Convergence correlation tests

<table>
<thead>
<tr>
<th>Indicators</th>
<th>Correlations</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Naturalization Policy Indicators</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CPI - Koop_Nat</td>
<td>0.86 (p=0.000)</td>
<td>18</td>
</tr>
<tr>
<td>CPI - MIPEX_Nat</td>
<td>0.91 (p=0.000)</td>
<td>15</td>
</tr>
<tr>
<td>Koop_Nat - MIPEX_Nat</td>
<td>0.88 (p=0.000)</td>
<td>11</td>
</tr>
<tr>
<td><strong>Overall Policy Indicators</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CPI - Koop_IE</td>
<td>0.82 (p=0.000)</td>
<td>18</td>
</tr>
<tr>
<td>CPI - LOI</td>
<td>0.80 (p=0.031)</td>
<td>7</td>
</tr>
<tr>
<td>CPI - MIPEX</td>
<td>0.69 (p=0.004)</td>
<td>15</td>
</tr>
<tr>
<td>Koop_IE - LOI</td>
<td>0.81 (p=0.014)</td>
<td>8</td>
</tr>
<tr>
<td>Koop_IE - MIPEX</td>
<td>0.81 (p=0.003)</td>
<td>11</td>
</tr>
<tr>
<td>MIPEX - LOI</td>
<td>0.85 (p=0.008)</td>
<td>8</td>
</tr>
<tr>
<td>Koop_CD - MCP</td>
<td>0.72 (p=0.000)</td>
<td>33</td>
</tr>
<tr>
<td>CPI - Koop_CD</td>
<td>0.37 (p=0.130)</td>
<td>18</td>
</tr>
<tr>
<td>CPI - MCP</td>
<td>0.63 (p=0.000)</td>
<td>28</td>
</tr>
<tr>
<td>Koop_CD - Koop_IE</td>
<td>0.59 (p=0.000)</td>
<td>44</td>
</tr>
<tr>
<td>Koop_CD - LOI</td>
<td>0.56 (p=0.147)</td>
<td>8</td>
</tr>
<tr>
<td>Koop_CD - MIPEX</td>
<td>0.55 (p=0.080)</td>
<td>11</td>
</tr>
<tr>
<td>Koop_IE - MCP</td>
<td>0.76 (p=0.000)</td>
<td>33</td>
</tr>
<tr>
<td>MCP - LOI</td>
<td>0.56 (p=0.148)</td>
<td>8</td>
</tr>
<tr>
<td>MCP - MIPEX</td>
<td>0.66 (p=0.004)</td>
<td>17</td>
</tr>
<tr>
<td>CIVIX - CPI</td>
<td>0.29 (p=0.292)</td>
<td>15</td>
</tr>
<tr>
<td>CIVIX - Koop_CD</td>
<td>0.01 (p=0.969)</td>
<td>20</td>
</tr>
<tr>
<td>CIVIX - Koop_IE</td>
<td>0.23 (p=0.322)</td>
<td>20</td>
</tr>
<tr>
<td>CIVIX - LOI</td>
<td>0.22 (p=0.605)</td>
<td>8</td>
</tr>
<tr>
<td>CIVIX - MCP</td>
<td>0.13 (p=0.486)</td>
<td>30</td>
</tr>
<tr>
<td>CIVIX - MIPEX</td>
<td>0.37 (p=0.170)</td>
<td>16</td>
</tr>
</tbody>
</table>

Notes: High correlations above 0.7 are indicated in bold. For the abbreviations see appendix Table A1
Table 3: Explanatory variables

<table>
<thead>
<tr>
<th>Short-term factor</th>
<th>Long-term factors</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Political Parties</strong></td>
<td><strong>Colonialism</strong></td>
</tr>
<tr>
<td>CPI</td>
<td>X</td>
</tr>
<tr>
<td>Koopmans</td>
<td>X</td>
</tr>
<tr>
<td>Janoski</td>
<td>X</td>
</tr>
</tbody>
</table>

Note: For the abbreviations see appendix Table A1.

Table 4: Correlations between policy indicators and public debates

<table>
<thead>
<tr>
<th>Policy indicators</th>
<th>Electoral campaigns</th>
<th>Public debates</th>
<th>Years</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPI</td>
<td>0.82</td>
<td>0.56</td>
<td>08</td>
<td>5</td>
</tr>
<tr>
<td>Koop_CD</td>
<td>0.13</td>
<td>-0.36</td>
<td>02 / 08</td>
<td>18 / 12</td>
</tr>
<tr>
<td>Koop_IE</td>
<td>0.17</td>
<td>0.11</td>
<td>02 / 08</td>
<td>18 / 12</td>
</tr>
<tr>
<td>LOI</td>
<td>-0.29</td>
<td>0.20</td>
<td>08</td>
<td>6</td>
</tr>
<tr>
<td>MCP</td>
<td>0.17</td>
<td>0.01</td>
<td>00 / 10</td>
<td>12</td>
</tr>
<tr>
<td>MIPEX</td>
<td>0.47</td>
<td>0.02</td>
<td>07</td>
<td>6</td>
</tr>
</tbody>
</table>

Note: For the abbreviations see appendix Table A1.

Table 5: Correlations between policy indicators and party manifestos

<table>
<thead>
<tr>
<th>Policy indicators</th>
<th>Manifestos (unweighted)</th>
<th>Manifestos (weighted)</th>
<th>Years</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPI</td>
<td>0.14</td>
<td>-0.01</td>
<td>80</td>
<td>15</td>
</tr>
<tr>
<td>Koop_CD</td>
<td>-0.09</td>
<td>-0.06</td>
<td>80 / 90 / 02</td>
<td>33</td>
</tr>
<tr>
<td>Koop_IE</td>
<td>0.03</td>
<td>0.01</td>
<td>80 / 90 / 02</td>
<td>33</td>
</tr>
<tr>
<td>MCP</td>
<td>-0.02</td>
<td>-0.04</td>
<td>80 / 00</td>
<td>40</td>
</tr>
<tr>
<td>MIPEX</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Note: For the abbreviations see appendix Table A1.
Table 6: Output and outcome comparison

<table>
<thead>
<tr>
<th>Indicators</th>
<th>Correlations</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPI – Nat. Rates</td>
<td>0.41 (p=0.057)</td>
<td>22</td>
</tr>
<tr>
<td>Koop_Nat – Nat. Rates</td>
<td>0.65 (p=0.001)</td>
<td>20</td>
</tr>
<tr>
<td>MIPEX_Nat – Nat. Rates</td>
<td>0.52 (p=0.067)</td>
<td>13</td>
</tr>
</tbody>
</table>

Graph 1: The citizenship policy process

(adapted from Czaika and de Haas (2011: 20))
## APPENDIX

Table A1: Citizenship indicators and sub-indicators

<table>
<thead>
<tr>
<th>Indicators</th>
<th>Names</th>
<th>Years</th>
<th>Main Publications</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>- Koop_CD</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Koop_Nat</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LOI</td>
<td>Index of Legal Obstacles to Integration</td>
<td>1995</td>
<td>Waldrauch/Hofinger (1997)</td>
</tr>
<tr>
<td></td>
<td>2010</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>- MIPEX_Nat</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>- naturalization</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>