Do policy instruments matter? Governments' choice of policy mix and higher education performance in Western Europe

Do Policy Instruments Matter?

Governments' choice of policy mix and higher education performance in Western Europe

ABSTRACT: Governments pursue their goals by adopting various mixes of policy instruments. This paper proposes a specific operationalization of these mixes and applies it to the analysis of reforms that many Western European governments have pursued, as they have adopted a similar policy design in their higher education systems (HESs) over the last twenty years. In fact, although these policies have similar templates, performance indicators exhibit remarkable variation between countries. Thus, by applying Qualitative Comparative Analysis to a large dataset containing all changes in policy instruments undertaken in the last twenty years in twelve HESs in Western Europe, this paper explores the possibility that differences in performance across national HESs could be associated – ceteris paribus – with different policy mixes. This paper finds not only that the common template has been applied through very different national policy mixes but also that only a few instruments are regularly linked to good teaching performance, regardless of the other components of the actual policy mix.

KEYWORDS: Policy Instruments – Higher Education – Governance Reforms – Western Europe – Performance – QCA

1. Introduction

Public policy studies offer different typologies of governance modes, each of which attempts to theoretically order the ways in which public policies are coordinated and steered (for all typologies, see: Considine and Lewis 1999; Treib, Bahr and Falkner 2007; Howlett 2011; Capano 2011; Tollefson, Zito, and Gale 2012). All these classifications share not only their adoption of the three fundamental, albeit differently arranged, principles of social coordination (hierarchy, market and network) but also the tendency to associate specific sets of policy instruments with each of the designed governance modes in a long-lasting or relatively long-lasting policy style. This tendency refers to the basic theoretical problem in which people select policy instruments that are congruent

with a governance mode and do so in a more or less predictable way, leading to a characteristic style or manner of formulating and implementing policies.

Any selection of policy instruments is characterized by *an intrinsic policy-mix trend* (Howlett 2005; Bressers and O'Toole 2005) and, as such, should be considered the result of a miscellany of different ideas, interests and technologies and deemed institutionalized in certain specific, recurrent contingencies.

Some mixes are thought to perform better than others, but it remains unclear why this is the case. For example, all studies of 'good governance' promoted by the Organization for Economic Co-operation and Development (OECD) and other international organizations tend to focus on the best mix of policy instruments adopted in pursuit of certain specific policy goals (OECD 2007, 2010), but they provide few indications on why the chosen tools constitute such an optimal arrangement.

Thus, there is an increasing problem in the governance literature regarding the real policy effectiveness of the process of continuous shifting that has characterized, in a comparative perspective and over the last three decades, governance modes in public policy and public administration. We know that governance modes have changed, and we know that these changes have taken shape via different policy mixes. However, we know very little about the characteristics and the actual consequences of these same mixes.

In this paper, we propose to unravel this double-sided problem by empirically focusing on governance reforms in Western European higher education systems (HESs). According to the mainstream literature related to governance changes in higher education (HE) (Huisman 2009; Paradeise *et al*. 2009; Shattock 2014; Capano and Regini 2014; Dobbins and Knill 2014), Western European governments have redesigned governance systems to make HE institutions more accountable by intervening with the introduction of rules that govern the allocation of public funding and tuition fees, the recruitment of academics, and the evaluation and accreditation of institutions. To accomplish this goal, these countries have turned to a similar policy formula (the so-called 'steering at a distance' governance arrangement).

However, according to contrasting results in the literature, there is no clear evidence regarding either the composition of the adopted policy formulas or whether and how the new governance template can be associated with satisfactory results. We address this gap by assuming a policy instrument perspective, meaning that the actual national

interpretation that each country has given to the common policy template in reforming HE governance can be assessed by focusing on the specific combinations of policy instruments that have been adopted at the national level. To understand these combinations and, thus, the specific content of the adopted policy design, we propose a specific operationalization of policy instruments whose constitutive logic could also be applied in other policy fields.

Furthermore, we explore the hypothesis that the way specific policy tools are set together matters for policy performance. We are fully aware that the link between policies' content (policy instruments) and their outcomes is indirect and limited (Koontz and Thomas 2012) and that policy performance is driven by many factors (in the case of HE policy, factors such as the percentage of public spending and the socio-economic, cultural background of families, external and internal shocks, and financial retrenchments matter). However, the main way governments can steer their policy systems is to adopt specific sets of policy tools to address the behavior of specific targets and beneficiaries; thus, the policy mixes that governments recurrently design could help to readdress the way that policies work and their performance. Therefore, policy mixes can be intended as possible explanatory conditions (among others), and the composition of the actual set of adopted policy instruments could make the difference or, at least, signal something of relevance from an explanatory point of view.

Being conscious of the intrinsic limitations of the research design, our main goal is to demonstrate the degree to which the research design we have followed can be promising in filling the existing knowledge gap concerning the assessment of (1) how policy instruments are really mixed together and (2) their association with the effects of governance reforms on public policy.

We have pursued this research strategy by collecting and analyzing data on the regulatory changes in HE in 12 countries over the last 20 years (1995-2014).

The paper is structured as follows. In the second section, we present our policy instrument framework, while in section three, we introduce the research design. The fourth section presents the results of the empirical analysis, which are then discussed in section five. Finally, the conclusion summarizes our preliminary results and proposes directions for future research.

2. Governance arrangements and systemic performance in HE: an instrumental perspective

2.1 Governance reforms in HE

Many scholars have underlined the apparent convergence towards the 'steering at a distance' mode in HE in recent decades (Braun and Merrien 1999; Paradeise *et al*. 2009; Huisman 2009; Shattock 2014). This governance arrangement is characterized by mixing the following instruments together: financial incentives to pursue specific outputs and outcomes in teaching and research, student loans, accreditation, ex post evaluation conducted by public agencies, contract benchmarking, and provisions by the law for greater institutional autonomy (Gornitzka *et al*. 2005; Lazzaretti and Tavoletti 2006; Maassen and Olsen 2007; Trakman 2008; Capano 2011; Capano and Turri 2017). However, this convergence certainly works to support general principles (more institutional autonomy, more evaluation, and more competition), while the concrete ways through which the policies are generated seem to be quite diverse (Capano and Pritoni 2018). Nonetheless, this diversity, at least in terms of the concrete composition of the adopted policy instruments, has never been clarified.

Furthermore, it should be noted that in terms of policy performance, the effects of these governance shifts have not been clearly assessed. In fact, the literature on HESs' performance has focused mainly on a few aspects as key determinants of policy success (or failure): the mechanism of funding (Liefner 2003), the total amount of public funding (Winter-Ebmer and Wirz 2002; Horeau, Ritze, and Marconi 2013; Williams, de Rassenfosse, Jensen and Marginson 2013), full institutional autonomy (Aghion *et al*. 2008), partisan dynamics (Ansell 2008), stratification (Willemse and De Beer 2012), and the type of loan system adopted (Flannery and O'Donoghue 2011). However, this research strategy simply assesses whether certain variables have the power to influence the probability of the outcome changing as expected on average at the population level, regardless of interactions. These types of explanatory results seem weak and generally risk remaining fairly superficial. The salient point here is that focusing on a single dimension to assess the performance capacity of a governance arrangement is quite misleading. For example, the effects of shifting from a centralized governance system to one in which universities are more autonomous cannot be analyzed without

contextualizing the change within its specific configuration, that is, by considering the other relevant instrumental dimensions (for example, how universities are funded, the system of degree accreditation, whether a national research evaluation assessment is present, etc.).

Overall, despite the significant governance shifts in HE, there is currently a lack of knowledge on both the actual nature of these changes in terms of existing combinations of adopted instruments and the policy performance of the new governance arrangements.

2.2. Governance arrangements as a set of policy instruments

Policies are steered by specific governance arrangements composed of rules, instruments, actors, interactions, and values (Howlett 2009; 2011; Tollefson, Zito, and Gale 2012; Capano, Howlett, and Ramesh 2015). The implicit assumption of the governance literature is that different governance modes or arrangements deliver different results in terms of policy outcomes. However, empirical evidence on this issue has been lacking, especially because the main analytical focus in public policy has been the process of changing governance arrangements in terms of their content with respect to the actors involved, the distribution of power, and the adoption of 'new' policy instruments. Thus, there has not been enough focus on the results of these governance shifts in mainstream public policy.

However, there is an increasing awareness that pure types of governance arrangements do not actually work; instead, the main principles of coordination (hierarchy, market and network) are combined in various ways. All governance arrangements are essentially hybrids and are characterized as working through policy mixes, that is, policy instruments belonging to 'different' instrument categories or pertaining to different policy paradigms/beliefs/systems/ideologies (Howlett 2005; Ring and Schroter-Schlaack 2010; Capano, Rayner, Zito 2012). Thus, an existing set of adopted policy instruments can be conceptualized as specific portfolios, settings, and combinations of different types of policy instruments associated with different working logics (Jordan *et al*. 2012; Schaffrin *et al*. 2014; Howlett and del Rio 2015).

However, how can the content of these policy mixes be described and understood, and how can their policy performance be assessed? In an attempt to fill these theoretical and empirical gaps, we adopt a bottom-up perspective by focusing on the basic unit of

any governance mode – the policy instruments that can be adopted – and their possible combinations. This approach seems quite realistic; policy instruments are the operational, performance-related dimensions of governance arrangements.

Accordingly, we operationalize systemic governance arrangements in terms of adopted policy instruments and therefore as specific sets of techniques or means by which governments try to affect the behavior of policy actors to direct them towards the desired results (Linder and Peters 1990; Vedung 1998; Howlett 2000; Salamon 2002).

There are numerous classifications by which policy tools can be ordered based on different criteria of analytical distinction, from coercion to the type of governmental source adopted (Ingram and Schneider 1990; Phidd and Doern 1983; Hood 1983; Vedung 1998; Howlett 2011). All these typologies suggest different families of instruments. Our research framework focuses on the capacity of policy instruments to induce specific behaviors; thus, we need to consider the nature of the instruments and examine the different ways through which they induce action towards the expected result. In conducting this examination, we are inspired by the classical theorization of Vedung (1998). When focusing on the nature of substantive policy instruments, Vedung grouped these instruments by the basic inducement on which they relied to foster compliance.

By adopting this perspective, we can delimit four distinct families of substantial policy instruments that have different, non-overlapping capacities to induce behaviors: *regulation*, *expenditure*, *taxation* and *information*[1]. Each family is associated with a specific inducement. Regulation induces behavior control; expenditure induces remuneration; taxation – depending on the way it is designed – can engender behavior control and remuneration; and information offers persuasion. Notably, all four families of tools can be employed by applying different methods of coercion that are dependent on how free individuals are to choose alternatives. Regulation can be quite strong or weak according to the type of behavioral prescriptions provided. Expenditure can lack coercion

[1] In our perspective, taxation can be considered an autonomous substantial instrument. We are aware that in other typologies, taxation is the chief component of broader instruments. Phidd and Doern (1983) considered taxation to be a means of regulation (as it implies high coercion), while Hood (1983), following the same reasoning, held partially to the 'authority' type (i.e., user charges) and partially to the treasury type (i.e., tax exemptions, tax expenditures). We believe that expenditure and taxation have different political and economic characteristics and present different ways of inducing or restraining institutional and individual behavior (Woodside 1983).

in the case of subsidies but be very demanding when delivering targeted funding. Taxation can be quite coercive when a general tax increase is established, but it can also have a low degree of coercion when many targeted tax exemptions exist. Information can be coercive when compulsory disclosure is imposed or lack strong coercion when monitoring is applied.

The four families of substantial policy instruments we have decided to consider, as well as the types proposed by other policy instrument classifications, represent very general instrumental principles (which need to take specific forms to be practically applied): they are 'families of individual instruments sharing similar characteristics' (Bouwma *et al*. 2016, p. 216)

Thus, according to Salamon (2002), the 'shape' through which the substantial instrument is designed to deliver the expected result is the important factor in terms of policy impact and potential performance. For every type of family of substantial policy instruments, there are different shapes of delivery that offer actual outlets through which those substantial instruments can affect reality. In addition, these instrumental shapes should be considered the basic analytical unit when assessing how policies are made and, thus, which types of governance arrangements actually work in terms of policy performance.

Accordingly, the important factors in detecting the adoption of a regulation instrument are the various forms through which regulations can be designed: for example, by imposing a specific behavior, enlarging the range of opportunities, or establishing specific public organizations. Expenditures can be delivered through grants, subsidies, loans, lump sum transfers, and targeted transfers, among other shapes. Taxation can imply fees, user charges, and exemptions, among others. Information can take the shape of neutral administrative disclosures, monitoring, diffusion, etc.

Each of these instruments' shapes carries specific potential effects that cannot be measured alone because they should be considered in relation to the other tools that compose the actual set of adopted policy instruments. Understanding the distinct shapes that various types of substantive policy instruments can take when delivered is essential

to grasping how governments change the instrumental side of governance arrangements over time[2]. There are three dimensions in which this distinction is helpful.

The first dimension is descriptive. By focusing on the different shapes of policy instruments, the usual description can acquire a more detailed reconstruction of shifts in governance compared to the usual distinction of more or less market, more or less hierarchy, more or less regulation, more or fewer expenditures, etc.

The second dimension is analytical. Due to the focus on the basic shapes through which substantial instruments are delivered, the policy mix concept can be operationalized in a very effective and realistic way. Thus, the eventual relevant differences can be assessed in terms of policy settings because the content of policy mixes can be grasped in a very detailed manner.

The third dimension is clearly exploratory, with a particular emphasis on the possible explanatory relevance of policy instruments for performance. If policy performance is assumed to be conditioned by – among other factors – the adopted set of policy instruments, a detailed operationalization of the substantial instruments should uncover which combinations of instruments actually work, which can lead to a better reconceptualization of the (many) determinants of policy performance and of their interactions.

This perspective seriously considers the suggestions of those scholars who have observed how the actual set of adopted policy instruments is the consequence of a diachronic accumulation. Thus, we need to analyze the full package instead of a single type of policy instrument (Hacker 2004; Pierson 2004; Huitema and Meijerink 2009; Tosun 2013; Schaffrin *et al*. 2014).

Therefore, from a theoretical perspective, the decision to focus on the instrumental side of governance arrangements to measure their potential impact on policy performance encourages proceeding by using combinatory logic: when the expected effect, namely, policy performance, is assumed to depend on – among other things – a combination of

[2] Here, we follow the suggestion of Ingram and Schneider (1990: 522, n. 5) both conceptually and in the operationalization presented below. They stated, 'most tools can be disaggregated into relatively small units and each unit then scored in terms of all behavioral dimensions of interest to the investigator. Even the smallest units, such as a single statement, may score "high" on more than one behavioral dimension. The units and their scores can then be re-assembled to produce a multivariate characterization of the original policy tool'.

multiple conditions (i.e., specific settings of policy instruments), then a complex causality principle is at work. For this reason, we decided to turn to a Qualitative Comparative Analysis (QCA)[3]. QCA is a configurational and set-theoretic approach in which relationships of necessity and sufficiency are tested and the idea of causality underpinning the approach is fundamentally characterized by equi-finality[4], conjunctural causation[5], and asymmetry[6]. Therefore, QCA aims at unravelling multi-causal rather than mono-causal explanations, focuses on combinations of conditions rather than on single variables, and does not assume that a unique solution (or equation) accounts for both the occurrence and non-occurrence of a particular outcome.

Furthermore, by applying this method, we can precisely assess which instruments or combinations of instruments are present when better performance is achieved. By using this method of operationalizing and methodologically treating policy instruments, a relevant analytical gap can be filled, producing a more precise design for research on the determinants of policy performance; these determinants are too often based on the direct effects of structural and contextual variables (comprising factors such as the socio-economic situation, social capital, and the political situation) or processual variables. Overall, policy makers change policy by choosing specific policy instruments, and thus, thanks to the adopted method, it is possible to better describe the content of these decisions. Showing which combinations of instruments are related to specific policy performances plays an enlightening role in readdressing the research on policy evaluation and improving the analysis of the links between policies and outcomes.

3. Research design

[3] QCA is a relatively new research approach (Ragin 1987; 2000; 2008; Rihoux and Ragin 2009; Schneider and Wagemann 2012). In recent years, QCA has drawn increasing attention within the social sciences, and some scholars consider QCA to already be a 'mainstream method' in political and sociological research (Rihoux *et al.* 2013).

[4] The idea of causation in QCA is equi-final in the sense that more than one causal pattern can lead to the outcome (Ragin 1987).

[5] The idea of causation in QCA is characterized by conjunctural causation in the sense that specific combinations of different conditions lead to the outcome (Rihoux and Ragin 2009).

[6] The idea of causation in QCA is asymmetric in the sense that there is no particular relationship between causal patterns leading to the presence of the outcome and the absence of the outcome. Conditions explaining the presence of the outcome are silent with respect to the absence of the outcome, and vice versa (Schneider and Wagemann 2012).

3.1 Case selection and timespan

This paper is based on a specific dataset of policy tools used in 12 Western European countries (Austria, Denmark, England, Finland, France, Greece, Ireland, Italy, the Netherlands, Norway, Portugal, and Sweden) between 1995 and 2014.

Regarding country selection, we initially intended to cover all the pre-2004 enlargement EU countries. We decided to exclude Eastern European countries due to the period of transition they experienced after 1989, a time of deep turmoil that mixed the communist legacy, the return to pre-communist governance and a kind of acceleration towards marketization (Dobbins and Knill 2009); these characteristics make it quite difficult to assess and code the characteristics of the adopted policy instruments.

However, we were forced to exclude, for different reasons, four additional countries: Luxembourg, due to its small size (one university); Belgium and Germany because of their federal structures, and Spain because of its very decentralized regionalism, which has a significant impact on the systemic governance of HE. These 11 pre-2004 enlargement countries reflect all historical types of university governance that have developed in Europe and can therefore offer sufficient differentiation in terms of policy legacy (Clark 1983; Braun and Merrien 1999; Shattock 2014) and the inherited set of policy instruments. We also included a non-EU country, Norway. Thus, all the Nordic countries, which are assumed to have adopted a welfarist approach to HE, could be considered, and it was possible to examine whether this common characteristic influenced the analyzed outcome.

In all the selected countries, the HESs have undergone structural changes in the last two decades. Accordingly, we decided to begin our analysis in approximately the mid-1990s to encompass all the major changes that involved HESs over the last 20 years[7]. Obviously, each country presents its own reform 'starting point' in the field, which means that some of the countries had already produced relevant legislation by the mid-1990s, while others began much later.

[7] Some of these reforms were implemented during the 1990s, while many others occurred or developed during the new millennium as a consequence of the Bologna Process.

3.2 Operationalization

As already explained in the theoretical section, we assumed that differences in HESs' teaching performance may be associated with – among other things – differences in the combinations of the adopted policy tools. Among the various possible indicators of HESs' performance – such as access, academic recruitment and careers, and third missions – we decided to focus only on teaching, which ultimately represents one of the main tasks of every HE institute. Furthermore, in the last thirty years, all governments have been committed to incentivizing universities to pay more attention to the socio-economic needs of their own country and to the need to increase the stock of human capital (currently, in many countries and at the EU level, increasing the number of citizens obtaining a tertiary degree is a major policy goal).

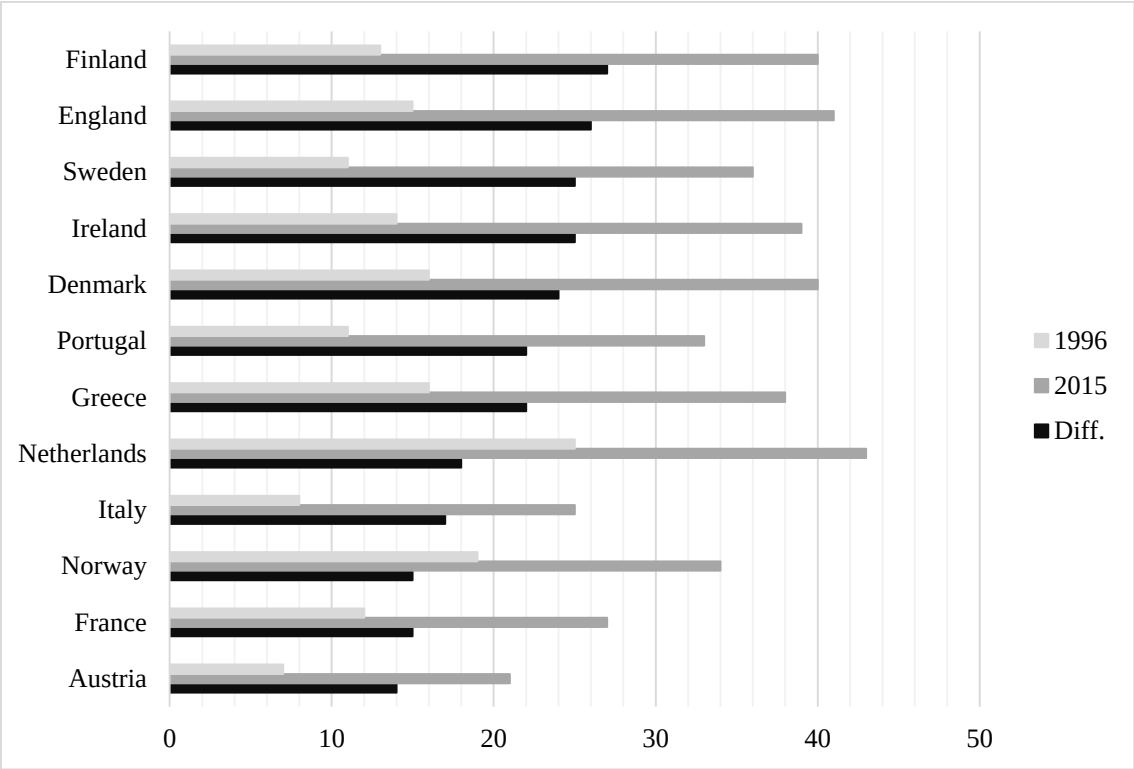
The most common indicator of teaching is the percentage (%) of people with a university-level degree[8]. As such, we operationalized teaching performance starting from the percentage of adults between 25 and 34 years old who have a university-level degree. Specifically, because many countries' education systems changed between the 1990s and the 2000s (possibly because of the Bologna Process), HE programs differ from those that existed 20 years ago. Thus, according to the OECD data, we chose our starting point as the percentage of the 25- to 34-year-old population with a 'university-level' education (level 5A in the ISCED 1997 classification) in 1996 compared to the percentage of people of the same age who had either a bachelor's or master's degree in 2015 (level 6 or 7 in the ISCED 2011 classification)[9]. These data were downloaded from the OECD archive (see also OECD 1998, 2016)[10] and are summarized in Figure 1 below:

[8] By 'university-level degree', we mean level 5A in the ISCED 1997 classification ('First stage of tertiary education: largely theoretically based programs intended to provide qualifications for gaining entry into more advanced research programs and professions with higher skills requirements') or level 6 (Bachelor's degree or equivalent: 'Programs designed to provide intermediate academic and/or professional knowledge, skills and competencies leading to a first tertiary degree or equivalent qualification') or 7 in the ISCED 2011 classification (Master's degree or equivalent: 'Programs designed to provide advanced academic and/or professional knowledge, skills and competencies leading to a second tertiary degree or equivalent qualification').

[9] First, we decided to examine the percentage of adults aged 30-34 years who had a university-level degree in each country, as this value was one of the targets of Europe 2020. However, those data account for both university and short-term tertiary degrees, while we coded only regulations related to universities.

[10] We opted for a simple indicator of teaching performance because teaching is one of the two main missions of HE, and increasing the number of citizens with an HE degree is a political goal. However, we

Figure 1 *University-level (ISCED 5A 1997 – ISCED 6 + ISCED 7 2011) attainment of 25- to 34-year-old adults: 1996 and 2015 in comparison*



However, the variation in this performance indicator does not directly represent our outcome because we could not account for two equally relevant considerations: first, improving results is easier when the starting point is a very low value than when it is a higher value, and second, results of university-level education are strongly linked to the structure of tertiary education as a whole. In other words, all else being equal, countries that offer short-cycle tertiary degrees (i.e., first-cycle degrees lasting less than 3 years: level 5B in the ISCED 1997 classification and level 5 in the ISCED 2011 classification) should be rewarded more than countries without them because in the former case, higher education institutions are subject to more competition for students (or, at least, they have

did not exclude a multidimensional index of performance, as suggested in the literature (Enders, de Boer and Weyer 2013), because we are perfectly aware that the concept of 'performance' in higher education is very complex and multifaceted (think, for example, about research performance and the so-called 'third mission'). Furthermore, we are also conscious that there is a qualitative dimension of performance, but it is currently virtually impossible to develop a reliable set of comparable data on the 'quality dimension'. The OECD has launched a research programme on this, but it is still in an initial stage, and the data are not yet very reliable.

a smaller catchment area) and, in turn, are less likely to improve their results. Consequently, we modified the data slightly following a two-step process. In the first step, we differentiated countries into three categories: countries below the mean of university degree attainment (25-34 years old) in 1996, countries above the mean but less than one standard deviation (s.d.) above the mean, and countries above one s.d. above the mean. Countries in the latter category had an increase in their performance (i.e., in the difference between their level of university degree attainment in 2015 and in 1996) that was equal to +50%, whereas countries in the intermediate category (above the mean but less than one s.d. above the mean) had an increase in performance that was between zero and +50%, depending on how much they were above the mean. Finally, countries below the mean did not show an increase.

In the second step, we adopted the same differentiation of countries with respect to the ratio between 25- to 34-year-old adults with a short-cycle tertiary degree and those with any tertiary degree in 2015. In this way, we could weigh the extent to which the opportunity to enroll in short-cycle degrees affected the HES as a whole. Again, countries more than one s.d. above the mean had a further increase in their performance that was equal to +50%, countries above the mean but less than one s.d. above the mean had an increase in performance that was between zero and +50%, depending on how far above the mean they were, and countries below the mean did not have an increase. Table 1 presents both the original OECD data and the country rankings based on our measure of teaching performance from 1996 to 2015.

Table 1 *Construction of the outcome (teaching performance between 1996 and 2015)*

Step 1

	Country	*University attainment (Level 5A) '96*	*University attainment (Level 6-7) '15*	*Difference 2015-1996*	*Increase 1*
	Austria	7	21	+14	+0% = +0.00
	Denmark	16	40	+24	+22% = +5.33
	England	15	41	+26	+12% = +3.00
	Finland	13	40	+27	+0% = +0.00
	France	12	27	+15	+0% = +0.00
	Greece	16	38	+22	+22% = +4.89
	Ireland	14	39	+25	+1% = +0.21

	Italy	8	25	+17	+0% = +0.00
	Netherlands	25	43	+18	+50% = +9.00
	Norway	19	34	+15	+50% = +7.50
	Portugal	11	33	+22	+0% = +0.00
	Sweden	11	36	+25	+0% = +0.00
	Mean	*13.92*		*+20.83*	
	St. deviation	*4.68*		*4.56*	

Step 2

	Country	*Short-cycle (Level 5) '15*	*Tertiary (Level 5-6-7) attainment '15*	*Short / Tertiary (%)*	*Increase 2*
	Austria	16	38	42.11	+50% = +7.00
	Denmark	4	43	9.30	+0% = +0.00
	England	8	48	16.67	+3% = +0.78
	Finland	0	41	0.00	+0% = +0.00
	France	17	44	38.64	+50% = +7.50
	Greece	1	40	2.50	+0% = +0.00
	Ireland	12	51	23.53	+27% = +6.75
	Italy	0	25	0.00	+0% = 0.00
	Netherlands	1	44	2.27	+0% = 0.00
	Norway	14	48	29.17	+43% = +6.45
	Portugal	0	33	0.00	+0% = +0.00
	Sweden	11	45	24.44	+27% = +6.75
	Mean			*15.72*	
	St. deviation			*14.94*	

Final outcome

	Country	*University attainment 2015-1996*	*Increase 1*	*Increase 2*	*Final outcome*
	Austria	+14	+0% = +0.00	+50% = +7.00	+21.00
	Denmark	+24	+22% = +5.33	+0% = +0.00	+29.33
	England	+26	+12% = +3.00	+3% = +0.78	+29.78
	Finland	+27	+0% = +0.00	+0% = +0.00	+27.00
	France	+15	+0% = +0.00	+50% = +7.50	+22.50
	Greece	+22	+22% = +4.89	+0% = +0.00	+26.89
	Ireland	+25	+1% = +0.21	+27% = +6.75	+31.96
	Italy	+17	+0% = +0.00	+0% = 0.00	+17.00
	Netherlands	+18	+50% = +9.00	+0% = 0.00	+27.00
	Norway	+15	+50% = +7.50	+43% = +6.45	+28.95
	Portugal	+22	+0% = +0.00	+0% = +0.00	+22.00
	Sweden	+25	+0% = +0.00	+27% = +6.75	+31.75

	Mean	*+20.83*	*+26.26*
	St. deviation	*4.56*	*4.46*

Source: our elaboration on OECD data.

The operationalization of the conditions of the next QCA, namely, the policy instruments, required greater theoretical reflection and greater effort in data gathering, as we will explain below. More precisely, we decided to operationalize the four families of substantial policy tools – regulation, expenditure, taxation and information – while considering a long list of shapes (24 in total), which are presented in Table 2.

Table 2 *Classification of policy instruments and their shapes*

Family of Policy instruments	*Condition*	*Shapes*	*Empirical examples*
Regulation	R1	Assessment, evaluation and accreditation (procedural rules)	Austria (1993): University of Applied Sciences Studies Act: an application for the accreditation of a degree program as a University of Applied Sciences degree programs had be addressed to the Agency for Quality Assurance
	R2	Agency for assessment, evaluation and accreditation	Denmark (2013): Establishment of Danish Agency for Higher Education as a merger between the Agency for Higher Education and Educational Support and the Agency for Universities and Internationalization
	R3	Content of curricula: more constraints	Norway (1995): Act on Universities and University Colleges: the degree structure, the individual higher education institution's study program portfolio, and the composition of professional studies is to be decided on and coordinated at the national level
	R4	Content of curricula: more opportunities	England (2004): Higher Education Act: the Director must protect academic freedom, including the freedom of institutions to determine the contents and manner of teaching of their courses
	R5	Academic career and recruitment: more constraints	Italy (2010): Law 240/2010: can apply calls for professorial recruitment launched by universities only if they have obtained a national qualification.
	R6	Academic career and recruitment: more opportunities	France (2013) Decree no. 305/2013: The treatment of teacher-researcher changes with his career seniority. Seniority bonuses can be awarded to lecturers and university professors that are voluntarily mobile
	R7	Regulation on students (admission	Greece (1997) Law no. 2525/1997: All upper-secondary graduates have access to HEIs after

		and taxation): more constraints	completing the Panhellenic university entrance examinations. The students compete for a fixed number of spaces irrespectively of demand.
	R8	Regulation on students (admission and taxation): more opportunities	Sweden (2012) Prop. 2012/13:1: The university or HE institution decides for themselves the requirement needed to be accepted to an education, course or program
	R9	Institutional and administrative governance: more constraints	Netherlands (2011) Stb. No. 95/2011: To regulate accurate decision making with respect to mergers (fusion), rules were revised. The representative bodies of the institutions (Councils consisting of students and staff) must agree with a merger in advance and must have had the opportunity to timely assess the effects of the merger ex ante
	R10	Institutional and administrative governance: more opportunities	Norway (2001) Whitepaper no. 27/2001: Institutions were given more freedom in academic, economic and organizational matters
	R11	Contracts	Austria (2002) Federal Act on the Organization of Universities and their Studies: The Federal Minister shall discuss every 2 years the total amount available for the universities - which will then be split for each university based on the performance agreements.
	R12	Rules on goals in teaching	Sweden (1993) The Higher Education Ordinance (no. 100/1993): the government and the parliament emphasized the need to follow up educational outcomes as well as examine and promote quality enhancement at universities and university colleges.
Expenditure	E1	Grants	Portugal (1997) Lei no. 113/1997: The law also stipulates the creation of social assistance for students of a lower socio-economic status, under the form of grants and other subsidies.
	E2	Subsidies and lump-sum funding	Norway (2001) Whitepaper no. 27/2001: The reform changes the funding of higher education institutions. The detailed and earmarked grants were removed, and higher education institutions' funding was changed to a lump sum grant at the institutions' disposal
	E3	Targeted funding	Netherlands (2010) Stb. No. 166/2010: the ministry introduces to so-called Sirius program, i.e. funding bachelors or masters programs of universities and UAS (in competition) focused towards excellence.
	E4	Loans	Denmark (1994) Amendment of Act relating to educational support: The act introduces a loan given to students who have used up the "punches" in their klippekort (a system similar to travel cards or season tickets that are punched for each use). These students

			are now given an additional loan in the last year of their education (SU-loven, 1994). The loan was intended to help these students finish their studies.
	E5	Performance based institutional funding	Austria (2014) UG 02 as of 2014 amendment: Structural funds are appointed to individual universities and are calculated in accordance with qualitative, quantitative and performance-based indicators
	E6	Standard cost per student	Ireland (2013) Higher Education System Performance Framework: The annual core grant is allocated as a block grant – based on a formula with a standard per capita amount
Taxation	T1	Tax exemption	Portugal (1992) Lei no. 20/1992: It stipulates that low-income students (based on a threshold which was to be defined each year by the Ministry of Finance) would be exempt from the tuition fee or receive a discount
	T2	Tax reduction for particular categories of students	Austria (2008) Decree no. 134/2008: Remission or Reimbursement of study fees extended for 50% disabled students or pregnant students or self-employed or working students
	T3	Service-based student fees	Portugal (2003) Lei n.o 37 de 22 de Agosto - Tuition fees are changed and are no longer a flat rate across all universities. Instead, they vary depending on the institution, the courses, and the relative quality
Information	I1	Transparency	England (1988) Education Reform Act: A higher education corporation shall have power to publish the results of the research or any other material arising out of or connected with it
	I2	Certifications	Sweden (2011) Prop. no. 133/2011: The operations that were managed by UHR are… assessment of foreign diplomas…
	I3	Monitoring and reporting	Ireland (2003) Official Languages Act: The Act specifies that each university must generate regular reports and other documents such as a strategic plan, quality assurance reports (s 35), an equality policy (s 36) and various financial reports (ss 37-42)

In this way, we tried to capture all the possible shapes that substantial HE policy instruments can take. We also avoided constructing categories that were too exclusive, which would have made the data collected in different countries difficult to compare.

Finally, it is important to note that there is a one-year lag between conditions and the outcome: conditions are operationalized based on data from 1995 to 2014, while the outcome compares (adjusted) teaching performance in 1996 and 2015. The reason is that

we believe that most changes in the policy instruments could have a 'quasi-immediate' impact on behaviors, and a one-year time lag carefully takes this process into account[11].

3.3 Data collection, dataset construction and coding

By following the lines of our theoretical framework – which focused on the different combinations of the existing set of adopted policy combinations – we collected, analyzed and coded all pieces of national legislation and regulations regarding HE in all 12 countries under analysis from the mid-1990s onwards. Hundreds of official documents and thousands of pages of national legislation were carefully scrutinized and hand-coded in the search for both substantial and procedural policy instruments. The coding procedure proceeded in three steps: first, we identified a list of relevant pieces of legislation in national HE policy, namely, laws, decrees, circulars and ministerial regulations that affected the HES of each country under scrutiny. Second, we reduced every piece of legislation to its main issues. Third, we attributed each of those issues to one of the shapes in which we classified the policy instrument repertoire in HE.

For the first two steps, the research strategy was twofold. With respect to Italy, France and both English-speaking countries – England and Ireland – the analysis was conducted 'in house', meaning that the three authors of this paper were responsible for entering the Italian, French, English and Irish pieces of legislation into the dataset. Linguistic barriers rendered the selection of regulations and their direct coding impossible for the other eight countries – Austria, Denmark, Finland, Greece, the Netherlands, Norway, Portugal and Sweden. Therefore, we contacted a highly reputable country expert for each case to achieve a perfectly comparable list of pieces of relevant regulation and, in turn, legislative provisions regarding HE for those countries.

The attribution of all the analyzed relevant decisions to the appropriate categories (substantial policy instruments and related shapes) was again conducted by the authors.

[11] We are aware that it can also be argued that policy changes in the area of HE *do not* have almost-immediate effects. The immediacy of effects clearly depends on the type of policy change, and some policy changes may require more time than others to produce effects (while a change in tuition fees or in the formula funding can have an immediate effect, a change in institutional governance can require several more years to show an impact). To account for this consideration, we conducted a robustness check with a 5-year time lag between policy instruments and the outcome (see the Online Supplementary Material). The results were not particularly different from those when we hypothesized an almost-immediate effect.

This final step of the coding procedure was developed as follows: first, each issue of each legislative provision in each country was coded separately by each author; second, contradictory cases – i.e., policy instruments placed in different categories by two or more coders (approximately 15% of the whole sample) – were solved jointly in a subsequent stage.

4. Linking policy instruments with teaching performance in Western Europe

4.1 Policy mixes in Western European HESs: an overview of the adopted policy shapes

While we refer to the Online Supplementary Material (Appendix A) for all technical details concerning our QCA, we first would like to provide a general picture of how the countries under scrutiny intervened in HE between 1995 and 2014. More precisely, Table 3 indicates how often each country in our sample recurred to all the shapes of policy instruments listed in Table 2.

Table 3 *HES governance reforms in Western Europe (1995-2014)*

	A	*D*	*E*	*FI*	*FR*	*G*	*IR*	*IT*	*NE*	*NO*	*P*	*S*	*Tot*
R1	3	4	1	4	9	3	9	14	4	16	8	5	*80*
R2	2	4	0	0	3	4	3	2	1	3	2	3	*27*
R3	1	2	1	6	4	5	0	4	4	3	0	5	*35*
R4	5	12	1	2	6	7	0	3	6	6	2	8	*58*
R5	0	0	1	1	2	3	2	6	0	1	3	1	*20*
R6	1	0	1	1	4	3	5	6	1	0	4	1	*27*
R7	3	2	1	0	1	1	1	0	7	3	0	3	*22*
R8	6	4	4	11	3	3	0	1	6	3	0	5	*46*
R9	1	3	4	8	6	7	8	6	3	2	1	2	*51*
R10	4	11	0	6	6	7	3	2	7	14	4	3	*67*
R11	4	2	1	0	2	1	1	1	0	1	0	0	*13*
R12	1	3	0	4	1	1	0	2	3	4	0	4	*23*
N Reg.	31	47	15	43	47	45	32	47	42	56	24	40	*469*
% Reg.	57.4	67.1	50.0	53.8	70.1	68.2	62.7	58.0	70.0	60.2	64.9	74.1	*63.1*
E1	2	3	1	4	2	3	0	2	5	5	4	3	*34*
E2	3	1	0	7	2	5	1	1	2	3	1	2	*28*
E3	1	7	0	8	4	2	2	1	2	2	0	2	*31*
E4	1	2	2	0	1	2	0	2	4	4	1	2	*21*

E5	1	1	4	8	2	3	4	5	0	10	2	0	*40*
E6	1	1	0	2	0	0	1	2	0	1	0	0	*8*
N Exp.	9	15	7	29	11	15	8	13	13	25	8	9	*162*
% Exp.	16.7	21.4	23.3	36.3	16.4	22.7	15.7	16.0	21.7	26.9	21.6	16.7	*21.8*
T1	2	1	0	1	1	0	1	1	0	0	0	0	*7*
T2	2	0	5	1	3	0	3	4	1	0	0	0	*19*
T3	0	0	0	2	0	0	0	1	0	2	3	0	*8*
N Tax.	4	1	5	4	4	0	4	6	1	2	3	0	*34*
% Tax.	7.4	1.4	16.7	5.0	6.0	0.0	7.8	7.4	1.7	2.2	8.1	0.0	*4.6*
I1	3	4	2	2	3	4	5	10	3	6	1	0	*43*
I2	1	0	1	1	1	0	0	0	0	1	0	3	*8*
I3	6	3	0	1	1	2	2	5	1	3	1	2	*27*
N Info.	10	7	3	4	5	6	7	15	4	10	2	5	*78*
% Info.	18.5	10.0	10.0	5.0	7.5	9.1	13.7	18.5	6.7	10.8	5.4	9.3	*10.5*
Tot	*54*	*70*	*30*	*80*	*67*	*66*	*51*	*81*	*60*	*93*	*37*	*54*	*743*

With regard to regulation – the most utilized family of policy instruments – Figure 2 reveals that two out of the three most adopted instrumental shapes are concerned with giving more opportunities to universities in terms of both the content of curricula and the institutional governance (which seems quite coherent with the common template pursued by all the examined countries: the 'steering at a distance' governance model). However, the most frequently utilized of all the policy instruments is related to assessment, evaluation and accreditation, which proves, from a quantitative point of view, what has been repeatedly argued qualitatively in the literature, namely, that Western European countries largely utilized evaluation tools over the course of the last twenty years (Neave 2012; Rosa and Amaral 2014).

Figure 2 *Regulation: which instruments are utilized the most (%)?*

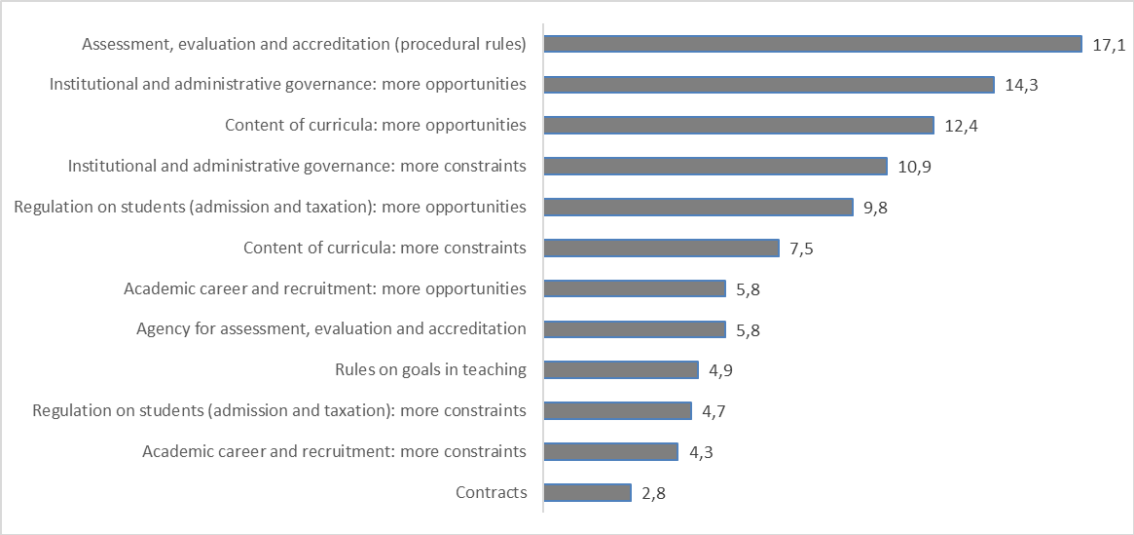
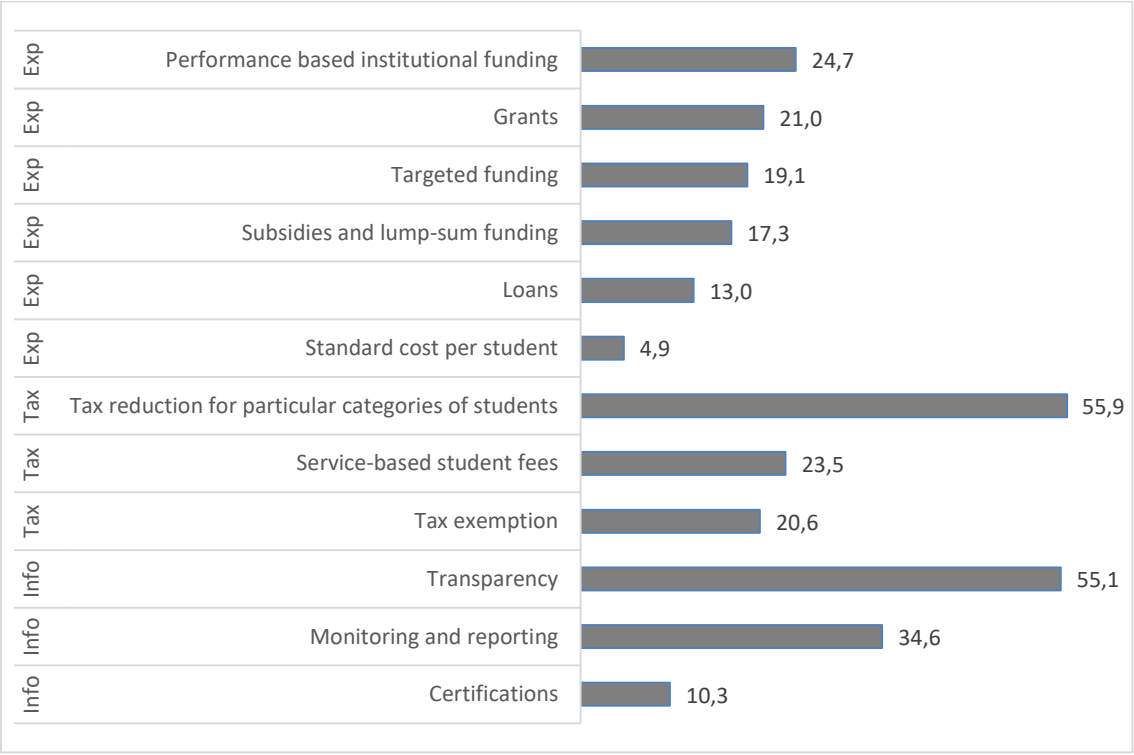


Figure 3 indicates which shapes were the most frequently used over the other three families of substantial instruments. Apparently, this general picture also confirms the trend towards a common template, but with some relevant specifications.

Figure 3 *Other families of substantial policy tools (expenditure, taxation, and information): which instruments are utilized the most (%)?*



In fact, expenditure was delivered not only through grants but also through targeted funding and, above all, through performance-based institutional funding. Thus, the most expected expenditure instrument in a pure 'steering at a distance' governance model – lump sum – was adopted less frequently than expected. Therefore, governments preferred a more coercive way to allocate public funding (together with an emerging attitude favoring the inclusion of families in paying HE loans) and thus demonstrated the will to maintain a certain degree of control over the behavior of universities.

Finally, regarding both taxation and information, it emerged that two instrumental shapes were particularly utilized: tax reduction for particular categories of students and transparency.

4.2 What can be associated with better teaching performance? A configurational analysis

We return to fuzzy-set QCA[12] to empirically explore the combinations of policy instruments and related shapes that might contribute to explaining teaching performance in Western European HESs between 1996 and 2015. While we refer to the Online Supplementary Material (Appendix A) for a careful discussion and justification of the thresholds chosen in the process of 'calibrating' the sets (both the conditions and the outcome) (Ragin 2008; Schneider and Wagemann 2012), this section presents the main results of our empirical analysis in terms of both the need for and sufficiency of relations between conditions, i.e., instrumental shapes, and the outcome, namely, (adjusted) teaching performance.

The analysis of necessary conditions for improving teaching performance was completed quickly: no condition (or its non-occurrence) was necessary for the outcome (or for its non-occurrence)[13]. In other words, we could not identify any conditions that needed to be present or absent to observe either a good teaching performance (presence of the outcome) or a bad teaching performance (absence of the outcome) between 1996 and 2015.

[12] We used the fs-QCA software downloaded from the website of Charles Ragin.

[13] All the consistency thresholds were lower than 0.9, which is the value above which empirical evidence supports the claim that a condition is necessary for the outcome (Schneider and Wagemann 2012, 278). See Appendix A for details.

However, much more interesting is the analysis of the *sufficient* conditions for improving teaching performance in HE, which was conducted with a 'truth table'. It clearly would not have been feasible to perform an analysis of sufficiency on 12 cases with dozens of different conditions (i.e., instrumental shapes) due to problems of limited diversity of the data (Schneider and Wagemann 2012; Vis 2012); thus, we turned to theoretical reflections to hypothesize a set of possible combinations to be tested in our fs-QCA.

First, we began with 45 theoretically relevant potential configurations of conditions that populated the literature on HES performance. These combinations of instrumental shapes build mainly on the literature emphasizing the adoption of similar policy instruments to follow the common template of the 'steering at a distance' model. The combinations also build on the empirical evidence and contradictions emerging from variable-oriented studies that focused on the determinants of performance in HE. Regarding the 'steering at a distance' literature, we have already emphasized the substantial literature underlying the ways governments have been changing governance in HE by improving institutional autonomy (and its dimensions, such as budgetary autonomy, degree of freedom in curricular content and autonomy in recruiting academic staff), quality assurance, accreditation, teaching and research assessments, monitoring, and varieties of funding mechanisms (Gornitzka *et al*. 2005; Cheps 2006; Lazzaretti and Tavoletti 2006; Maassen and Olsen 2007; Trakman 2008; Huisman 2009, Shattock 2014, Capano, Regini, and Turri 2016).

This literature clearly addressed our choice to consider the shapes for each family of substantial policy instruments that seemed to best represent the operative dimensions of the main categories of government intervention. This choice was reinforced by the contributions that focused on the real effects of performance-based funding on institutional autonomy and on the degree of centralization of the governance system. This literature produced contrasting empirical evidence and thus suggested a strategic dimension that we should take into consideration. For example, with respect to performance and targeted funding as a cause of high graduation and research rates, the relevant studies showed the weak performative capacity of these instruments. Most of these studies focused on the United States, where many states introduced performance criteria to determine the allocation of extra resources beginning at the end of the 1970s

(Rabovsky 2012; Tandberg and Hillman 2014; Volkwein and Talberg 2008; Rutheford and Rabovsky 2014). This evidence was quite contradictory compared to the widespread use of these types of instruments by the governments of our analyzed countries. Thus, we considered all targeted expenditure tools as relevant in our combinations.

Regarding institutional autonomy, contradictory evidence has emerged from reputable studies. For example, in their comparison between the EU and the US, Aghion and colleagues (2010) found that high institutional autonomy (and a competitive environment) was positively correlated with high performance both in educational attainment and in research. By contrast, Braga and colleagues (2013) showed that high institutional autonomy negatively impacted the level of educational attainment.

With respect to the level of the centralization of the governance system, Knott and Payne (2004) considered systemic centralization in US states to be high, intermediate, or low depending on the scope of the decision-making powers held by state boards. They tested systemic centralization as a condition affecting an array of resource and productivity measures, including the size of university revenue and the number of published articles. They concluded that state governance matters and that flagship universities are penalized by centralization, which reduces their total revenues, research funding, and number of published articles. However, centralization was also found to reduce tuition revenues and – presumably – the cost of enrolment. The worst overall performance occurs under mild centralization. However, this interpretation contrasts with a qualitative study of five US states conducted by Richardson and Martinez (2009). They argued that universities in centralized systems might perform better than those with decentralized designs with respect to access and graduation rates.

Thus, according to the literature, and assuming that governments have tried to pursue the 'steering at a distance' model in accordance with their own national identity, we attempted to combine these shapes in such a way that 45 different policy mixes – 30 consisting of four instrumental shapes belonging to different families of instruments and 15 consisting of five of those same instrumental shapes – were established (see Table AA4 in the Online Supplementary Material for those 45 different policy mixes).

Then, we performed 45 different QCAs – one for each of the selected policy mixes – to ascertain the combination of conditions that were more theoretically convincing and empirically robust. To accomplish this, we followed a two-step process. First, we focused

on the values of consistency and the coverage of different (intermediate) final solutions. In fact, in the QCA literature, it is generally argued that 'coefficients of consistency and coverage provide important numeric expressions for how well the logical statement contained in the QCA solution term fits the underlying empirical evidence and how much it can explain' (Schneider and Wagemann 2010, 414). Therefore, scholars have reviewed the coefficients to measure the goodness of fit of the tested theoretical models (Ragin 2006). This first step allowed us to realize that two combinations existed with both consistency and coverage values that were much higher than any other, as well as higher (or, at least, equal to) than 0.90, which is a notable value[14]. However, limiting our comparison to consistency and coverage alone would be insufficient (Braumoeller 2015; Rohlfing 2018); it is important to also account for the theoretical plausibility of solutions and the (groups of) countries that are identified by each solution term. For this reason, in our second step, we carefully scrutinized those same two combinations of instrumental shapes, looking for results that appeared to be more theoretically plausible and empirically grounded. In applying these further selection criteria, we realized that one combination was preferable to the other [15].

Thus, at the end of the above-described analytical process, the best combination of instrumental shapes that emerged from the QCA for explaining the teaching performance of the analyzed sample of countries between 1996 and 2015 consisted of R1 (assessment, evaluation and accreditation); R10 (institutional and administrative governance: more opportunities); R12 (rules on goals in teaching); E5 (performance-based institutional funding); and T3 (service-based student fees).

Table 4 *Explaining teaching performance: the best policy mix (intermediate solution)*

Solution terms	*Raw coverage*	*Unique coverage*	*Consistency*	*Cases covered*

[14] These combinations were 'Combination 32' and 'Combination 45'. For details about all configurations of conditions, see the Online Supplementary Material.

[15] For example, the third solution term of the (intermediate) final solution characterizing the 'Combination 32' consisted of three absences of instrumental shapes: it was the conjunct absence of R1 (Assessment, evaluation and accreditation), E6 (Standard cost per student) and T3 (Service-based student fees) that was associated with good teaching performance in England and Sweden. However, explaining the *presence* of the outcome with only *absences* of instrumental shapes appears to be fairly unsatisfactory: it seems that some factor remained out of the picture.

R12	0.62	0.52	0.88	Den (0.73, 0.93); Fin (0.95, 0.77); Ned (0.73, 0.77); Nor (0.95, 0.91); Swe (0.95, 0.98)
E5*~T3	0.39	0.29	0.91	Eng (0.86, 0.95); Gre (0.65, 0.76); Ire (0.86, 0.98)

Solution coverage (proportion of membership explained by all paths identified): 0.908847

Solution consistency ('how closely a perfect subset relation is approximated') (Ragin 2008, 44): 0.895641

Raw coverage: proportion of memberships in the outcome explained by a single path.

Unique coverage: 'proportion of memberships in the outcome explained solely by each individual solution term' (Ragin 2008, 86).

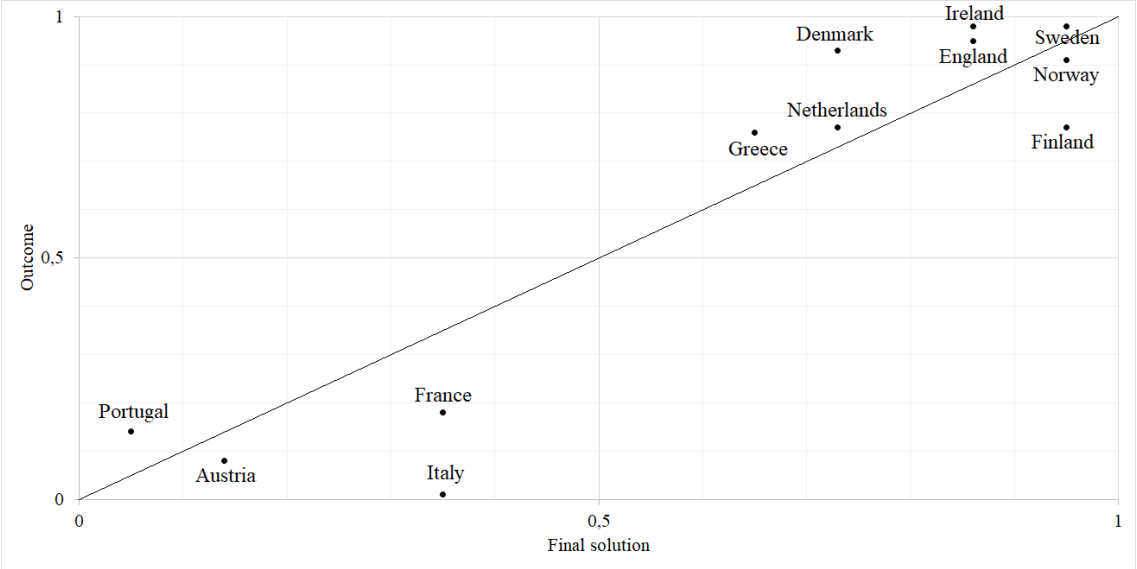
Empirically contradictory cases are shown in bold.

Complex solution: R12*~R1*~E5*~T3 + E5*~R10*~R12*~T3 + E5*~R1*~R12*~T3 + R10*R12*E5*T3 (coverage 0.85; consistency 0.95).

Parsimonious solution: R12 + R10*~R1 + E5*~T3 + T3*~R1 + R10*T3 + R1*~R10*~T3 (coverage 0.95; consistency 0.85).

The consistency value of the intermediate solution was impressive (0.90), and the coverage of the (intermediate) solution formula was even better (0.91). There was no 'deviant case for coverage' (Schneider and Rohlfing 2013, 585). As shown in Figure 4, six cases (i.e., Denmark, England, Greece, Ireland, the Netherlands and Sweden) were above the diagonal in the upper-right corner – and thus were 'typical cases' – whereas two cases (i.e., Finland and Norway) were 'deviant cases in consistency of degree.' Finally, the four cases (namely, Austria, France, Italy and Portugal) in the lower-left quadrant – being a good example of neither the solution terms nor the outcome – did not merit particular attention.

Figure 4 *The 'best policy mix': final XY plot*



The above analysis of policy mixes thus means that in our sample of countries, it was possible to identify two distinct paths associated with an increase in teaching performance. First, in the Scandinavian countries – Denmark, Finland, Norway and Sweden – together with the Netherlands, it emerged that the rules on systemic goals in teaching were the prevailing policy instrument, which clearly indicates the aims to be achieved by institutions. Second, in the Anglo-Saxon countries – England and Ireland – together with Greece, improved performance was associated with performance-based institutional funding; moreover, those three countries were also similar in that they did not pay very much attention to service-based student fees.

5. Discussion of the findings

This analysis showed how instrumental shapes were composed in the 12 analyzed countries, and we also treated these mixes with fuzzy-set QCA to assess which combinations of instrumental shapes were associated with better systemic performance in teaching. The empirical evidence is quite promising in both areas.

Regarding the composition of the national packages of instrumental shapes, the different distributions of substantial instruments and their possible shapes revealed that there was significant variance when mixing the same substantial instruments and their shapes. This variance, of course, calls for a better understanding of which combination(s) of instrumental shapes can be associated with better HES performance in terms of

teaching. Thus, this emerging variance in the composition of policy mixes incontrovertibly shows how the common template, the 'steering at a distance' mode of governing HESs, was applied in a very loose way, that is, as a generic framework that each country interpreted in a very specific and idiosyncratic way.

The findings emerging from our QCA treatment are relevant with respect to both the specific issue of identifying the most effective policy mixes in governing HESs and the more general theoretical and empirical problem of operationalizing governance shifts and their potential policy effectiveness.

The first issue is that among the eight cases showing the outcome, there is a clear-cut divide based on the presence of only one condition (instrumental shape). On one hand, there are countries in which the performance funding shape prevails, and thus, the government 'steers at a distance' by financially addressing the behavior of the institution. This is the case for England, Greece and Ireland, where the conjunction between performance funding and the absence of service-based student fees is sufficient for the outcome. Here, it is surprising that this solution groups together countries belonging to different traditions in governing HE.

On the other hand, the four Nordic countries plus the Netherlands are clustered together by the presence of a significant governmental role in indicating the systemic goals to be achieved, which is a sufficient condition for the outcome. In these countries, the government has significantly followed an autonomistic policy but has counterbalanced it through a clear indication of what the overall system is expected to deliver in terms of teaching performance.

This clear-cut result raises a relevant issue because it shows how the mix of instrumental shapes makes a difference regardless of the 'quantity' of shapes of the same family of substantial instruments that has been introduced. For example, in the Greek case, the association of performance funding with the outcome occurs in the presence of a high percentage of adopted regulatory instrumental shapes, while in the Norwegian and Finnish cases, the impact of a specific regulatory shape (R12, rules on goals in teaching) appears more relevant with respect to expenditure shapes, notwithstanding the finding that these two Nordic countries scored first in terms of the percentage of adoption of expenditure instrumental shapes.

This evidence raises a more general consideration that emerges from our analysis: the fact that each solution term highlights the relevance of the presence of only one instrumental shape. This finding suggests taking into consideration the hypothesis that some specific shapes of policy instruments make a difference regardless of the other shapes of policy instruments with which they work. This empirical evidence could appear simply because we chose only the most frequently adopted instruments according to the specialized literature or because the final results show the cumulative effect of different compositions of policy mixes adopted over time. However, the hypothesis that a small combination of instruments can make a significant difference irrespective of the other instruments that can be part of the actual adopted policy mix also deserves more empirical attention because – if confirmed – it could have a relevant impact on the actual trend in the literature on policy instruments and on policy performance. Obviously, this evidence could depend on different contexts and on different implementation practices (in terms of policy instruments, this means primarily the rules and organizational procedures through which instrumental shapes are implemented); it could also depend on a specific temporal sequence of adoption of policy instruments (for example, the relevance of both systemic goals and performance funding could require the presence of significant and effective previously adopted institutional autonomy on curricula and recruitment at work). However, while accepting this consideration, the emergence of the hypothesis that only a small number of instrumental shapes – or even only one shape, regardless of the other adopted shapes – can be associated with the presence of a performance improvement also appears intriguing and promising.

6. Conclusions and future research

We devoted this paper to addressing a general problem in analyzing governance shifts in public policy by empirically focusing on HE as an exemplary field. We assumed that the complexity of understanding whether and how governance changes should be analyzed from a detailed perspective that begins from the basic component of governmental actions and governance arrangements: policy instruments. We proposed a classification of substantial instruments (regulation, expenditure, taxation and information) to grasp the complete spectrum of induced behavior that can be addressed.

Then, we operationalized these substantial instruments according to 24 different shapes. We used this long list to code the instrumental choices made in 12 European countries over the last 20 years in their HES governance arrangements. Afterwards, we developed a fuzzy-set QCA to assess which combination of instrumental shapes is associated with good systemic performance in teaching, operationalized by taking into account the (adjusted) increase in the percentage of 25- to 34-year-old adults with a university-level degree. The empirical analysis allowed us to identify the most theoretically convincing and empirically robust combination of instrumental shapes associated with the teaching performance of the analyzed countries between 1996 and 2015. In more detail, it was possible to identify two distinct paths associated with the outcome: while in the Scandinavian countries and the Netherlands, rules on goals in teaching were the prevailing policy instrument, in Anglo-Saxon countries and Greece, good teaching performance was associated mainly with the presence of performance-based institutional funding, together with the absence of service-based student fees.

Of course, we are aware of the intrinsic limitations of our research design: first, the link between policy instruments and teaching performance is indirect; second, policy performance is driven by many factors that interact with the adopted instrumental shapes, and we were not able to assess the role of external factors *ex ante*. For example, expenditure instrumental shapes may have a different impact on teaching performance depending on whether a drastic reduction in public funding is occurring concomitantly; a peak in enrollment can increase the opportunity to produce more graduates, especially if this figure is linked to an increase in assigned grants; or an increase in institutional autonomy (and thus of greater competition among universities in their teaching offerings) may have different effects according to the general socio-economic situation of the country; and similar considerations can be drawn with regard to many more instrumental shapes.

Nonetheless, we are convinced that policy instruments can be intended as possible explanatory conditions (among others) and thus that a specific focus on them is absolutely needed to better understand and order what really occurs when governments decide to intervene in a policy field with the only means they have at disposal: by choosing specific instrumental shapes or combinations of them. Accordingly, the results of our exploration are quite interesting both for the study of performance in HE and, more generally, for the

study of the effects of governance shifts in public policy and, in turn, with regard to the portfolios/mix of adopted policy instruments.

With regard to the literature evaluating the performance of university systems, our empirical evidence shows that the same outcome is associated with a specific configuration of conditions (shapes of policy instruments) that must be present or absent to work. This way of thinking and, thus, this reading of the way governance arrangements work is explained by different combinations of instrumental shapes that can be associated with similar effects and, thus, by the way these combinations should be properly contextualized. In this sense, the empirical evidence presented in this paper shows that the evaluative literature on HE performance should find a third method that lies between the variable-oriented research strategy and the dense description of case-study analysis to fully grasp what may be important in terms of performance.

Furthermore, with respect to the literature on HE focusing on the bundle of changes that have been discussed from a comparative perspective towards a common template — the 'steering at a distance' model — there have clearly been certain national paths that have merged when translating the common template focused on certain instruments over others. Why these paths, these specific combinations of instruments, have been chosen is not of interest in this paper, but it could be taken into consideration in the future for a better understanding and explanation of the process of governance shifts, their features, their drivers and their decisional outputs.

Finally, regarding the broader literature on policy instruments mixes, this paper proposes a research design through which it is possible to overcome its actual limitations and thus a new way towards: 1) how to conceptualize and operationalize policy instruments to reach a better description of both the content of policy mixes, as well as 2) how they really work in terms of performance.

Regarding the content of policy mixes, we have shown how reasoning in terms of instrumental shapes can be useful for better grasping the multidimensionality of policy instruments. According to this way of operationalizing policy instruments, a more detailed picture of real adopted policy instruments can be obtained; thus, more fine-grained analyses, including analyses in comparative terms, are possible. This evidence calls for the suggestion of Ingram and Schneider (1990) and Salamon (2002) to be taken seriously: to grasp the real composition of policy mixes (and of their potential effects).

The various types of policy instruments should be disaggregated into smaller units, and thus, the shapes through which policy instruments are actually adopted when decision-makers design policies should be designed.

Regarding the functioning of policy mixes with respect to policy performance, this paper has shown (or, at least, has raised the intriguing hypothesis) that often, only a limited combination of instrumental shapes may truly be associated with the outcome of interest, despite the larger number of instrumental shapes that compose the adopted policy mix. Thus, it could be that only a few instrumental shapes can make a difference, regardless of the more articulated composition of the actual policy mix.

We are perfectly aware that another limitation of our analysis is that we did not consider the institutional 'impact' of instruments (or the content of the implementation process, where different institutional interpretations and strategies can be at work), but we were unable to do this because the analytical focus was on policy design and thus on the formal decisions made at the national level. Although we have focused on a specific dimension of policy instrument mixes, we believe that this could be a first necessary step toward a deeper analysis in terms of changes obtained over time. Overall, the way decision-makers design policies by arranging different instrumental shapes should matter.

There are obviously different possible paths for further research starting from the approach we have presented in this paper. We describe four of them here.

The first path would be to extend the research while including countries belonging to different geo-political contexts and with different legacies in governing HESs and, more ambitiously, to broaden the scope of the research by including and comparing different policy fields. This extension could also allow us to test whether there could be a way to find a list of instrumental shapes that have an inter-sectorial analytical use.

The second path would be to deepen the analysis in order to investigate the working rules of the shapes through which the substantial instruments are used (implementation practices). This step would also mean working on the dimensions of the rules through which each shape is designed. We refer to the rules through which decisional powers and competences are attributed; accountability rules are fixed when the shapes are designed. For example, regarding the use of performance funding, the focus should be on the percentage of public funding allocated, the criteria for allocating performance funding, the means and timing of evaluation, and so on. Obviously, this path

would be very complex, but it could be a very interesting and promising way to definitively elucidate how instruments work in day-to-day policy dynamics.

The third path would be to focus on a few exemplary national cases for a deeper analysis of the diachronic sequence of instrumental choices made by governments and of their cumulative effects in terms of association with performance. This analytical deepening could allow us to clarify whether some instrumental shapes matter more than others or whether this step is simply the final emergence of a specific sequence of instrumental choices.

The fourth path would be an analysis of the diachronic interaction between structural/environmental factors and policy mixes in determining the policy performance to better assess the 'association' (and maybe to upgrade it to the 'explanatory' level) between some instrumental shapes and improved performance that we found in our research.

We are definitively convinced that policy instruments matter, and in the future, more empirical work should be done to better understand how these instruments do their job.

REFERENCES

Aghion, Philippe, Mathias Dewatripont, Caroline Hoxby, Andreu Mas-Colell, and André Sapir. 2008. *Higher Aspirations: An Agenda for Reforming European Universities*. Brussels: Bruegel.

Aghion, Philippe, Mathias Dewatripont, Caroline Hoxby, Andreu Mas-Colell, and André Sapir. 2010. The Governance and Performance of Universities: Evidence from Europe and the US. *Economic Policy* 25(61):7–59.

Ahola, Sakari, and David Hofmann (eds.) 2012. *Higher education research in Finland*. Jyvaskyla: Jyvaskyla University Press.

Ahola, Sakari, and JaniMesikammen. 2003. Finnish Higher Education Policy and the Ongoing Bologna Process**.** *Higher Education in Europe* 28(2): 217-227.

Ansell, Ben W. 2008. University Challenges: Explaining Institutional Change in Higher Education. *World Politics* 60(1):189–230.

Bouwma, Irene, Duncan Liefferink, Rob Van Apeldoorn and Bas Arts. 2016. Following Old Paths or Shaping New Ones in Natura 2000 Implementation? Mapping Path Dependency in Instrument Choice. *Journal of Environmental Policy & Planning* 18(2): 214-233.

Braumoeller Bear F. 2015. Guarding against false positives in Qualitative Comparative Analysis. *Political Analysis* 23(4): 471–487.

Braun, Dietmar, and François-Xavier Merrien, eds. 1999. *Towards a New Model of Governance for Universities?*. London: Jessica Kingsley Publishers.

Braga, Michele, Daniele Checchi, and Elena Meschi. 2013. Educational Policies in a Long-Run Perspective. *Economic Policy* 28(73):45–100.

Bressers H., T.A. Laurence and J. O'Toole. 2005. Instrument Selection and Implementation in a Networked Context, in *Designing Government: From Instruments to Governance*, eds. P. Eliades, M. Hill and M. Howlett, pp. 132-153; Montreal & Kingston: McGill-Queen's University Press.

Capano, Giliberto. 2011. Government Continues to Do Its Job. A Comparative Study of Governance Shifts in the Higher Education Sector. *Public Administration* 89(4):1622–1642.

Capano, Giliberto, Jeremy Rayner, and Anthony R. Zito. 2012. Governance from the Bottom up: Complexity and Divergence in Comparative Perspective. *Public Administration*90(1):56–73.

Capano, Giliberto, and Marino Regini. 2014. Governance Reforms and Organizational Dilemmas in European Universities. *Comparative Education Review* 56(1):73–103.

Capano, Giliberto, Michael Howlett, and M. Ramesh, eds. 2015. *Varieties of Governance*. London: Palgrave.

Capano, Giliberto, Marino Regini, and Matteo Turri. 2016. *Changing Governance in Universities. Italian Higher Education in Comparative Perspective*. London: Palgrave MacMillan.

Capano, Giliberto, and Matteo Turri. 2017. Same Governance Template but Different Agencies. Types of Evaluation Agencies in Higher Education. Comparing England, France, and Italy. *Higher Education Policy* 30(2): 225–243.

Capano, Giliberto and Andrea Pritoni. 2019. Varieties of hybrid systemic governance in European Higher Education. *Higher Education Quarterly* 73(1).

CHEPS. 2006.*The extent and impact of higher education governance reform across Europe.* Final Report to the Directorate General for Education and Culture OfEuropean Commission. http://ec.europa.eu/education/**.**

Clark, Burton R. 1983. *The Higher Education System. Academic Organization in Cross National Perspective*. Berkeley: University of California Press.

Considine, Mark, and Jenny M. Lewis. 1999. Governance at Ground Level: The Frontline Bureaucrat in the Age of Markets and Networks". *Public Administration Review* 59 (6): 467–480.

Dobbins, Michael, and Christoph Knill. 2014. *Higher Education Governance and Policy Change in Western Europe*. Basingstoke: Palgrave Macmillan.

Enders, Jurgen, Harry De Boer, and Elke Weyer. 2013. Regulatory Autonomy and Performance: The Reform of Higher Education Revisited. *Higher Education*, 65(1): 5–23.

Flannery, Darragh, and Cathal O'Donoghue. 2011. The Life-cycle Impact of Alternative Higher Education Finance Systems in Ireland. *The Economic and Social Review* 42(3): 237–270.

Gornitzka, Åse, Maurice Kogan, and Alberto Amaral, eds. 2005. *Reform and Change in Higher Education*. Dordrecht: Springer.

Hacker Jacob S. 2004. Reform Without Change, Change Without Reform: The Politics of US Health Policy Reform in Comparative Perspective. In *Transatlantic Policymaking in an Age of Austerity: Diversity and Drift*, eds. Martin A. Levin and Martin Shapiro, 13–63. Washington, DC: Georgetown University Press.

Hood, Christopher. 1983. *The Tools of Government*. London: Macmillan.

Horeau, Cecile, Jo Ritzen, and Gabriele Marconi. 2013. Higher education and economic innovation, a comparison of European countries. *IZA Journal of European Labor Studie*s, 2(1): 2-24.

Howlett, Michael. 2000. Managing the 'Hollow State': Procedural Policy Instruments and Modern Governance. *Canadian Public Administration* 43(4):412–431.

Howlett, Michael. 2005. What is a policy instrument? Policy tools, policy mixes and policy-implementation styles. In *Designing government. From instruments to governance*, eds. Pearl Eliadis, Margaret M. Hill, and Michael Howlett, 311–350. Montreal and Kingston: McGill-Queen's University Press.

Howlett, Michael. 2009. Governance Modes, Policy Regimes and Operational Plans: A Multi-Level Nested Model of Policy Instrument Choice and Policy Design. *Policy Science*s 42(1): 73-89.

Howlett, Michael. 2011. *Designing Public Policies: Principles and Instruments*. London: Routledge.

Howlett, Michael, and Pablo Del Rio. 2015. The Parameters of Policy Portfolios: Verticality and Horizontality in Design Spaces and Their Consequences for Policy Mix Formulation. *Environment and Planning C: Politics and Space* 33:1233–1245.

Huisman, Jeroen, eds. 2009. *International Perspectives on the Governance of Higher Education*. London: Routledge.

Huitema, Dave, and Sander Meijerink, eds. 2009. *Water Policy Entrepreneurs: A Research Companion to Water Transitions Around the Globe*. Camberley, UK – Northampton, USA: Edward Elgar Publishing.

Knott, Jack H., and Abigail Payne. 2004. The Impact of State Governance Structures on Management and Performance of Public Organizations: A Study of Higher Education Institutions. *Journal of Policy Analysis and Management* 23(1): 13–30.

Koontz, Tomas M. and Craig W. Thomas. 2012. Measuring the performance of public-private partnerships: A systematic method for distinguishing outputs from outcomes. *Public Performance & Management Review* 35(4): 769–786.

Ingram, Helen, and Anne I. Schneider. 1990. The Behavioral Assumptions of Policy Tools. *The Journal of Politics,* 52(2): 510–529.

Joao Rosa, Maria and Alberto Amaral (eds.) 2014. *Quality Assurance in Higher Education*. London: Palgrave.

Jordan, Andrew, David Benson, Anthony Zito, and Rüdiger K.W. Wurzel. 2012. Environmental Policy: Governing by Multiple Policy Instruments? In *Constructing a Policy State? Policy Dynamics in the EU*, eds. Jeremy Richardson, 104–124. Oxford: Oxford University Press.

Lazzaretti, Luciana, Ernesto and Tavoletti. 2006. **Governance Shifts in Higher Education: A Cross National Comparison.** *European Educational Research Journal* 5(1):18–37.

Liefner, Ingo. 2003. Funding, Resource Allocation, and Performance in Higher Education Systems. *Higher Education* 46(4):469–489.

Linder, Stephen, and Guy B. Peters. 1990. An Institutional Approach to the Theory of Policymaking: The Role of Conscious Choice in Policy Formulation. *The Journal of Theoretical Politics* 2:59–83.

Maassen, Peter, and Johan P. Olsen, eds. 2007. *University Dynamics and European Integration*. Dordrecht: Springer.

Oecd. 2007.*Instruments Mixes for Environmental Policy*. Paris:Oecd.

Oecd. 2010.*OECD Science,TechnologyandIndustryOutlook 2010*. Paris: Oecd.

Paradeise, Catherine, EmanuelaReale, Ivar Bleiklie, and Ewan Ferlie, eds. 2009. *University Governance. Western European Comparative Perspectives*. Dordrecht: Springer.

Pérez DuránIxcel and Rodriguez Menés Jorge. 2017. Explaining accountability for public policies: an fsQCA analysis of health policy in Spain. *European Political Science*. Online publication first. DOI: 10.1017/S1755773915000405.

Phidd, Richard W., and Doern G. Bruce, eds. 1983. *Canadian Public Policy: Ideas, Structure, Process*- Toronto: Methuen.

Pierson, Christopher. 2004- *The Modern State*, London: Routledge.

Rabovsky, Thomas M. 2012. Accountability in Higher Education: Exploring Impacts on State Budgets and Institutional Spending Patterns. *Journal of Public Administration Research and Theory*. 22(4):675–700.

Ragin, Charles C. 1987. *The Comparative Method*. Berkeley: University of California Press.

Ragin, Charles C. 2000. *Fuzzy-set Social Science*. Chicago: The University of Chicago Press.

Ragin, Charles C. 2006. Set relations in social research: Evaluating their consistency and coverage. *Political Analysis* 14(3): 291–310.

Ragin, Charles C. 2008. *Redesigning Social Inquiry: Fuzzy Sets and Beyond*. Chicago: The University of Chicago Press.

Richardson, Richard, and Mario Martinez. 2009. *Policy and Performance in American Higher Education: An Examination of Cases Across State Systems*. Baltimore: The John Hopkins University Press.

Rihoux, Benoît, and Charles C. Ragin, eds. 2009. *Configurational Comparative Methods: Qualitative Comparative Analysis (QCA) and Related Techniques*. Los Angeles (USA): Sage.

Rihoux, Benoît, Priscilla Álamos-Concha, Damien Bol, Alex Marx, and Ilona Rezsöhazy. 2013. From Niche to Mainstream Method? A Comprehensive Mapping of QCA Application in Journal Articles from 1984 to 2011. *Political Research Quarterly* 66(1): 175–184.

Ring, Irene, and Christoph Schroter-Schlaack. 2010. *Instrument Mixes for Biodiversity Policies*. Leipzig: Helmholtz Centre for Environmental Research.

Rohlfing, Ingo. 2018. Power and False Negatives in Qualitative Comparative Analysis: Foundations, Simulation and Estimation for Empirical Studies. *Political Analysis* 26(1): 72-89.

Rosa, Maria Joao and Alberto Amaral (eds). 2014. *Quality Assurance in Higher Education*. Basingstoke, UK: Palgrave-Macmillan.

Rutherford, Amanda, and Thomas Rabovski. 2014. Evaluating Impacts of Performance Funding Policies on Student Outcomes in Higher Education. *The Annals of The American Academy of Political and Social Science* 655(1): 185–209.

Salamon, Lester M., ed. 2002. *The Tools of Government: A Guide to the New Governance*. Oxford: Oxford University Press.

Schaffrin, André, Sebastian Sewerin, and Sibylle Seubert. 2014. The Innovativeness of National Policy Portfolios – Climate Policy Change in Austria, Germany, and the UK. *Environmental Politics* 23(5):860–883.

Schneider, Carsten Q., and Ingo Rohlfing. 2013. Combining QCA and Process Tracing in Set-Theoretic Multi-Method Research. *Sociological Methods & Research* 42(4):559-597.

Schneider, Carsten Q., and Claudius Wagemann. 2010. Standards of Good Practice in Qualitative Comparative Analysis (QCA) and Fuzzy-Sets. *Comparative Sociology*, 9(3): 397-418.

Schneider, Carsten Q., and Claudius Wagemann. 2012. *Set-Theoretic Methods for the Social Sciences*. Cambridge: Cambridge University Press.

Shattock, Michael L., ed. 2014. *International Trends in University Governance*. London: Routledge.

Tandberg, David, and Nicholas Hillman 2014. State Higher Education Performance Funding: Data, Outcomes, and Policy Implications. *Journal of Education Finance* 39(3): 222–243.

Tollefson, Chris, Anthony R. Zito, and Fred Gale. 2012. Symposium Overview: Conceptualizing New Governance Arrangements. *Public Administration* 90(1): 3–18.

Tosun, Jale. 2013. *Risk Regulation in Europe. Assessing the Application of the Precautionary Principle*. New York: Springer.

Trakman, Leon. 2008. Modelling University Governance. *Higher Education Quarterly* 62(1-2):63–83.

Treib, Oliver, Holger Bahr, and Gerda Falkner. 2007. “Modes of Governance: Towards a Conceptual Clarification.” *Journal of European Public Policy* 14 (1): 1–20.

Vedung, Evert. 1998. Policy Instruments: Typologies and Theories. In *Carrots, Sticks, and Sermons: Policy Instruments and Their Evaluation*, eds. Marie-Louise Bemelmans-Videc, Ray C. Rist, and Evert Vedung, 21–58. New Brunswick, NJ: Transaction.

Vis, Barbara. 2012. The Comparative Advantages of fsQCA and Regression Analysis for Moderately Large-N Analysis. *Sociological Methods & Research* 41(1) 168–198.

Volkwein, J. Fredericks, and David Tandberg. 2008. Measuring Up: Examining the Connections Among State Structural Characteristics, Regulatory Practices, and Performance. *Research in Higher Education* 49(2):180–197.

Willemse, Nienke, and Paul De Beer. 2012. Three Worlds of Educational Welfare States? A Comparative Study of Higher Education Systems Across Welfare States. *Journal of European Social Policy* 22(2):105–117

Williams, Ross, Gaetan de Rassenfosse, Paul Jensen and Simon Marginson. 2013. The Determinants of Quality National Higher Education Systems. *Journal of Higher Education Policy and Management* 35(6), 599–611.

Winter-Ebmer, Rudolf, and Aniela Wirz. 2002. Public Funding and Enrolment into Higher Education in Europe. *IZA Discussion Papers 503*, Institute for the Study of Labor (IZA).

Woodside, Kenneth. 1983. The Political Economy of Policy Instruments: Tax Expenditures and Subsidies. In *The Politics of Canadian Public Policy*, eds. Michael M. Atkinson and Marsha A. Chandler, 173–197. Toronto: University of Toronto Press.

Online Supplementary Material: Appendix A – Tables of the QCA

Calibration

The first step in each fs-QCA is the 'calibration' of sets (both the conditions and the outcome). In this fundamental process, which should be as transparent as possible and discussed in detail, it is crucial to specify qualitative anchors for full membership (1), full non-membership (0) and the point of maximum ambiguity (0.5)[16]. Table AA1 summarizes these decisions:

Table AA1 *Calibration of sets (conditions and the outcome) and fuzzy values of countries*

	Countries												*Thresholds*		
	A	*D*	*E*	*FI*	*FR*	*G*	*IR*	*IT*	*NE*	*NO*	*P*	*S*	*1*	*0.5*	*0*
R1	0.14	0.23	0.05	0.23	**0.82**	0.14	**0.82**	**0.98**	0.23	**0.99**	**0.73**	0.35	12	6	1
R2	**0.65**	**0.95**	0.05	0.05	**0.86**	**0.95**	**0.86**	**0.65**	0.27	**0.86**	0.65	**0.86**	4	1.5	0
R3	0.14	0.35	0.14	**1**	**0.95**	**0.99**	0.05	**0.95**	**0.95**	**0.73**	0.05	**0.99**	4	2.5	0
R4	**0.68**	**1**	0.05	0.12	**0.82**	**0.9**	0.02	0.27	**0.82**	**0.82**	0.12	**0.95**	8	4	1
R5	0.05	0.05	0.27	0.27	**0.65**	**0.86**	**0.65**	**1**	0.05	0.27	**0.86**	0.27	4	1.5	0
R6	0.18	0.05	0.18	0.18	**0.88**	**0.73**	**0.95**	**0.98**	0.18	0.05	**0.88**	0.18	5	2	0
R7	**0.86**	**0.65**	0.27	0.05	0.27	0.27	0.27	0.05	**1**	**0.86**	0.05	**0.86**	4	1.5	0
R8	**0.95**	**0.65**	**0.65**	**1**	0.39	0.39	0.05	0.11	**0.95**	0.39	0.05	**0.86**	6	3.5	0
R9	0.05	0.18	0.32	**0.99**	**0.82**	**0.95**	**0.99**	**0.82**	0.18	0.1	0.05	0.1	7	5	1
R10	0.32	**0.97**	0.02	**0.65**	**0.65**	**0.77**	0.18	0.1	**0.77**	**1**	0.32	0.18	10	5	1
R11	**1**	**0.95**	**0.73**	0.05	**0.95**	**0.73**	**0.73**	**0.73**	0.05	**0.73**	0.05	0.05	2	0.5	0
R12	0.14	**0.73**	0.05	**0.95**	0.14	0.14	0.05	0.35	**0.73**	**0.95**	0.05	**0.95**	4	2.5	0
E1	0.35	**0.65**	0.14	**0.86**	0.35	**0.65**	0.05	0.35	**0.95**	**0.95**	**0.86**	**0.65**	5	2.5	0
E2	**0.65**	0.14	0.05	**1**	0.35	**0.95**	0.14	0.14	0.35	**0.65**	0.14	0.35	5	2.5	0
E3	0.12	**1**	0.05	**1**	**0.82**	0.27	0.27	0.12	0.27	0.27	0.05	0.27	5	3	0
E4	0.27	**0.65**	**0.65**	0.05	0.27	**0.65**	0.05	**0.65**	**0.95**	**0.95**	0.27	**0.65**	4	1.5	0
E5	0.14	0.14	**0.86**	**1**	0.35	**0.65**	**0.86**	**0.95**	0.05	**1**	0.35	0.05	5	2.5	0
E6	**0.73**	**0.73**	0.05	**0.95**	0.05	0.05	**0.73**	**0.95**	0.05	**0.73**	0.05	0.05	2	0.5	0
T1	**1**	**0.95**	0.05	**0.95**	**0.95**	0.05	**0.95**	**0.95**	0.05	0.05	0.05	0.05	1	0.5	0
T2	**0.65**	0.05	**0.99**	0.27	**0.86**	0.05	**0.86**	**0.95**	0.27	0.05	0.05	0.05	4	1.5	0
T3	0.05	0.05	0.05	**1**	0.05	0.05	0.05	**0.95**	0.05	**1**	**1**	0.05	1	0.5	0
I1	0.39	**0.65**	0.22	0.22	0.39	**0.65**	**0.86**	**1**	0.39	**0.95**	0.11	0.05	6	3.5	0
I2	**0.73**	0.05	**0.73**	**0.73**	**0.73**	0.05	0.05	0.05	0.05	**0.73**	0.05	**0.99**	2	0.5	0
I3	**1**	**0.86**	0.05	0.27	0.27	**0.65**	**0.65**	**0.99**	0.27	**0.86**	0.27	**0.65**	4	1.5	0
Out.	0.08	**0.93**	**0.95**	**0.77**	0.18	**0.76**	**0.98**	0.01	**0.77**	**0.91**	0.14	**0.98**	30	25	20

Cases that have membership in a specific condition (>0.5) are shown in bold.

[16] We use the direct method of calibration (Ragin 2008, 85): once qualitative anchors have been chosen, QCA software applies a logarithmic function and attributes fuzzy values to the remaining cases.

Hence, the table below (AA2) summarizes the calibration decisions for the conditions for the 'best' configuration of conditions/policy shapes, as well as the outcome. These decisions are also discussed in detail below.

Table AA2 *Calibration of conditions and the outcome*

Conditions/outcome	*Full membership (1)*	*Point of maximum ambiguity (0.5)*	*Full non-membership (0)*
Outcome: Indicator of HE teaching performance	+30 Cases: IR; S	+25 1<cases<0.5: E; D; F; G; NE; NO 0.5<cases<0: F; P; A	+20 Cases: IT
R1: Assessment, evaluation and accreditation (procedural rules)	12 Cases: NO; IT	6 1<cases<0.5: FR; IR; P 0.5<cases<0: A; D; FI; G; NE; S	1 Cases: E
R10: Institutional and administrative governance: more opportunities	10 Cases: NO; D	5 1<cases<0.5: FI; FR; G; NE 0.5<cases<0: A; IR; IT; P; S	1 Cases: E
R12: Rules on goals in teaching	4 Cases: FI; NO; S	2.5 1<cases<0.5: D; NE 0.5<cases<0: A; FR; G; IT	0 Cases: E; IR; P
E5: Performance based institutional funding	5 Cases: FI; IT; NO	2.5 1<cases<0.5: E; G; IR 0.5<cases<0: A; D; FR; P	0 Cases: NE; S
T3: Service-based student fees	1 Cases: F; IT; NO; P	/	0 Cases: A; D; E; F; G; IR; NE; S

Calibration

The calibration of conditions and the outcome needs to be justified at length, especially where the point of maximum ambiguity is concerned. However, due to the limited space at our disposal, we decided to explain our choices only with regard to the conditions for the combination of policy instruments we present in the empirical section

of the paper. Otherwise, this appendix would have been extremely long and very difficult to read and understand.

Regarding R1 (the rules of accreditation, evaluation and assessment), we were quite selective in choosing 6 as the point of maximum ambiguity. This choice was made according to the theoretical assumption that these types of regulatory activities are based on the will of the decision makers, who have decided to delegate some autonomy to universities and to monitor and check whether those universities are behaving as expected. Thus, this approach could be considered a specific application of the principal/agent theory, in which there is a constant risk of moral hazard (Moe 1984): universities, in fact, can interpret the accreditation of degree courses simply as a compliance process, without any real attention to substantial quality, or can be capable of finding 'opportunistic' strategies to address the periodic evaluation of research (Lane and Kivisto 2008). Thus, to avoid this risk, decision makers are expected to intervene often on the rules of accreditation, evaluation and assessment (Schwarz and Westerheijden 2007; Joao Rosa and Amaral 2014): these recurrent interventions, in fact, are assumed to amend any unwanted results. According to these assumptions, we rewarded the countries (Norway, Italy, France, Ireland and Portugal) that have shown attention over time to the rules and regulations of accreditation, evaluation and assessment, thus charging the related instrumental shape of specific regulative relevance in their HE policies.

Regarding R10 (more opportunities for universities concerning institutional governance and administrative procedures), according to the comparative literature on the diachronic evolution of institutional autonomy (Huisman 2009; Shattock 2014; Capano, Regini and Turri 2016), we based our calibration on the assumption that the process of 'autonomization' for universities (in countries where universities were historically not autonomous, that is, in 10 out of 12 in our sample) is not a one-shot game but requires more than one decision to find an equilibrium point, whereas exactly because institutional autonomy is given by the state, the state can intervene at various times to adjust the autonomistic policy. Importantly, the process of autonomization requires time and various interventions to reach a stable equilibrium. Thus, we properly rewarded all the countries that, in the considered period of time, repeatedly intervened in enlarging the margins of institutional autonomy for universities (Norway, Denmark, Finland, France, Greece and the Netherlands). We fixed 5 as the point of maximum ambiguity based on

the consideration that these rules need continuous calibration and that having one intervention every 4 years or more can be considered an indicator of carelessness or bad governing. It should be noted that among the six countries below the point of maximum ambiguity, Sweden significantly increased the institutional autonomy of its universities before 1995, while institutional autonomy in England was already very high in the 1980s.

The calibration of instrumental shape R12 (rules on teaching goals) required particular reflection, especially because the specialized literature on HE was not helpful (no empirical or analytical attention has been given to this dimension of governmental steering). However, according to the 'steering at a distance' template (Van Vught 1989; Capano 2011), governments should be more interested in fixing the systemic goals than in controlling procedures and processes. Thus, governments are expected to fix some systemic goals that address universities' behavior. For this reason, we proceeded by rewarding the countries that made more decisions to address systemic goals in teaching (Finland, Norway and Sweden). Then, we fixed the point of maximum ambiguity to reward those countries that reiterated the decision on this issue more than twice (Denmark and the Netherlands), showing that this choice was part of a conscious medium-run strategy and was not based on chance or contingency.

The instrumental shape E5 (performance-based institutional funding) has also been used a few times by certain countries. Here, the calibration was based on the following assumptions (Frølich, Schmidt and Roma 2010; Flannery D. & C. O'Donoghue 2011; Jonkers and Zacharewicz 2016; Zacharewicz *et al*. 2018):

1. Performance funding is one of the fundamental characteristics of the 'steering at a distance' template.
2. Performance funding is a suitable instrumental shape through which governments can address institutional behavior.
3. Performance funding can undergo moral hazard and can show some unexpected outcomes (if designed in an inappropriate way).
4. Performance funding is based on goals that can change over time.

According to these assumptions, we should expect that governments can intervene at various times to change the actual rules of performance funding (to correct design mistakes and unexpected results and to overcome moral hazard due to changing goals).

Thus, we rewarded the countries that intervened more than one time per decade (this approach justifies 2.5 as the point of maximum ambiguity).

The calibration of T3 is very problematic to theorize. It was included in the list of instrumental shapes according to the main types of fees working in other policy fields, but we expected it to be almost non-existent in HE, where tuition fees are based on criteria such as covering all or part of the costs or are income based. For this reason, we assumed that the use of this type of instrumental shape should be rare and almost experimental; thus, we theorized it in a dichotomous way: it is either absent or present. In other words, this condition is more crisp than fuzzy.

Overall, the calibration of conditions did not stem only from theoretical considerations but was also conducted by looking at whether countries remained above or below the point of maximum ambiguity. Although QCA scholars strongly recommend returning exclusively to 'theoretical' rather than empirical or even 'arithmetical' considerations when locating thresholds for calibration (Schneider and Wagemann 2010, 403), with regard to policy tools, it was difficult to find theoretical reasons suggesting a particular value in the literature, and we were forced to propose our own theorizations. In contrast, many qualitative studies assess whether a given country devoted particular attention to a set of policy instruments when shaping its HES, whereas other countries followed a different path. Thus, our choices originated from a mix of theoretical reasons and reasons based on previous empirical research.

Finally, for the outcome, we considered an increase of 30 percentage points in our indicator of teaching performance as the threshold for full membership, an increase of 25 percentage points as the point of maximum ambiguity, and an increase of 20 percentage points in the average value of the prior twenty years as the threshold for full non-membership. This choice was not simple because no study addresses this issue. We assumed that the countries that were capable of improving their performance by an average of at least 1.5 percentage points per year could be very effective. In addition, the point of maximum ambiguity was difficult to maintain because it affected the final assessment of the analyzed countries. Overall, the selection procedure placed all the countries that started with higher scores in 1996 above the ambiguity point to reward their efforts related to the outcome. Moreover, the gap between Greece (26.89) and France (22.50) appeared to be a 'natural' place to put the maximum ambiguity point (25.00).

Finally, in a context of diffused increases, a pace of 1.0 point per year seemed a reasonable threshold for full non-membership.

Analysis of the necessary conditions for improving teaching performance

Once the sets were calibrated, the second step of each QCA involves analyzing the necessary relations and should always be conducted *before* analyzing the sufficiency conditions (Schneider and Wagemann 2010, 404). See Table AA3 on this topic:

Table AA3 *Analysis of necessary conditions. Outcome: (adjusted) variation in the % of 25-34-year-old adults attaining a university degree (1996-2015)*

	Outcome		*~Outcome*	
Condition	*Consistency*	*Coverage*	*Consistency*	*Coverage*
R1	0.45	0.59	0.78	0.62
~R1	0.71	0.84	0.48	0.35
R2	0.68	0.66	0.78	0.46
~R2	0.45	0.77	0.43	0.44
R3	0.65	0.67	0.64	0.40
~R3	0.41	0.65	0.47	0.45
R4	0.64	0.73	0.60	0.42
~R4	0.49	0.67	0.61	0.51
R5	0.40	0.57	0.73	0.63
~R5	0.74	0.82	0.50	0.33
R6	0.39	0.54	0.80	0.67
~R6	0.76	0.86	0.45	0.31
R7	0.58	0.79	0.45	0.37
~R7	0.54	0.61	0.75	0.52
R8	0.65	0.75	0.53	0.38
~R8	0.46	0.62	0.65	0.53
R9	0.49	0.66	0.58	0.48
~R9	0.61	0.71	0.59	0.41
R10	0.64	0.81	0.51	0.39
~R10	0.51	0.63	0.75	0.56
R11	0.58	0.64	0.69	0.46
~R11	0.51	0.73	0.46	0.40
R12	0.62	0.88	0.34	0.29
~R12	0.50	0.56	0.86	0.58
E1	0.67	0.73	0.63	0.42

~E1	0.47	0.68	0.60	0.53
E2	0.49	0.74	0.49	0.45
~E2	0.64	0.67	0.72	0.46
E3	0.46	0.76	0.45	0.46
~E3	0.67	0.67	0.76	0.46
E4	0.64	0.79	0.49	0.37
~E4	0.49	0.61	0.72	0.55
E5	0.63	0.73	0.56	0.40
~E5	0.49	0.65	0.63	0.51
E6	0.45	0.65	0.52	0.46
~E6	0.63	0.68	0.61	0.40
T1	0.43	0.53	0.72	0.54
~T1	0.63	0.79	0.38	0.29
T2	0.38	0.56	0.70	0.62
~T2	0.74	0.80	0.51	0.33
T3	0.30	0.51	0.54	0.57
~T3	0.75	0.73	0.53	0.32
I1	0.58	0.74	0.62	0.48
~I1	0.59	0.72	0.66	0.49
I2	0.49	0.75	0.47	0.43
~I2	0.62	0.66	0.72	0.47
I3	0.63	0.69	0.75	0.50
~I3	0.55	0.78	0.53	0.46

Coefficients indicating a necessity relation (>0.9) are shown in bold.
Logical No (~) refers to the absence of a condition.

Table AA3 demonstrates that no condition (and its non-occurrence, which is indicated with a tilde [~]) is necessary for the outcome.

Analysis of the sufficient conditions for improving teaching performance

Table AA4 *Policy instrument mixes from the literature*

Policy mix	*Combination of instrumental shapes*
Four conditions	
Combination 1	R1 (Assessment, evaluation and accreditation); R10 (Institutional and administrative governance: more opportunities); E2 (Subsidies and lump-sum funding); I1 (Transparency)
Combination 2	R2 (Agency for assessment, evaluation and accreditation); E4 (Loans); E5 (Performance based institutional funding); I1 (Transparency)

Combination 3	R2 (Agency for assessment, evaluation and accreditation); R4 (Content of curricula: more opportunities); R10 (Institutional and administrative governance: more opportunities); E5 (Performance based institutional funding)
Combination 4	R4 (Content of curricula: more opportunities); R6 (Academic career and recruitment: more opportunities); E1 (Grants); E3 (Targeted funding)
Combination 5	R1 (Assessment, evaluation and accreditation); R10 (Institutional and administrative governance: more opportunities); R12 (Rules on goals in teaching); E3 (Targeted funding)
Combination 6	R1 (Assessment, evaluation and accreditation); R10 (Institutional and administrative governance: more opportunities); R12 (Rules on goals in teaching); E5 (Performance based institutional funding)
Combination 7	R2 (Agency for assessment, evaluation and accreditation); R10 (Institutional and administrative governance: more opportunities); R12 (Rules on goals in teaching); E3 (Targeted funding)
Combination 8	R2 (Agency for assessment, evaluation and accreditation); R10 (Institutional and administrative governance: more opportunities); R12 (Rules on goals in teaching); E5 (Performance based institutional funding)
Combination 9	R10 (Institutional and administrative governance: more opportunities); R11 (Contracts); T2 (Tax reduction for particular categories of students); I1 (Transparency)
Combination 10	R1 (Assessment, evaluation and accreditation); R10 (Institutional and administrative governance: more opportunities); E5 (Performance based institutional funding); T1 (Tax exemption)
Combination 11	R8 (Regulation on students admission and taxation: more opportunities); T2 (Tax reduction for particular categories of students); I1 (Transparency); I3 (Monitoring and reporting)
Combination 12	R10 (Institutional and administrative governance: more opportunities); T2 (Tax reduction for particular categories of students); I1 (Transparency); I3 (Monitoring and reporting)
Combination 13	R1 (Assessment, evaluation and accreditation); R9 (Institutional and administrative governance: more constraints); E3 (Targeted funding); E5 (Performance based institutional funding)
Combination 14	E3 (Targeted funding); E5 (Performance based institutional funding); I1 (Transparency); I3 (Monitoring and reporting)
Combination 15	R4 (Content of curricula: more opportunities); E3 (Targeted funding); I1 (Transparency); I3 (Monitoring and reporting)
Combination 16	R4 (Content of curricula: more opportunities); E5 (Performance based institutional funding); I1 (Transparency); I3 (Monitoring and reporting)
Combination 17	R10 (Institutional and administrative governance: more opportunities); E3 (Targeted funding); I1 (Transparency); I3 (Monitoring and reporting)
Combination 18	R10 (Institutional and administrative governance: more opportunities); E5 (Performance based institutional funding); I1 (Transparency); I3 (Monitoring and reporting)

Combination 19	R1 (Assessment, evaluation and accreditation); R4 (Content of curricula: more opportunities); R10 (Institutional and administrative governance: more opportunities); E5 (Performance based institutional funding)
Combination 20	R1 (Assessment, evaluation and accreditation); R4 (Content of curricula: more opportunities); R10 (Institutional and administrative governance: more opportunities); E3 (Targeted funding)
Combination 21	R1 (Assessment, evaluation and accreditation); R2 (Agency for assessment, evaluation and accreditation); R10 (Institutional and administrative governance: more opportunities); R12(Rules on goals in teaching)
Combination 22	R9 (Institutional and administrative governance: more constraints); R10 (Institutional and administrative governance: more opportunities); E5 (Performance based institutional funding); E6 (Standard cost per student)
Combination 23	R9 (Institutional and administrative governance: more constraints); R10 (Institutional and administrative governance: more opportunities); E3 (Targeted funding); E5(Performance based institutional funding)
Combination 24	R4 (Content of curricula: more opportunities); R6 (Academic career and recruitment: more opportunities); R10 (Institutional and administrative governance: more opportunities); E5 (Performance based institutional funding)
Combination 25	R4 (Content of curricula: more opportunities); R6 (Academic career and recruitment: more opportunities); R10 (Institutional and administrative governance: more opportunities); E3 (Targeted funding)
Combination 26	R10 (Institutional and administrative governance: more opportunities); I1 (Transparency); I2 (Certifications); I3 (Monitoring and reporting)
Combination 27	R10 (Institutional and administrative governance: more opportunities); E3 (Targeted funding); E5 (Performance based institutional funding); I3 (Monitoring and reporting)
Combination 28	R8 (Regulation on students admission and taxation: more opportunities); R10 (Institutional and administrative governance: more opportunities); T1 (Tax exemption); T3 (Service-based student fees)
Combination 29	R8 (Regulation on students admission and taxation: more opportunities); R10 (Institutional and administrative governance: more opportunities); E5 (Performance based institutional funding); E6 (Standard cost per student)
Combination 30	R10 (Institutional and administrative governance: more opportunities); R12 (Rules on goals in teaching); E5 (Performance based institutional funding); I3 (Monitoring and reporting)
Five conditions	
Combination 31	R1 (Assessment, evaluation and accreditation); R2 (Agency for assessment, evaluation and accreditation); R10 (Institutional and administrative governance: more opportunities); R12 (Rules on goals in teaching); E5 (Performance based institutional funding)
Combination 32	R9 (Institutional and administrative governance: more constraints); R10 (Institutional and administrative governance: more opportunities); E5 (Performance based institutional funding); E6 (Standard cost per student); T3 (Service-based student fees)
Combination 33	R9 (Institutional and administrative governance: more constraints); R10 (Institutional and administrative governance: more opportunities); R11 (Contracts); E3 (Targeted funding); E5 (Performance based institutional funding)

Combination 34	R4 (Content of curricula: more opportunities); R6 (Academic career and recruitment: more opportunities); R10 (Institutional and administrative governance: more opportunities); E5 (Performance based institutional funding); I3 (Monitoring and reporting)
Combination 35	R4 (Content of curricula: more opportunities); R6 (Academic career and recruitment: more opportunities); R10 (Institutional and administrative governance: more opportunities); E3 (Targeted funding); E5 (Performance based institutional funding)
Combination 36	R10 (Institutional and administrative governance: more opportunities); R11 (Contracts); I1 (Transparency); I2 (Certifications); I3 (Monitoring and reporting)
Combination 37	R10 (Institutional and administrative governance: more opportunities); R12 (Rules on goals in teaching); E3 (Targeted funding); E5 (Performance based institutional funding); I3 (Monitoring and reporting)
Combination 38	R1 (Assessment, evaluation and accreditation); R10 (Institutional and administrative governance: more opportunities); R12 (Rules on goals in teaching); E5 (Performance based institutional funding); I1 (Transparency)
Combination 39	R1 (Assessment, evaluation and accreditation); R10 (Institutional and administrative governance: more opportunities); R12 (Rules on goals in teaching); E5 (Performance based institutional funding); I3 (Monitoring and reporting)
Combination 40	R1 (Assessment, evaluation and accreditation); R10 (Institutional and administrative governance: more opportunities); T2 (Tax reduction for particular categories of students); I1 (Transparency); I3 (Monitoring and reporting)
Combination 41	R1 (Assessment, evaluation and accreditation); R4 (Content of curricula: more opportunities); R10 (Institutional and administrative governance: more opportunities); E3 (Targeted funding); E5 (Performance based institutional funding)
Combination 42	R1 (Assessment, evaluation and accreditation); R4 (Content of curricula: more opportunities); R10 (Institutional and administrative governance: more opportunities); E5 (Performance based institutional funding); I3 (Monitoring and reporting)
Combination 43	R4 (Content of curricula: more opportunities); R6 (Academic career and recruitment: more opportunities); E3 (Targeted funding); E5 (Performance based institutional funding); I3 (Monitoring and reporting)
Combination 44	R9 (Institutional and administrative governance: more constraints); R10 (Institutional and administrative governance: more opportunities); E3 (Targeted funding); E5 (Performance based institutional funding); T3 (Service-based student fees)
Combination 45	R1 (Assessment, evaluation and accreditation); R10 (Institutional and administrative governance: more opportunities); R12 (Rules on goals in teaching); E5 (Performance based institutional funding); T3 (Service-based student fees)

Table AA5 *Configurations of instrumental shapes in a comparative perspective: consistency, coverage, solution terms and cases covered*

Mix	*Cons*	*Cov*	*Sol_1*	*Cases*	*Sol_2*	*Cases*	*Sol_3*	*Cases*	*Sol_4*	*Cases*
4 conditions										

Combination 1	0.82	0.79	~R1	**A**; D; E; FI; G; NE; S	R10* E2* I1	G; NO	/	/	/	/
Combination 2	0.81	0.73	Six (6) solution terms: too many							
Combination 3	0.80	0.62	Coverage value is too low							
Combination 4	0.87	0.75	~R4* ~R6	E; FI	R4* E1	D; G; NE; NO; S	/	/	/	/
Combination 5	0.85	0.75	R12	D; FI; NE; NO; S	R10* ~R1	D; FI; G; NE	R10* ~E3	G; NE; NO	/	/
Combination 6	0.89	0.83	R12	D; FI; NE; NO; S	E5* R1	E; FI; G	/	/	/	/
Combination 7	0.87	0.70	R10* R12	D; FI; NE; NO	R2* R12	D; NO; S	R2* R10* ~E3	G; NO	/	/
Combination 8	0.87	0.69	Coverage value is too low							
Combination 9	0.74	0.81	Consistency value is too low							
Combination 10	0.87	0.81	~R1* ~T1	E; G; NE; S	R10* ~R1	D; FI; G; NE	R10* E5	FI; G; NO	/	/
Combination 11	0.89	0.79	Five (5) solution terms: too many							
Combination 12	0.89	0.85	I3* ~T2	D; G; NO; S	R10* ~T2	D; FI; G; NE; NO	T2* ~R1* ~I3	E	R10* I1*I3	D; G; NO
Combination 13	0.94	0.60	Coverage value is too low							
Combination 14	0.79	0.64	Coverage value is too low							
Combination 15	0.74	0.77	Consistency value is too low							
Combination 16	0.77	0.89	R4	**A**; D; **FR**; G; NE; NO; S	E5* ~I1	E; FI	E5* ~I3	E; FI	/	/
Combination 17	0.79	0.83	~I3	E; FI; **FR**; NE; **P**	R10	D; FI; **FR**; G; NE; NO	/	/	/	/

Combination 18	0.87	0.65	Coverage value is too low							
Combination 19	0.90	0.74	E5* ~R1* ~R4	E; FI	R4* ~R1* R10	D; G; NE	R4* R10* E5	G; NO	/	/
Combination 20	0.92	0.74	~R1* ~R4	E; FI	R10* ~R1	D; FI; G; NE	R4* R10* ~E3	G; NE; NO	/	/
Combination 21	0.86	0.87	~R2	E; FI; NE	R12	D; FI; NE; NO; S	R10* ~R1	D; FI; G; NE	/	/
Combination 22	0.87	0.74	E5* ~R9	E; NO	E5* ~E6	E; G	R10* ~R9	D; NE; NO	/	/
Combination 23	0.86	0.63	Coverage value is too low							
Combination 24	0.90	0.72	E5* ~R4* ~R6	E; FI	R4* ~R6* R10	D; NE; NO	R4* R10* E5	G; NO	/	/
Combination 25	0.93	0.74	~R4* ~R6	E; FI	R10* ~R6	D; FI; NE; NO	R4* R10* ~E3	G; NE; NO	/	/
Combination 26	0.83	0.77	R10	D; FI; **FR**; G; NE; NO	I2* ~I3	E; FI; **FR**				
Combination 27	0.89	0.77	E5* ~I3	E; FI	R10* ~E3	G; NE; NO	R10* I3	D; G; NO	/	/
Combination 28	0.91	0.61	Coverage value is too low							
Combination 29	0.87	0.63	Coverage value is too low							
Combination 30	0.87	0.81	E5* ~I3	E; FI	R12* I3	D; NO; S	R10* E5	FI; G; NO	R10* R12	D; FI; NE; NO
5 conditions										
Combination 31	0.85	0.82	Five (5) solution terms: too many							
Combination 32	0.90	0.92	R10* ~R9	D; NE; NO	R10* E5	FI; G; NO	~R1* ~E6* ~T3	E; S	R9* E5* ~T3	G; IR
Combination 33	0.86	0.71	Coverage value is too low							
Combination 34	0.88	0.78	E5* ~R6	E; FI; NO	E5* ~I3	E; FI	R10* ~R4* ~R6	D; NE; NO	R4* R10* E5	G; NO

Combination 35	0.91	0.74	E5* ~R4* ~R6	E; FI	R4* ~R6* R10	D; NE; NO	R10* E3* E5	FI	R4* R10* E5	G; NO
Combination 36	0.89	0.74	R10* ~R11	FI; NE	I2* I3* ~R11	S	R10* I1*I3	D; E; G; NO	R11* ~R1* I2* ~I3	E
Combination 37	0.87	0.81	E5* ~I3	E; FI	R12* I3	D; NO; S	R10* E5	FI; G; NO	R10* R12	D; FI; NE; NO
Combination 38	0.85	0.83	R12	D; FI; NE; NO; S	E5* ~R1	E; FI; G	E5* ~I1	E; FI	R10* E5	FI; G; NO
Combination 39	0.87	0.81	Five (5) solution terms: too many							
Combination 40	0.89	0.86	Five (5) solution terms: too many							
Combination 41	0.90	0.74	E5* ~R1* ~R4	E; FI	R4* R10* ~R1	D; G; NE	R10* E3*E 5	FI	R4* R10* E5	G; NO
Combination 42	0.89	0.76	E5* ~R1	E; FI; G	E5* ~I3	E; FI	R4* R10* ~R1	D; G; NE	R4* R10* E5	G; NO
Combination 43	0.88	0.80	Seven (7) solution terms: too many							
Combination 44	0.89	0.82	E5* ~T3	E; G; IR	R10* ~R9	D; NE; NO	R10* E5	FI; G; NO	/	/
Combination 45	0.90	0.91	R12	D; FI; NE; NO; S	E5* ~T3	E; G; IR	/	/	/	/

Empirically contradictory cases are shown in bold.

Table AA6 *Combination 45: truth table*

R1	*R10*	*R12*	*E5*	*T3*	*Number*	*outcome*	*Raw consist.*	*PRI consist.*	*SYM consist.*
0	1	1	1	1	1 (66%)	1	0.99	0.98	0.98
0	1	1	0	0	2 (16%)	1	0.97	0.95	0.95
0	0	0	1	0	1 (91%)	1	0.96	0.93	0.93
0	1	0	1	0	1 (75%)	1	0.95	0.91	0.91
0	0	1	0	0	1 (83%)	1	0.95	0.90	0.90
1	1	1	1	1	1 (25%)	1	0.92	0.87	0.91
1	0	0	1	0	1 (50%)	1	0.84	0.75	0.75
1	1	0	0	0	1 (33%)	0	0.65	0.35	0.35

0	0	0	0	0	1 (100%)	0	0.64	0.28	0.28
1	0	0	0	1	1 (58%)	0	0.49	0.10	0.10
1	0	0	1	1	1 (41%)	0	0.41	0.07	0.07
…… (20 more combinations of conditions without empirical cases) ……									
0	0	0	0	1	0 (100%)	/	/	/	/

Theoretical assumptions: all conditions should contribute to the outcome when they are present.

Complex solution: R12*~R1*~E5*~T3 + E5*~R10*~R12*~T3 + E5*~R1*~R12*~T3 + R10*R12*E5*T3 (coverage 0.85; consistency 0.95).

Parsimonious solution: R12 + R10*~R1 + E5*~T3 + T3*~R1 +R10*T3 + R1*~R10*~T3 (coverage 0.95; consistency 0.85).

Twenty-one logical remainders exist. Thus, twenty-one combinations of conditions are not characterized by any empirical case, and problems of limited diversity are present in our data. Consequently, the solution formulas – complex, parsimonious and intermediate – are not interchangeable: indeed, in QCA, solution formulas differ based on the assumptions on logical remainders. The complex solution does not include remainder rows when minimizing the consistent rows with cases. The parsimonious solution treats remainders as 'do not care', stimulating outcome values to obtain parsimony. Conversely, the intermediate solution evaluates the plausibility of remainders in accordance with the researcher's simplifying assumptions based on theoretical or substantive empirical knowledge. Even though the most advanced methodological literature does not reach consensus on this issue (Thiem 2016; Thiem *et al*. 2015), it is generally suggested that one should consider the intermediate solution in these cases to lower the risk of drawing incorrect inferences about the automatic counterfactuals used in the parsimonious and complex solution[17] (Ragin 2008, 175; Jano 2016, 15).

[17] For the sake of transparency, as Thiem (2016) suggests, we also present both the complex solution and the parsimonious solution.

Online Supplementary Material: Appendix B – Robustness check

Table AB1 *Analysis of necessary conditions (5-year time lag). Outcome: (adjusted) variation in the % of 25-34-year-old adults attaining a university degree (1996-2015)*

	Outcome		*~Outcome*	
Condition	*Consistency*	*Coverage*	*Consistency*	*Coverage*
R1	0.53	0.61	0.78	0.54
~R1	0.60	0.82	0.44	0.36
R2	0.70	0.82	0.49	0.35
~R2	0.44	0.59	0.74	0.60
R3	0.62	0.70	0.62	0.43
~R3	0.49	0.68	0.57	0.48
R4	0.63	0.77	0.58	0.43
~R4	0.53	0.68	0.69	0.53
R5	0.32	0.71	0.44	0.60
~R5	0.82	0.71	0.79	0.41
R6	0.49	0.64	0.74	0.59
~R6	0.69	0.81	0.55	0.40
R7	0.42	0.73	0.40	0.42
~R7	0.68	0.65	0.75	0.44
R8	0.47	0.76	0.46	0.45
~R8	0.66	0.67	0.76	0.46
R9	0.49	0.82	0.38	0.39
~R9	0.64	0.63	0.83	0.50
R10	0.08	0.59	0.14	0.63
~R10	**0.95**	0.64	0.81	0.38
R11	0.49	0.61	0.66	0.50
~R11	0.59	0.74	0.48	0.37
R12	0.60	0.81	0.43	0.36
~R12	0.53	0.60	0.77	0.54
E1	0.72	0.80	0.54	0.36
~E1	0.42	0.60	0.70	0.61
E2	0.57	0.80	0.49	0.42
~E2	0.59	0.66	0.77	0.52
E3	0.42	0.76	0.41	0.45
~E3	0.70	0.66	0.79	0.45
E4	0.62	0.81	0.48	0.38
~E4	0.53	0.63	0.76	0.55
E5	0.71	0.85	0.43	0.31

~E5	0.42	0.55	0.79	0.62
E6	0.26	0.74	0.32	0.55
~E6	0.84	0.67	0.85	0.41
T1	0.41	0.73	0.35	0.38
~T1	0.65	0.62	0.74	0.44
T2	0.27	0.68	0.39	0.59
~T2	0.84	0.69	0.79	0.40
T3	0.30	0.66	0.35	0.46
~T3	0.75	0.65	0.74	0.39
I1	0.35	0.68	0.54	0.63
~I1	0.81	0.74	0.72	0.41
I2	0.50	0.84	0.32	0.32
~I2	0.60	0.59	0.85	0.51
I3	0.43	0.66	0.63	0.58
~I3	0.73	0.76	0.63	0.40

Coefficients indicating a necessity relation (>0.9) are shown in bold.
Logical No (~) refers to the absence of a condition.

Here, it was observed that the absence of R10 (more institutional autonomy) was necessary for the outcome. This evidence raises relevant questions about whether increasing universities' institutional autonomy (in terms of self-governance powers) plays a positive role by having a direct effect on improving the systemic performance in HE, as is very often underlined by the specialized literature.

Table AB2 *Combination 38 (5 year-time-lag): truth table*

R1	*R10*	*R12*	*E5*	*I1*	*Number*	*outcome*	*Raw consist.*	*PRI consist.*	*SYM consist.*
0	0	1	1	0	1 (91%)	1	0.96	0.94	0.94
0	0	1	0	0	2 (50%)	1	0.92	0.87	0.87
0	0	0	1	0	2 (66%)	1	0.90	0.85	0.85
1	0	1	1	0	2 (16%)	1	0.85	0.78	0.78
1	0	0	1	1	1 (83%)	1	0.81	0.70	0.70
0	0	0	0	1	1 (100%)	0	0.68	0.41	0.41
1	0	1	0	1	1 (75%)	0	0.51	0.17	0.17
1	0	0	0	0	2 (33%)	0	0.48	0.26	0.26
1	1	1	1	1	0 (100%)	/	/	/	/
…… (22 more combinations of conditions without empirical cases) ……									
					0 (100%)	/	/	/	/

Theoretical assumptions: all conditions should contribute to the outcome when they are present.

Complex solution: R12*~R1*~R10*~I1 + E5*~R1*~R10*~I1 + E5*~R10*R12*~I1 + R1*~R10*~R12*E5*I1 (coverage 0.75; consistency 0.89).

Parsimonious solution: R10 + E5 + ~R1*~I1 + R12*~I1 + R12*~R1 + R1*~R12*I1 (coverage 0.92; consistency 0.84).

Table AB3 *Combination 38 (5 year-time-lag): intermediate solution*

Solution terms	*Raw coverage*	*Unique coverage*	*Consistency*	*Cases covered*
E5	0.71	0.31	0.85	Eng (0.86, 0.95); Fin (0.86, 0.77); Gre (0.65, 0.76); Ire (0.65, 0.98); Nor (1, 0.91); Swe (0.65, 0.98)
R12*~R1	0.38	0.00	0.94	Den (0.65, 0.93); Fin (0.68, 0.77); Ned (0.65, 0.77)
R12*~I1	0.56	0.04	0.89	Den (0.65, 0.93); Fin (0.78, 0.77); Ned (0.65, 0.77); Nor (0.61, 0.91); Swe (0.95, 0.98)

Solution coverage (proportion of membership explained by all paths identified): 0.871314

Solution consistency ('how closely a perfect subset relation is approximated') (Ragin 2008, 44): 0.856390

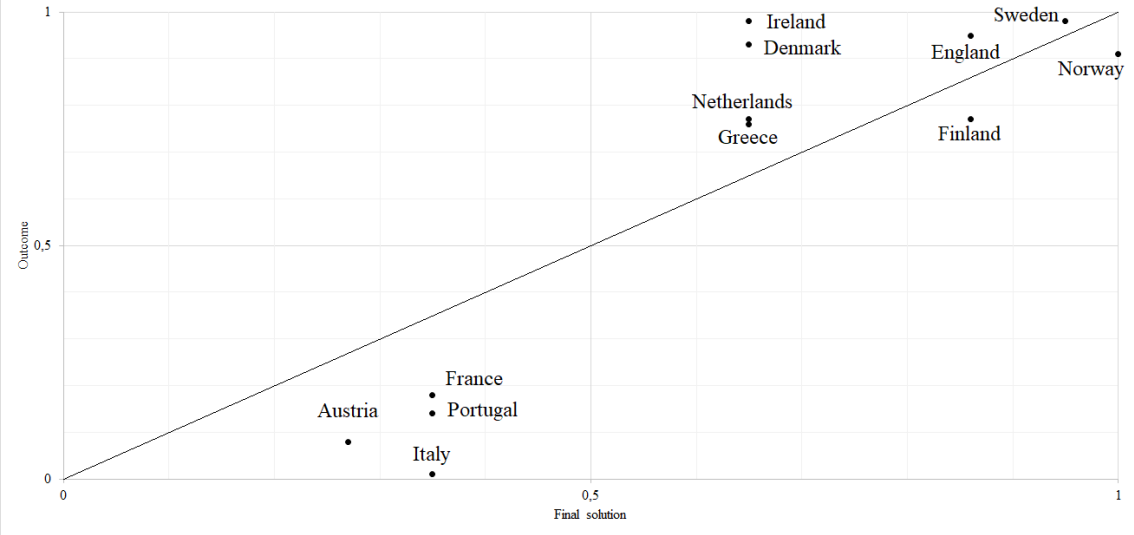
Raw coverage: proportion of memberships in the outcome explained by a single path.

Unique coverage: 'proportion of memberships in the outcome explained solely by each individual solution term' (Ragin 2008, 86).

Empirically contradictory cases are shown in bold.

This robustness check strongly demonstrated that R12 and E5 matter (by assuming an almost immediate effect of the newly adopted instruments, similar to the results of the analysis we conducted in the paper). The presence of three solution terms rather than two – as in the best combination that emerged in the paper – simply shows that the presence of R12 requires the absence of specific other conditions to work; at the same time, it emerges that E5 in this combination also includes three Nordic countries. These two differences do not undermine the combination chosen in the paper but demonstrate an observation that was clearly underlined in the paper, especially in the discussion and conclusion: that the actual condition can be the last step in a specific sequence of choices.

Figure AB1 *The 'best policy mix' (5 year-time-lag): final XY plot*



REFERENCES

Capano, Giliberto. 2011. Government Continues to Do Its Job. A Comparative Study of Governance Shifts in the Higher Education Sector. *Public Administration* 89(4):1622–1642.

Capano, Giliberto, Marino Regini, and Matteo Turri. 2016. *Changing Governance in Universities. Italian Higher Education in Comparative Perspective*. London: Palgrave MacMillan.

Flannery, Darragh, and Cathal O'Donoghue. 2011. The Life-cycle Impact of Alternative Higher Education Finance Systems in Ireland. *The Economic and Social Review* 42(3):237–270.

Huisman, Jeroen (ed) (2009) *International Perspectives on the Governance of Higher Education*. London: Routledge.

Jano, D. (2016). Compliance with EU Legislation in the Pre-accession Countries of South East Europe (2005-2011): A Fuzzy-set Qualitative Comparative Analysis. *Journal of European Integration*, 38(1): 1-22.

Joao Rosa, Maria and Alberto Amaral (eds.) 2014. *Quality Assurance in Higher Education*. London: Palgrave.

Jonkers Koen, and Thomas Zacharewicz 2016. Research Performance Based Funding Systems: A Comparative Assessment. Brussels: European Commission, Joint Research Centre. https://rio.jrc.ec.europa.eu/en/file/9514

Lane, Jan Erik and Jussi Kivisto. 2008. The role of interests, information, and incentives in higher education: A review of principal–agent theory and its potential applications to study of higher education governance. In *Higher education: Handbook of theory and research*, ed. Jan Smart, Vol. XXIII (pp. 141–179). Dordrecht: Springer.

Moe, Terry M. 1984. The new economics of organization. *American Journal of Political Science* 28(4): 739–777.

Ragin, Charles C. 2008. *Redesigning Social Inquiry: Fuzzy Sets and Beyond*. Chicago: The University of Chicago Press.

Schneider, Carsten Q., and Claudius Wagemann. 2010. Standards of Good Practice in Qualitative Comparative Analysis (QCA) and Fuzzy-Sets. *Comparative Sociology*, 9(3): 397-418.

Schwarz, Stefanie and Don Westerheijden (eds.). 2007. *Accreditation and evaluation in the European higher education area*. Dordrecht: Springer.

Shattock, Michael L., ed. 2014. *International Trends in University Governance*. London: Routledge.

Thiem, Alrik (2016). Standards of Good Practices and the Methodology of Necessary Conditions in Qualitative Comparative Analysis. *Political Analysis*, 24: 478-484.

Thiem, Alrik, Baumgartner M, Bol D (2015). Still Lost in Translation! A Correction of Three Misunderstandings Between Configurational Comparativists and Regressional Analysts. *Comparative Political Studies*, online publication first, doi: 10.1177/0010414014565892.

Van Vught, Frans (ed.) 1989. *Governmental Strategies and Innovation in Higher Education*. London: Jessica Kingsley.

Zacharewicz, Thomas, Benedetto Lepori, Emanuela Reale, Koen Jonkers; Performance-based research funding in EU Member States—a comparative assessment. 2018. *Science and Public Policy*, online first 18 June, https://doi.org/10.1093/scipol/scy041.