

AperTO - Archivio Istituzionale Open Access dell'Università di Torino

Exploring the determinants of higher education performance in Western Europe: A qualitative comparative analysis

This is the author's manuscript

Original Citation:

Availability:

This version is available <http://hdl.handle.net/2318/1694712> since 2021-03-09T12:56:19Z

Published version:

DOI:10.1111/rego.12244

Terms of use:

Open Access

Anyone can freely access the full text of works made available as "Open Access". Works made available under a Creative Commons license can be used according to the terms and conditions of said license. Use of all other works requires consent of the right holder (author or publisher) if not exempted from copyright protection by the applicable law.

(Article begins on next page)

Exploring the determinants of higher education performance in Western Europe: A qualitative comparative analysis

ABSTRACT: As a result of domestic pressure or international prescription, many national higher education systems (HESs) in Europe have undergone structural changes over the last thirty years. These changes have occurred primarily to enhance the overall performance – defined as students’ access, quality of teaching and excellence in research – of universities. As such, almost all of these countries have decided to adopt similar policy strategies to foster institutional autonomy and differentiation and greater managerial steering. However, although similar policy patterns have been replicated, performance indicators are remarkably variable across Western European countries. Thus, many of the proposed explanations may be over-simplifications of reality. This issue leads to our main research question: which factors are associated with (teaching) performance improvement in higher education (HE)? This paper focuses on university-level education to explore the possibility that this association is conjunctural in nature. In other words, when considering the performance of university systems (i.e., teaching), it is important to identify the most important combination of institutional autonomy, evaluation, internal governance and public funding. Qualitative comparative analysis (QCA) was employed to test this expectation on developments (1988-2008) in twelve HESs in Western Europe: Austria, Denmark, England, Finland, France, Greece, Ireland, Italy, the Netherlands, Norway, Portugal, and Sweden. The results suggest that giving more institutional autonomy to universities is much less important than expected, whereas severe evaluation, generous public funding and verticalized governance emerge as more relevant.

KEYWORDS: Higher education – Performance – QCA – Western Europe – Governance reforms

1. Introduction

Over the last three decades, governments have recurrently intervened in higher education (HE). Over time, significant changes have occurred in inherited national governance modes. In Europe, these governmental policies have attempted to abandon the inherited Continental governance mode, which is characterized by hierarchical coordination through state-centred policies, a lack of institutional autonomy, the powerful, all-pervasive authority of academic guilds, and faculties and schools as “confederations of chair-holders” (Clark 1983), in favour of the model used in English-speaking countries. These reforms have been characterized as “autonomistic” because universities have been granted more institutional autonomy at various levels and intensities. However, institutional autonomy does not stand alone. The other side of this phenomenon has been the changing role of governments in leading their HE systems,

particularly their university systems. Governments have drastically reduced the use of the traditional direct command and control strategy in favour of leading from a distance based on national standards, procedures for monitoring and evaluation, criteria for financial rewards and changing internal institutional governance arrangements (Lazzaretti and Tavoletti 2006; Huisman 2009; Paradeise *et al.* 2009; Enders, De Boer, and Weyer 2013). In contrast, in the Anglo-Saxon world, governments have increased their intervention and regulation despite a historical tradition of institutional autonomy for universities (El-Khawas 2005; Pick 2006; McLendon and Hearn 2009; Schuetze, Bruneau and Grosjean 2012; Jones 2012).

Governments in Continental Europe have attempted to use institutional autonomy as a fundamental reform tool, whereas the existing and deeply rooted institutional autonomy in Anglo-Saxon countries has been subject to governmental regulatory attention. In the Western world, the last three decades of HE policies have considered the issue of institutional autonomy and its use as an instrument for improving overall HES performance (and thus the performance of university-level education), specifically for “politically” perceived social needs. However, it remains unclear whether giving universities more institutional autonomy is a necessary and/or sufficient condition for ensuring better performance in HE.

Because the research on this question has produced disputable or inconclusive results (Christensen 2011; Enders, De Boer and Weyer 2013), we address this question from a different perspective and adopt a “conjunctural” logic to explore whether, how, and, ultimately, which other conditions (such as the structure of institutional governance, evaluative tools, and the amount of public funding) are associated with improvements in systemic performance. We are deeply conscious that other factors, such as the socio-economic context, the process of massification, and external and internal shocks, influence it. However, because we assume that governments have room to intervene in HE processes, it is necessary to explore how they can do their job and whether and how their job can be considered to have an effect on the outcome of interest.

We pursue this research strategy by collecting data on the evolution of institutional autonomy in university systems in 12 countries and performing a detailed analysis of the national regulations adopted between 1988 and 2008. The evolution of national provisions adopted by governments to develop evaluative tools (i.e., research and teaching evaluation, quality assurance in teaching, performance and targeted funding) and changes in the structure of institutional governance and public funding for the same time period (1988-2008) are also investigated. This dataset allows us to explore whether – by adopting a conjunctural analytical logic – convincing, or at least promising, results can be obtained. We develop our research strategy through a qualitative comparative analysis (QCA) in which the outcome builds on the comparison between the percentage of the 25-

to 34-year-old population with a “university-level” education in 1996 and the percentage of the same cohort that had either a bachelor or master’s degree in 2015. Our decision to differentiate time periods (i.e., 1988-2008 for the conditions and 1996-2015 for the outcome) depends on the assumption that legislation does not have an “immediate” impact on everyday institutional behaviour, and that the impact of governance reforms (indeed, every reform) on performance requires several years to develop.

The evidence that emerged from our QCA suggests that growing institutional autonomy is less important in improving teaching outputs in HE than has been previously suggested in the rhetoric of reforms and in academics’ beliefs. This occurs only when granting more institutional autonomy to universities is connected to other conditions.

This result is not unexpected to those who have emphasized that “more institutional autonomy” may mean a different approach to determine the behaviour of universities, but it conflicts with other empirical research and with the rhetoric of reforms as well as with the firm beliefs of academics and many scholars in HE and international stakeholders and think-tanks.

This paper is structured in the following manner. In the second section, we present an analysis of the literature on the relationship among performance and autonomy, governance, evaluation and public funding and describe the adopted theoretical framework. The third and fourth sections include the research design and descriptive statistics. Section five is devoted to the empirical analysis of our data through QCA, and the results are discussed in section six. The conclusion summarizes our preliminary results and addresses the applicability and usefulness of our dataset and a theoretical and methodological framework for future research in this field.

2. Explaining systemic performance in higher education: From single variable(s) to combination(s) of conditions?

2.1. The state of the art

Reforms in HE are often pursued to improve systemic performance. In this context, especially in the last three decades, government policies have homogeneously considered institutional autonomy an important trigger that can be activated and a fundamental factor that can be addressed from a distance. Attention to institutional autonomy has also been strongly supported as a pillar for reform strategies in supra-national and international organizations (European Commission 2006; OECD 2008; World Bank 2008). However, there is contrasting empirical evidence regarding the actual capacity of institutional autonomy to improve systemic performance.

Studies using similar research designs, such as reducing the complexity of the investigated phenomenon by focusing on single determinants of variations for specific performance indicators or constructing a single composite index for a set of indicators, have produced conflicting results. In their famous study, Aghion *et al.* (2010) compared US public research universities with those of European countries and showed that high institutional autonomy, which is associated with high systemic competition for resources (given public expenditures), is significantly correlated with drastic improvements in universities' research performance. In contrast, Braga, Checchi and Meschi (2013) investigated the impact of seven decades of reforms in education and HE on educational attainment (defined as completed years of education) in 24 European Union (EU) countries and found a negative correlation between institutional autonomy and educational attainment; that is, the more autonomous universities are rendered, the less inclusive the HES becomes.

Similar contrasting evidence has emerged from qualitative studies. For example, Richardson and Martinez (2009) examined five US states and demonstrated that those with a centralized governance system and thus less institutional autonomy performed better in terms of access and graduation rates than states with a decentralized state governance design. Their conclusion conflicts with a previous study based on a multivariate analysis of all PhD-granting universities that concluded that decentralized state governance resulted in better performance in contexts where the political structure and economic conditions are the main drivers of systemic performance (Knott and Payne 2004). Furthermore, many case studies demonstrate how institutional autonomy has been introduced or enlarged in all Continental European countries, although it has been implemented in various ways (Huisman 2009; Paradeise *et al.* 2009; Dobbins and Knill 2014; Shattock 2014). For example, Dutch universities have gained greater managerial than institutional autonomy (Enders, De Boer, Weyer 2013), whereas universities in Denmark enjoy the second-highest level of institutional autonomy after English institutions (Wright and Oberg 2008). In contrast, notwithstanding the governmental rhetoric, Italy's institutional autonomy was compressed two decades after it was introduced (Donina, Meoli and Paleari 2015). Unfortunately, these analyses did not examine the relationship between institutional autonomy and systemic performance. Therefore, the "Holy Grail" of institutional autonomy has been pursued without paying attention to its actual policy effectiveness.

The inconclusive empirical evidence on the role of institutional autonomy also characterizes other factors that may be significant drivers of the systemic performance of university systems. For example, research focusing on single determinants has shown that the amount of public funding is important for performance in terms of broader access and better attainment. This correlation shows that public subsidies activate individual

investments; thus, an individual's trade-off is between the marginal cost for additional years of education and the marginal benefits for his or her life opportunities. Governments should also be interested in investing in higher education to derive its net social benefit (Viaene and Zilcha 2013). This correlation has been tested, for example, by Winter-Ebmer and Wirz (2002), who analysed 14 European countries and found a strong correlation between public funding and larger enrolment in HE, although this relation could be reversed when tuition fees were increased. The same positive correlation between public funding and university performance was found in Horeau *et al.*'s (2013) analysis of 32 European countries. However, it is important to note that other studies have demonstrated no straightforward negative correlation between increasing private sharing of the cost of higher education and demand (Usher 2005; Johnstone 2006; Flannery and O' Doughue 2001). For example, Williams, de Rassenfosse, Jensen and Marginson (2013) employed a comparative large-N research study and revealed that those HESs that performed better on all three university missions were primarily those with high levels of public and private funding. More specifically, in contrast to Aghion *et al.* (2010), research performance is better in countries with higher governmental expenditure on research as a percentage of the gross domestic product (GDP), whereas participation rates are higher when there is higher total funding (public plus private). This study analysed 48 countries and identified interesting exceptions (i.e., countries with lower funding but high overall performance). Thus, the role of governance should be assessed because the countries with the best performance have different funding systems and systemic governance.

Performance funding, contracts, evaluation, and quality assurance are other policy tools that policy-makers consider viable for improving universities' performance in teaching and research and that can be adopted to better address institutional autonomy. Again, the empirical evidence in this area is problematic. Regarding performance funding, a large amount of research has focused on the US, where, since the end of the 1970s, many states have introduced performance funding to allocate extra resources by focusing on input goals (i.e., more access). Research (Volkwein and Talberg 2008; Robovsky 2012; Rutheford and Rabovsky 2014; Tandberg and Hillman 2014)¹ shows that the direct impact of these policy innovations has been weak, and they have often had no actual impact. When some impact is visible, it is difficult to ascertain whether it was

¹ It should be noted that all of these studies focused on the so-called "first wave of performance funding" in the US, when some states decided to allocate extra funding as a small percentage of the overall amount of public funding for HE. Subsequently, many states abandoned this policy. Starting in 2009, a second wave of performance funding policy was initiated. Some states introduced performance funding into the funding formula by focusing on output goals (i.e., the percentages of retention and graduation) For example, since 2009, Tennessee has allocated all public funds through a performance funding formula. However, there are no reliable data on the effects of this new policy (Dougherty *et al.* 2014).

caused by performance funding or other factors, such as changes in the state governance arrangement or changes in tuition policies.

In Europe, similar rationales are used to justify different funding systems (Frolich, Schmidt, and Rosa 2010). The results of contracts, for example, in increasing access or graduation rates are quite different (according to international statistics) and thus always return to the specific configuration of actual conditions (Cheps 2015). For example, governments have adopted very strict contracts in Denmark that include clear performance goals for institutions, which have positively affected access and graduation. In contrast, Italy and other countries have softer and generic contracts and have achieved contrasting results. Quality assurance systems have been developed worldwide, but they measure very different effects in relation to expected goals (Dill and Berkens 2010; Rosa and Amaral 2014). Evaluation research has been introduced in many countries, but only England and Italy significantly use their research results when allocating public funding. Furthermore, in some cases, the results of university evaluations have affected universities' social reputation and impacted their ability to secure competitive private funding, as in France (Musselin 2013).

Research on governance reforms in HE and their effectiveness indicates that some factors seem to matter at some times or that there is an intrinsic national idiosyncrasy in HE policy that makes every national case complex. Thus, it is only possible to provide ad hoc explanations that are based on thick descriptions and do not allow for reliable generalizations.

2.2. Searching for combinations of conditions

To decrease this divergence between the focus on single determinants and the specificity of single national cases, we attempt to adopt a different approach to assess whether and how the combination of the presence/absence of specific factors is associated with a specific outcome of interest, specifically, systemic performance in teaching.

We identify four conditions on which to base our analysis according to the above literature and evidence from the content of governmental policies that were chosen for reform. Empirical research has focused on increasing institutional autonomy, developing evaluation tools and generously funding HE to explain their impact on good teaching performance. Verticalized institutional governance is the fourth condition, which was chosen due to the focus of many governments in their interventions.

Institutional autonomy is defined as the university's level of substantial and procedural freedom in deciding on matters of interest (Berdhal 1990; Verhoest *et al.* 2004). On the basis of our argument regarding the "autonomistic turn" that has interested many Western European countries over the course of the last thirty years, our first

condition can thus be labelled “growing institutional autonomy”. Evaluation involves the instruments through which governments have designed different approaches to address institutional behaviour and thus constrain the institutional autonomy of universities (i.e., by abandoning the command and control system in Continental Europe and the traditional institutional self-government in Anglo-Saxon countries). Research evaluation, quality assurance in teaching, performance and target funding, and contracts are evaluative instruments that governments have adopted to make autonomous universities more accountable². Thus, this condition reflects the increase in the evaluative state through which institutional autonomy is redesigned, addressed and practically bordered (Neave 2012) and is defined as ‘severe evaluation’.

An analytical clarification is needed to justify why ‘growing institutional autonomy’ and ‘severe evaluation’ should be considered separate conditions. In fact, through evaluation tools, governments can steer university institutions. Evaluation practices, especially with regard to research and teaching, were originally introduced in England, where university institutions have enjoyed a high level of autonomy. Furthermore, the activation of evaluative tools has accompanied the concession of greater autonomy to universities in Continental Europe. From this perspective, evaluation instruments can be considered substitutes for regulative instruments and thus offer ways to address, if not to limit, the institutional autonomy of universities. For example, many accreditation procedures assess the quality of recruitment processes or impose ex-ante standards for opening new degree programmes, which means that more institutional autonomy in formal terms can clash with more restriction in real terms due to the evaluation instruments adopted (Christensen 2011; Neave 2012; Enders, De Boer and Weyer 2013). However, the logic of the autonomist policy is to oblige universities to behave like corporate organizations and not like confederations of chair-holders, as they have been in the past, at least in Continental Europe (Clark 1983; Capano 2008). In this sense, by providing more autonomy to universities, at least from a formal perspective, governments have obliged universities to perceive themselves as institutions by reversing the previous logic of the command and control strategy that was based on the principle that “you can do only what is centrally decided”. Currently, autonomous institutions are not obliged by evaluative tools to behave in a specific way but have room to choose their own strategy by taking into consideration issues arising from or constraints established by evaluation as well as their own context. Thus, the consideration of institutional autonomy and evaluation as separate conditions is analytically justified, although they

² The evaluation of research is considered because, although it does not have a direct impact on teaching, it represents a potential factor that can have a significant impact on teaching by addressing institutional behavior (e.g., some institutions could focus more on research performance and pay less attention to the quality of teaching, as shown by the first Teaching Excellence Framework in the UK).

may have some operational overlap or interdependence³, because institutional autonomy may theoretically have significant effects in aggregate terms.

As we have shown, public funding is always a pillar of policy performance, although it is not convincingly explanatory on its own. As such, “generous public funding” is our third condition.

Finally, we select “verticalized internal institutional governance” as our fourth condition because in most countries where public law regulates HE, governments have intervened to make the structure of their internal governance more hierarchical. For example, governments have imposed a shift from electing to appointing the rector/president, increased the presence of external members of the board, and/or strengthened the powers of the rectors (Shattock 2014). These interventions have occurred in all ten countries of Continental Europe: Sweden (1993), the Netherlands (1997), Austria (2002), Norway (2005), France (2007), Portugal (2007), Finland (2009), Italy (2010), and Greece (2011). These interventions into internal governance are viewed as a necessary tool for reinforcing the capacity of universities to behave as corporate actors and to positively react to pressure from the evaluative state (Bleiklie 1998; Whitley 2008; Whitley and Glaser 2014).

The focus on these four conditions allows us to explore and perhaps illuminate the elusive world of systemic performance in HE, to clarify the ambiguous evidence from the literature on its drivers, and, overall, to assess the real means that governments have at their disposal to address HE policy. The choice to focus on ‘good teaching performance’ as our outcome of interest was made for two main reasons.

First, the waves of reforms that have characterized HE policies in the last thirty years 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. For example, one of the goals of the Thatcher government was to increase the number of students admitted by UK universities, and one of the five targets of Europe 2020 was to consistently increase the number of citizens with a tertiary degree. Thus, universities indeed have different functions and missions, but “teaching” also has a

³ Even the increasing verticalized governance can be considered partially interdependent from institutional autonomy; many governments have imposed verticalization in internal institutional governance exactly because it is considered a fundamental source for behaving as a corporate actor. However, these three dimensions (autonomy, evaluation, internal governance) deserve to be conceptualized separately because they are different ways for governments to address the behaviour of the system through the behaviour of the institutions. Separating these dimensions allows us to better understand, through a QCA, whether and how they are present in association with the outcome of interest.

particular political salience, and governments have been pushing universities to increase teaching performance as much as possible.

Second, reliable comparative data on performance in teaching are relatively easier to find, whereas comparable data on research as well as on other dimensions of systemic performance are more difficult to find and are less often shared within the scientific community⁴. We are aware that the choice of this teaching indicator of performance can be criticized because a significant correlation exists between the increase in enrolment and the number of citizens with a tertiary degree. However, this correlation is only partial (as we show in *Appendix A* in the Online Supplementary Material), and there is room for other conditions to work and to drive or influence performance. These conditions should be linked to the potential of governments to steer HE systems and the content of their policies.

This operationalization, although rough, considers from a quantitative perspective the capacity of HESs to increase the stock of national human capital. From a comparative perspective, this is considered a political goal after the massification of higher education. Obviously, the question of the quality of the delivered education can be raised, but this question follows the issue of increasing the numbers.

3. Research design

3.1 Case selection and time period

This paper is based on an accurate dataset of the government regulations implemented in 12 Western European HESs (Austria, Denmark, England, Finland, France, Greece, Ireland, Italy, the Netherlands, Norway, Portugal and Sweden) between 1988 and 2008. These countries were selected with the intention of obtaining a complete population. First, we sought to cover all pre-2004 enlargement EU countries. However, we were forced to exclude four additional countries for different reasons: Luxembourg, due to its small size (one university); Belgium and Germany because of their federal structures, and Spain because of its very decentralized regionalism (Biela *et al.* 2012), 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 thus can offer sufficient differentiation in terms of policy legacy

⁴ The literature on systemic performance in HE remains limited and very divided, with a focus on single aspects of systemic performance. The adoption of educational attainment is the most common (see, for example, St. Aubyn *et al.* 2009; Agasisti 2011; Obadić and Aristovnik 2011; Braga, Checchi and Meschi 2013). Furthermore, the use of a holistic approach that considers different indicators for the different missions of HE remains undeveloped and subject to many methodological problems.

(Clark 1983; Braun and Merrien 1999; Shattock 2014) and the inherited set of policy instruments. We also included a non-EU country, Norway. In this way, all the Nordic countries, which are assumed to have adopted a welfarist approach to HE, could be considered. Thus, we were able to examine whether this common characteristic influenced the analysed outcome.

In all of the selected countries, HESs have undergone structural changes over the past two to three decades. Accordingly, we decided to begin our analysis of institutional autonomy dynamics at the end of the 1980s to encompass the major changes involving HE policy over the last decades. In fact, the paradigm shift that began in the 1980s and aimed to dismantle welfare state policies and move towards more efficient and market-oriented institutions strongly affected HESs. Some of these reforms – primarily those oriented towards granting major autonomy to universities – were implemented during the 1990s, whereas others occurred or were developed in the new millennium, because of the Bologna Process, which has been a kind of “script” adopted in different ways in different countries (Capano and Piattoni 2011). Accordingly, the time span is justified by the need to analyse a period characterized by large legislative transformations in all of the included countries. Clearly, each country has its own reform starting point in the field; some had legislation at the beginning of the 1990s, whereas others began much later. It is probably more accurate to discuss starting points because many countries embarked separately on the so-called “autonomistic turn” with different regulations/laws.

Finally, we selected the year 2008 as the final point of our data collection because we are convinced that legislation does not have an “immediate” impact on everyday institutional behaviour, and that the impact of every reform needs time to manifest. It could also be argued that policy changes in the area of HE have quicker effects and, more generally, that a time lag of 8 years can be considered arbitrary. To account for this consideration, we also conducted a robustness check with a 5-year time lag between the conditions and the outcome (see *Appendix B* in the Online Supplementary Material).

3.2 Operationalization

As explained in the literature review, we assume that differences in university systems’ (teaching) performance, *ceteris paribus*, might be associated with differences in the combinations of growing institutional autonomy given to universities, the evaluation tools imposed by governments, systems of internal governance and the amount of public funding devoted to university education. Regarding the outcome (specifically, system performance), among the various indicators of performance (e.g., scientific research, access, academic recruitment and careers), we focus on teaching, which is one of the main tasks of every university and, as we noted above, has been a recurrent political goal in the last three decades.

The most common indicator of teaching is the percentage (%) of people with a university-level degree. As such, we operationalized teaching performance starting from the percentage (%) of adults between 25 and 34 years old who have a university-level degree. Because many countries' education systems changed between the 1990s and the 2000s, university programmes differ from those that existed 20 years ago. Thus, according to the Organization for Economic Co-operation and Development (OECD) data, we chose our starting point as the percentage of the 25- to 34-year-old population with a "university-level" education in 1996 compared to the percentage of the same cohort that had either a bachelor or master's degree in 2015⁵. These data were downloaded from the OECD archive (see also OECD 1998, 2016)⁶ and are summarized in Figure 1 below.

Figure 1 *University-level (BA+MA) attainment of 25-34 y.o. adults*

However, the variation in this performance indicator does not directly represent our outcome. This is 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 starting with a higher value;
- Second, results in 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 should be rewarded more than countries without these degrees because in the former case, university institutions are subject to more competition for students (or, at least, have a smaller catchment area) and are less likely to improve their results. Consequently, we slightly modified the data 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

⁵ First, we decided to examine the percentage of adults aged 30-34 who had a tertiary degree in each country as this value was one of the targets of Europe 2020. However, those dates account for both university and short-term tertiary degrees, whereas we only coded regulations related to universities.

⁶ 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 do not exclude a multidimensional index of performance, as suggested in the literature (Enders, de Boer and Weyer 2013).

countries in the intermediate (above the mean but less than one s.d. above the mean) category 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 in 2015 between 25- to 34-year-old adults with a short-cycle tertiary degree, on the one hand, and 25- to 34-year-old adults with a tertiary degree, on the other hand. In this way, we can weigh the extent to which the opportunity to enrol in short-cycle degrees affects the HES as a whole. Again, countries more than one standard deviation 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 much they are above the mean, 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. Our manipulation of the original rough data allows a better weighting of the real systemic performance because the final indicator takes into consideration significant systemic characteristics and thus assesses the systemic performance dynamics in a more realistic way. In fact, our manipulation allows us to decrease the historical differences between countries by “cleaning” the performance from the asymmetric conditions of departure.

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

Operationalizing the conditions for the next QCA (i.e., growing institutional autonomy, severe evaluation, verticalized internal governance and generous public funding) required theoretical reflection and data collection, as explained below. First, it is important to remember that there is a time lag between conditions and the outcome; conditions are operationalized based on data from 1988 to 2008, whereas the outcome compares teaching performance in 1996 and 2015. We have already explained the rationales on which this choice is based.

To empirically detect (and, in turn, “measure”) the institutional autonomy of the universities in different countries, we constructed an Index of Institutional Autonomy ranging from 0 (i.e., no institutional autonomy) to 1 (i.e., full institutional autonomy) based on three sub-dimensions (i.e., financial autonomy, organizational autonomy, and teaching autonomy) and seven specific indicators. This index is presented in Table 2. Concrete examples of actual legislation that have been coded in each sub-dimension and indicator are provided in Appendix C (Table AC2).

Table 2 *Index of Institutional Autonomy*

We chose these three sub-dimensions and their operationalization because we believe that they reflect the essential characteristics of institutional autonomy. In contrast, a complex and articulated index (e.g., the scorecard proposed by the European University Association, which is composed of 4 dimensions and 19 indicators) could generate measurement problems that may not be directly related to national regulations (EUA 2017). The proposed operationalization captures all the possible relevant dimensions of institutional autonomy and better fits the content of national regulations.

There are several ways of constructing an index. First, the choice must be made between a theoretical and an empirical approach. In the former case, variables are combined in relation to weights that are hypothesized theoretically; in the latter case, variables are generally weighted using a principal components analysis (Bartholomew and Knott 1999). Since we believe that all sub-dimensions of institutional autonomy are equally important, we opted for the less problematic decision of attributing the same weight to each sub-dimension (i.e., financial, organizational and teaching autonomy). As a result, the Index of Institutional Autonomy represents the mean of the three sub-indices, which are a mean of the coefficients attributed to all of their specific indicators⁷.

However, we do not resort to the absolute values that this index assumes in different countries. Instead, the condition that we will calibrate later is the variation in the “amount” of institutional autonomy. Thus, the expectation is that shifting the institutional setting towards more autonomy for universities influences improvements in teaching performance more than the absolute levels of institutional autonomy. In other words, we aim to test the “common” assumption that more autonomy is a good prescription for obtaining positive results.

Additionally, for the severe evaluation condition, we focus on variations in the data rather than the absolute values. Our choice is easily justifiable because at the end of the 1980s (when our data collection begins), there was evaluation of universities’ activities by the State only in France and in England due to the establishment of contracts and to the first research exercise, respectively. In all other countries, no teaching or research assessments or other forms of evaluation were implemented to increase institutional accountability. Thus, the national regulations that changed those aspects (e.g., teaching

⁷ Each policy measure that we coded has an impact on the level of institutional autonomy in the sense that it increases/decreases institutional autonomy depending on the positive/negative coefficient that coders jointly attribute to it. Similar considerations hold true for the Index of Evaluation. For more details on the coding procedure, please see Appendix C (Table AC1).

assessments, research assessments, performance funding and contracts) have been coded to detect and measure the impact of those decisions on the potential for universities to address their own objectives and practices. This led to an Index of Evaluation that ranges from 0 (no evaluation at all) to 1 (the presence of all evaluation tools and the maximum impact of those tools on public funding criteria, recruitment and career strategies) and that takes into account all variations that occurred in the period under scrutiny regarding the policy instruments included in Table 3⁸.

Table 3 *Index of Evaluation*

In this case, as noted previously with regard to the Index of Institutional Autonomy, we preferred not to attribute different weights to different factors. Therefore, the index varies equally depending on the regulations that countries approved with respect to any of the specific indicators.

For verticalized internal governance, the indicator is more straightforward: it is equal to 1 in countries where university governance is characterized by a hierarchical arrangement. In some ways, the institutional architecture is designed to follow the principles of New Public Management, where the appointment system prevails in choosing the executive roles as well as board members. Conversely, it is equal to 0 in countries where university governance has maintained the traditional arrangements (i.e., where the rector/president is elected rather than appointed, as are the members of the board) and where the system of power is as polycentric as possible. However, we also considered the possibility of hybrid scenarios. In the first case, a value equal to 0.67 indicates that the system of governance is more similar to the verticalized ideal than to the traditional type (democratic/polycentric), whereas the opposite is true when the value is equal to 0.33. To account for the empirically verified possibility that a specific country moved from one scenario to another during our time period, this indicator represents the average of all coefficients for every year from 1988 to 2008.

Finally, for the level of public funding devoted to tertiary education, we opted to consider the percentage of public funding relative to the GDP, which is operationalized as the average amount of public funding on GDP devoted to tertiary education between 1988 and 2008⁹. Similar to the outcome data, data for public funding were gathered from OECD yearly reports.

⁸ As we did with regard to the Index of Autonomy, in Appendix C (Table AC3), we provide concrete examples of actual legislation that we coded that had an impact on the Index of Evaluation.

⁹ The percentage of public funding relative to the GDP devoted to tertiary education is not the only possible proxy to operationalize the condition of “public funding”. The annual expenditure on tertiary

3.3 Data collection, dataset construction and coding

Consistent with our theoretical framework, which focused on the dynamics of institutional autonomy in Western Europe, we collected, analysed and coded all legislation related to HE in the twelve countries under analysis from the late 1980s onwards. Every act, every decree, and every formal disposition was entered into our dataset. Hundreds of official documents and thousands of pages of national legislation were carefully scrutinized and hand-coded to identify measures that increased or decreased universities' institutional autonomy and introduced evaluation tools into HESs. Specifically, the coding procedure was developed in four steps. First, we identified a list of relevant pieces of legislation in HE national 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 those issues to one of the dimensions on which we structured our Index of Institutional Autonomy (see above) or Index of Evaluation (see before). Fourth, we attributed a positive (in the case of more institutional autonomy or evaluation) or negative (in the case of less institutional autonomy or evaluation) coefficient to those issues.

For the first two steps, the research strategy was two-fold. For Italy, similar to France and both English-speaking countries (England and Ireland), the analysis was conducted "in house". That is, the authors of this paper were responsible for the Italian, French, English and Irish pieces of legislation entered into the dataset. For the other eight countries (i.e., Austria, Denmark, Finland, Greece, the Netherlands, Norway, Portugal and Sweden), linguistic barriers rendered the selection of regulations and their direct coding impossible. Therefore, we contacted a highly reputed country expert for each case to develop a comparable list of pieces of relevant regulation and legislative provisions for HE in those countries.

The legislative prescriptions were attributed to the appropriate categories by the authors, and positive and negative coefficients were chosen by the authors together with three other scholars with expertise in HE policy. Specifically, this second step of the coding procedure was developed as follows. First, each issue for each legislative provision in each piece of regulation for each country was coded separately by each author. Second, contradictory cases, legislative issues included in different categories by

educational institutions per student as well as the percentage of tertiary education public funding relative to the total public expenditure are two equally usable measures. As a robust test, we ran three different QCAs with all potential proxies for public funding. In all cases the intermediate solution was the same (with only coefficients of consistency and coverage changing slightly), whereas complex solutions were very similar to one another. This is because indicators of public funding generally co-vary. This is also a clear signal that our empirical analysis should be considered highly reliable.

the two coders (approximately 15% of the entire sample), were jointly resolved. Third, the attribution of the coefficients was shared by all members of a panel of experts that included the two authors together with three other scholars with expertise in HE policy¹⁰.

4. Descriptive statistics

This section presents descriptive statistics to provide a general picture of how the countries under scrutiny intervened in HE between 1988 and 2008. Specifically, Table 4 indicates how, when and the pace at which each country engaged in the so-called “autonomistic turn”, developed evaluative policy tools, (eventually) changed universities’ governance and publicly founded its HES.

Table 4 Autonomy, evaluation, governance and public funding in Western Europe (1988-2008)

By examining Table 4, many conclusions can be drawn. Let us start with institutional autonomy. First, as expected, England appears to be an outlier among all other countries at the beginning of the period under scrutiny, and in more recent years, English universities have been much more autonomous than their counterparts in Continental Europe. Second, almost all countries from the late 1980s onwards experienced a clear shift towards more autonomy for their universities; the only country that contradicts this trend is Ireland. Finally, there is a clear tendency towards convergence among countries. Apart from England, as mentioned previously, and France and Greece, where universities are comparably less autonomous than in other countries, most of the HESs in Western Europe appear to be similar to one another, at least with regard to how much institutional autonomy their universities enjoy.

The reforming processes that occurred in Western European HE policies over the course of the last decades not only addressed institutional autonomy; they were also characterized by a clear tendency towards the introduction of many evaluation tools by the government. Their aim was to counterbalance the increased freedom that universities began to enjoy; in other words, evaluation can be conceptualized as the price that institutions paid to be more autonomous in choosing their objectives and developing their own strategies. Regarding this topic, as Table 4 clearly indicates, all countries introduced several evaluation measures in their legislation regarding HE, such as teaching assessments, research evaluations, and performance-based public funding. However, countries can be realistically differentiated into two sub-groups: France, the Netherlands,

¹⁰ For a clearer idea of how this complex coding procedure resulted in positive and negative coefficients attributed to all legislative issues, please see Appendix C (Table AC1).

Norway and Sweden have higher Index of Evaluation values in 2008, whereas Austria, Denmark, England¹¹, Finland, Greece, Ireland, Italy and Portugal present “relatively” lower values. With regard to the pace at which governments introduced evaluation tools, trends among countries are quite similar. However, there is a clear exception: Sweden started very early with its “evaluation shift” (i.e., at the beginning of the 1990s).

Regarding the internal governance of universities, Table 4 confirms what has been repeatedly argued in the literature: England and Ireland are characterized by high verticalization in internal governance for the time period under scrutiny, but many other countries (i.e., Sweden (1992-1993), the Netherlands (1997), Austria (2002) and Denmark (2003)) followed their example and implemented governance reforms that shifted internal power towards executive bodies. The other countries introduced less-constraining reforms while attempting to maintain the traditional consensual governance system.

With respect to public funding devoted to tertiary education, two observations arise. First, Scandinavian countries (i.e., Denmark, Finland, Norway and Sweden) are much more likely to devote high public funding to tertiary education than are the other countries in our sample. The percentage of public expenditures relative to GDP that these countries invest in their HESs is approximately 2%, whereas in Austria and the Netherlands, these values are slightly below 1.5%; in all other countries, the investment barely reaches 1.25% (less in England, Portugal and, in particular, Italy). The second observation relates to the diachronic trends in the table: all countries except Italy increased the percentage devoted to tertiary education between the late 1980s and the present.

Overall, our data indicate that between 1988 and 2008, *i*) universities in Western Europe received much more autonomy (with the exception of Irish universities), *ii*) governments have widely adopted evaluation tools to counterbalance or to address that autonomy, *iii*) many countries opted for governance reforms that produced verticalization in institutional governance, and *iv*) countries increased the amount of public funding devoted to tertiary education (with the exception of Italy).

5. QCA

We adopt a fuzzy set QCA (fsQCA)¹² to empirically explore whether and to what extent our four conditions are associated with the outcome. The first step in each fsQCA is the “calibration” of sets (both the conditions and the outcome) (Ragin 2008; Schneider

¹¹ Despite the well-known RAEs (Research Assessment Exercises) that were performed in 1986, 1989, 1992, 1996, 2001 and 2008 and the more recent REF (Research Excellence Framework) in 2014.

¹² We use the software fsQCA 3.0 (Ragin and Davey 2017).

and Wagemann 2012). In this fundamental process, which should be as transparent as possible and discussed in detail (Schneider and Wagemann 2010, 403), it is crucial to specify qualitative anchors for full membership (1), full non-membership (0) and the point of maximum ambiguity (0.5)¹³. Table 5 summarizes these decisions, which are discussed at length in Appendix C of the Online supplementary material.

Table 5 *Calibration of sets (conditions and the outcome) and fuzzy values of countries*

Once the sets are calibrated, the second step of each QCA involves analysing the necessity relations and should always be conducted *before* analysing the sufficiency conditions (Schneider and Wagemann 2010, 404); see Table 6.

Table 6 *Analysis of necessary conditions. Outcome: variation in the % of 25-34 y.o. adults attaining a university degree (1996-2015)*

Table 6 demonstrates that no condition (and its non-occurrence, which is indicated with a tilde [~]) is necessary for good teaching performance¹⁴. This evidence supports the logic of our analysis, which is to focus on combining dimensions to better understand the real effects of governmental steering in HE. Furthermore, it must be emphasized that growing institutional autonomy, consensual internal governance and low public funding emerge as necessary conditions for the *absence* of the outcome. This empirical evidence is quite striking, not only because it can be a prelude to the sufficiency analysis but also, and above all, because the decision to give “more” institutional autonomy to universities – if adopted alone, without other conditions that can countervail it – appears to be a counterproductive policy tool. This evidence counters the common view of policymakers, the rhetoric of political discourse on reforming HE, and academic deep values. The increase of institutional autonomy has been the pillar of all the reforms pursued in Continental Europe, but it seems that this emphasis was inappropriate.

Subsequent to the analysis of necessity, an empirical test of sufficiency set-relations between combinations of conditions and the outcome is conducted with a “truth table”. Specifically, the process proceeds as follows. *i)* We convert the data matrix into the abovementioned truth table. *ii)* Single truth table rows are assessed for their consistency

¹³ We use the direct method of calibration (Ragin 2008, 85): once qualitative anchors have been chosen, the QCA software applies a logarithmic function and attributes fuzzy values to the remaining cases.

¹⁴ All of the consistency thresholds are 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).

scores regarding whether they count as sufficient conditions for the result. Finally, *iii*) if they count as sufficient conditions, they are included in the “Boolean minimization process”; otherwise, they are not.

Table 7 *Truth table*

First, notably, eight logical remainders exist. Thus, eight 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. In QCA, solution formulas differ on the basis of 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 “don’t 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. In these cases – 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 to reduce the risk of drawing incorrect inferences about the automatic counterfactuals used in the parsimonious and complex solution¹⁵ (Ragin 2008, 175; Jano 2016, 15). However, given the exploratory nature of this study, in what follows, we present and discuss both the intermediate solution and the complex solution since the latter is better able to show complex conjunctions of conditions among empirical cases.

Intermediate solution: $F + G + E^* \sim A$.

Table 8a *Intermediate solution: solution terms, consistency, coverage and cases covered*

Complex solution: $A^*E^*G + A^*E^*F + E^* \sim A^* \sim G^* \sim F + G^* \sim A^* \sim E^* \sim F + F^* \sim A^* \sim E^* \sim G$.

Table 8b *Complex solution: solution terms, consistency, coverage and cases covered*

Both the consistency value (0.90) and the coverage coefficient (0.96) of the intermediate solution formula are impressive. There are no contradictory cases in the

¹⁵ For the sake of transparency, as Thiem (2016) suggests, we also present the parsimonious solution, which is as follows: $\sim A + G + F$ (consistency 0.90; coverage 0.96).

lower-right quadrant or “deviant cases for coverage” (Schneider and Rohlfing 2013, 585) in the upper-left quadrant. Two cases (i.e., the Netherlands and Sweden) are above the diagonal in the upper-right corner and are “typical cases”, whereas five cases (i.e., Denmark, England, Finland, Ireland, and Norway) are “deviant cases consistent in degree”, and Greece is placed precisely on the diagonal. Finally, the cases in the lower-left quadrant (i.e., Austria, France, Italy and Portugal) are not good examples of either the solution terms or the outcome and do not merit particular attention.

Figure 2a *Intermediate solution: XY final plot*

The intermediate solution means that in three Scandinavian countries (i.e., Denmark, Finland and Norway), good teaching performance is associated with generous public funding to tertiary education, whereas the second solution term means that in three other countries (i.e., England, Ireland and the Netherlands) a verticalized internal governance exists that goes hand in hand with good teaching performance. Moreover, Sweden appears in both solution terms. This means that the Swedish case is characterized by both generous public funding and verticalized governance, and both conditions are associated with the outcome. Finally, the Greek case belongs to the third solution term, meaning that Greece is characterized by both the absence of the “autonomistic turn” and the presence of evaluation.

If we do not base our analysis on simplifying assumptions, we find a “complex” (rather than “intermediate”) solution. In this case, both parameters of fit – coverage (0.88) and consistency (0.92) – are impressive. Once more, there are no contradictory cases in the lower-right quadrant or “deviant cases for coverage” (Schneider and Rohlfing 2013, 585) in the upper-left quadrant. Six cases (Denmark, England, Finland, Ireland, the Netherlands and Sweden) are above the diagonal in the upper-right corner and are “typical cases”, whereas Norway is a “deviant case consistent in degree”, and Greece is placed precisely on the diagonal. Finally, the cases in the lower-left quadrant (i.e., Austria, France, Italy and Portugal) are not good examples of either the solution terms or the outcome and do not merit particular attention.

Figure 2b *Complex solution: XY final plot*

The complex solution means that growing institutional autonomy is associated with good teaching performance *if and only if* it is present together with evaluation and governance (first solution term: Netherlands and Sweden) or with evaluation and public funding (second solution term: Denmark, Norway and Sweden). The other solution terms mean that – in the absence of all other conditions – teaching improvements are associated

with evaluation in Greece, verticalized internal governance in England and Ireland, and generous public funding in Finland.

6. Discussion of the findings and future research

The findings of our QCA are interesting and must be interpreted and discussed in the context of further research.

First, providing more institutional autonomy does not appear to be a necessary or sufficient condition for good teaching performance. These results contrast not only some evidence offered in the literature (Aghion *et al.* 2010), but also the political rhetoric of higher education reform as well as the guidelines consistently proposed by universities at the European level (EUA 2017). On the contrary, we confirm the conclusions of the few studies that have suggested that more institutional autonomy may have negative effects on educational attainment (Braga, Checchi and Meschi 2013).

Furthermore, as shown above in the analysis of necessary conditions, giving more autonomy to universities can even be a necessary condition for the *absence* of the outcome. This evidence is clearly demonstrated by all the combinations that compose both the intermediate and the complex solution. In fact, although the intermediate solution shows that growing institutional autonomy is not relevant in two solution terms, in the third one (Greece), its absence in the presence of severe evaluation is necessary to achieve good teaching performance. The complex solution shows that growing autonomy is present in two solution terms (covering four countries) but always together with other conditions (always severe evaluation, together with either verticalized governance or generous funding), whereas in the other three solution terms, its absence is a condition for achieving the outcome. Thus, giving more institutional autonomy to universities appears irrelevant or, according to the mainstream literature on governance shifts in higher education (Huisman 2009; Shattock 2014; Dobbins and Knill 2014), implies one element of the pursued reforms that, if present, is always accompanied and better addressed by other relevant interventions.

As far as the two solutions – intermediate and complex – are concerned, it is quite clear that the former risks oversimplifying the reality, although it indicates some potentially interesting trends. For example, it shows that generous funding is the condition that explains performance in the Nordic countries (as expected, according to the welfarist nature of their HESs), whereas the verticalization of internal governance is a successful condition not only in England and Ireland (as expected) but also in the Netherlands as well as in Sweden, the first two Continental countries where the verticalization of internal governance was introduced during the 1990s. However, it is

interesting to observe the Swedish case, which is explained by both of these solution terms. Finally, the Greek case is explained by severe evaluation in the absence of growing institutional autonomy.

However, the complex solution looks more promising with respect to the exploratory nature of this analysis. With its five different solution terms, it offers material for a much deeper reflection on how good systemic teaching performance in HE can be pursued by governments and on the role of the four conditions in this effort. It is not the case that the complex solution, with its more fine-grained analysis, shows only one “deviant case consistent in degree” (Norway) when compared to the five cases shown in the intermediate solution.

In the first two solution terms (A*E*G and A*E*F), growing institutional autonomy is present together with two other conditions. The combination A*E*G includes the Netherlands and Sweden, two countries that had the maximum increase in the analysed period in verticalized internal governance and severe evaluation. However, their final scores differ slightly on institutional autonomy, and with regard to the funding dimension, the Netherlands is just below the ambiguity point for 1988-2008, whereas Sweden is much higher (see Table 4). The two countries are also typical cases in both the intermediate and complex solutions. This confirms the characteristics of the development of HE policy in Sweden and in the Netherlands: they were the first countries in Continental Europe to start the autonomistic policy (during the 1980s) and, in the following decades, to accompany this shift with radical changes in internal institutional governance and with the progressive introduction of evaluative tools (Huisman and Hendriks 2013; Helken, Frolich and Reymert 2016).

The combination A*E*F includes Denmark, Norway and, again, Sweden. It is interesting that all three countries had a significant increase in institutional autonomy but a more significant increase in evaluation, and all three countries scored very high in funding (although it should be noted that the increase in Denmark and Norway was consistent, whereas Sweden was already a generous funder of its HES in 1988). This solution term is quite interesting with respect to some empirical trends that have been observed. On the one hand, the presence of Sweden confirms that in this country, HE policy has been developed by combining the four conditions very specifically; thus, a strong governmental commitment exists to push universities towards strategic institutional behaviour (Silaander and Haaker 2017). On the other hand, it must be noted that in Denmark and in Norway, the verticalization of internal governance was introduced in 2003 and 2005, respectively, indicating that more than five years are needed to express the potential effects of this condition.

Taken together, the first two solution terms of the complex solution lead to some observations.

First, they show that growing institutional autonomy can matter, but only if severe evaluation is present, as expected, together with other conditions. Second, Nordic countries, except for Finland (where generous funding seems to have been the main relevant driver), have been characterized in the reform of HE by the capacity to combine the different instruments at their disposal in an equilibrated way. Particularly relevant is the Swedish case, in which all four conditions work together to pursue the outcome. Furthermore, it is interesting to note that the Netherlands seems to have been capable of performing very well by combining three out of the four conditions, with a relatively lower value in funding. This could be due to the specific way autonomistic policy has been pursued by the government through a combination of negotiations with universities and a very strong regulative use of evaluation (Enders, De Boer, and Weyer 2013).

The other three solution terms are characterized by the presence of only one condition, which becomes a kind of sufficient condition, in the absence of the other three. In Greece, where a new wave of institutional reform started only in 2011, severe evaluation alone is associated with good teaching performance. This is an interesting result because the presence of Greece in this solution term is surprising given that the introduction of significant evaluative practices and the establishment of a national agency of evaluation started only in 2005, while in the previous decade, the establishment of evaluation was strongly resisted by universities, rendering it ineffective (Billiris 2007). However, there is empirical evidence that, notwithstanding the initial academic opposition, the 2005 reforms introducing evaluation have been working for more than ten years (Zmas 2015)

In England and Ireland, verticalized internal governance emerges as the sufficient condition for the outcome. This is not unexpected because both countries have a low score on the other three conditions, whereas internal governance has been verticalized since the beginning of the analysed time span. However, this result should be considered with caution. In fact, unlike other countries, institutional autonomy, according to our data, was already higher in these countries, and evaluation, especially in England, has had a significant role in governmental policies (Tapper 2007; Shattock 2012; Loxley, Seery and Walsh 2014). Thus, institutional autonomy may not be relevant for these two countries because it “already worked”. This raises the question of whether other conditions, or other combinations of conditions, could explain the outcome. However, the combinations of conditions for these two countries may be idiosyncratic and therefore related to national specificity, or they may depend on the fact that our analysis, based on the variation of the four conditions, cannot gauge the actual impact of a level of institutional autonomy that was higher than average during the considered period. Alternately, it may depend on our means of assessing the changes in evaluation or on a different intensity of impact of governmental evaluation regulations.

Finally, the fifth solution term shows that Finland has achieved good teaching performance in the presence of generous funding, although the other three conditions are absent. Notably, notwithstanding the small increase in the analysed period, Finland had some institutional autonomy in 1988, and some evaluation has been introduced, whereas governance has remained consensual. Finland started a new wave of reforms in 2009, when the verticalization of internal governance was introduced¹⁶ (Välilä 2005; Ahola *et al.* 2015).

Overall, according to the complex solution, the only slightly deviant case in consistency is Norway, which could be due to a specific, idiosyncratic interactions among the three conditions that compose its solution term.

The evidence emerging from the complex solution is relevant for the exploratory aim of this analysis.

First, as already emphasized, growing institutional autonomy matters for good teaching performance only in combination with other conditions (E, F, G), thus confirming that it must be addressed by other conditions that structure the game that universities can play. There is a need to balance the macro-analysis with a micro-analysis by focusing on how the reforms have been able to change the institutional traditions and historically embedded practices of individual universities (Maassen, Gornitzka, and Fumasoli 2017).

Furthermore, our exploratory analysis confirms that the autonomistic policy is more characterized by changes in the methods used to address HESs than by the provision of autonomy to institutions. Enders, De Boer, and Weyer (2013) analysed the development of HE policy in the Netherlands and defined this phenomenon as “regulatory autonomy”. Moreover, the literature that emphasizes the shift of HE governance towards steering from a distance highlights that this new governmental strategy does not indicate a retreat from governing HE but is simply another method of influence (Paradise *et al.* 2009; Huisman 2009; Shattock 2014). Thus, the relationship between institutional autonomy and the other conditions (not only evaluation) appears to be asymmetric. Steering from a distance is a method by which HE institutions, which are more or less autonomous, are directed to regulate their behaviour based on other kinds of tools that oblige them to be accountable for expected goals.

Second, evaluation appears to be a prevailing condition when it is introduced in a consistent (“severe”) way. As emphasized, in the other two solution terms, its absence can be justified by the fact that the impact of evaluative regulations may depend on the national implementation process or by other ways that governments steer the systems. In

¹⁶ Interestingly, in 2016-2018, public funding to the Finnish university system has undergone substantial cuts (approximately 10%).

fact, what is striking in the last two solution terms of the complex solution (which include England, Ireland and Finland) is that there has apparently been no governmental steering (usually enforced through evaluation rules, at least in funding). Further empirical analysis should be pursued. Overall, severe evaluation emerges as a different, sophisticated way to achieve effects that cannot be achieved through strong, traditional regulation.

Third, the fact that different combinations work for good teaching performance may be related to three orders of factors:

- the fact that national policy styles matter;
- the way policies in higher education have been designed in terms of the combination of policy instruments adopted;
- the timing of governmental interventions and reforms.

First, analysing the characteristics of the political-administrative systems as well as the national or sectoral policy style would be useful to better appreciate the real way that these four conditions work (Bleiklie and Michelsen 2013). Second, the design of the adopted policy mixes is relevant. Thus, analysing how the national “reform packages” have been designed over time is important (Howlett 2011; Peters *et al.* 2018). Reviewing the regulative and technical details by which evaluation and the verticalization of internal governance have been introduced as well as the different possible grades of institutional autonomy could provide a more solid understanding of how the four analysed conditions interact with each other. The specific combination and context of the adopted policy tools may be very important for motivating HE institutions to achieve good teaching performance. Third, it is important to reconstruct the historical sequence through which the different conditions have been introduced or changed and to conduct a diachronic analysis of these interactions. For example, it would be useful to determine whether there is coherence and consistency after provisions are implemented to provide more institutional autonomy and specific evaluative tools are established, or if coherence exists between a particular level of public funding and the eventual adoption of institutional autonomy measures or changes in the evaluation tools (Howlett and Rayner 2007).

In sum, this area of research highlights the possibility that some specific conditions or sets of conditions should be institutionalized over time to justify severe evaluation as a sufficient condition for improving systemic performance. In-depth case studies could illuminate the dynamics and the diachronic interactions.

7. Conclusion

This paper explored the determinants of good teaching performance in HE in 12 European countries by focusing on the national regulations adopted in the last decades to

address institutional autonomy, funding, the internal governance of universities, and evaluation. Through a fuzzy set QCA, we found that providing universities with more institutional autonomy is less associated with good teaching performance than expected, whereas generous public funding, verticalized internal governance and severe evaluation are more relevant. This unexpected result dramatically complicates prior empirical research as well as the belief of stakeholders and policymakers that the decision to increase universities' institutional autonomy is an important trigger for good teaching performance. Furthermore, the results of the analysis call for a fine-grained analysis of the content of reform packages and their interactions over time.

The research design that we have developed can shed light on the ambiguities and partial inconclusiveness of previous research on the causal paths to good performance in HE. The in-house dataset allows us to consider all decisions relating to HE from 1988 to 2008. Thus, the dynamics and content of the waves of reforms pursued by governments over time have been reconstructed. We plan to increase the number of countries and to develop an in-depth analysis by focusing on the details of the adopted provisions through their interpretation as policy instruments. In this way, we will be able to address the surprising finding of the low importance of growing institutional autonomy. Unpacking national regulations and reforms for policy instruments and focusing on those combinations through which institutional autonomy is granted, evaluation is operationalized, and funding is delivered could facilitate a better understanding of the complex path towards good teaching performance in HE.

Overall, our exploration shows that the model of steering at a distance has been translated in different ways at the national level. Thus, it seems that there may be very important differences in its adoption that make it conceptually useless or misleading. Overall, giving more institutional autonomy to universities is not the panacea that many scholars and policymakers believe it to be.

References

- Agasisti T (2011). Performances and spending efficiency in higher education: A European comparison through non-parametric approaches. *Education Economics*, 19(2), 199-224.
- Aghion P, Dewatripont M, Hoxby C, Mas-Colell A, Sapir A (2010) The Governance and Performance of Universities: Evidence from Europe and the US. *Economic Policy* 25, 7–59.
- Ahola, S, Hedmo T, Thomsen JP, Vabø A. (2014) *Organisational fractures of higher education: Denmark, Finland, Norway and Sweden*. Oslo: NIFU.

- Bartholomew DJ, Knott M. (1999) *Latent Variable Models and Factor Analysis*. Arnold, London.
- Biela, J., Hennl, A., Kaiser, A. (2012) Combining federalism and decentralization: comparative case studies on regional development policies in Switzerland, Austria, Denmark, and Ireland. *Comparative Political Studies*, 45(4), 447-476.
- Bleiklie I (1998) Justifying the Evaluative State: New Public Management Ideals in Higher Education. *Journal of Public Affairs Education* 4, 87–100.
- Bleiklie, I, Michelsen, S. (2013) Comparing HE policies in Europe. *Higher Education*, 65(1), 113-133.
- Berdahl R (1990) Academic-freedom, Autonomy and Accountability in British Universities. *Studies in Higher Education*, 15, 169–180.
- Billiris H (2007) The National System of Higher Education in Greece: Waiting for a Systematic Quality Assurance System. In: Schwarz S, Westerheijden D (eds.) *Accreditation and evaluation in the European higher education area*, pp. 197-206 Dordrecht: Springer.
- Braga M, Checchi D, Meschi E (2013) Educational Policies in a Long-Run Perspective. *Economic Policy*, 28, 45–100.
- Braun D, Merrien FX (eds) (1999) *Towards a New Model of Governance for Universities?* Jessica Kingsley Publishers, London.
- CHEPS (2015), *Performance-Based Funding and Performance Agreements in Fourteen Higher Education Systems*. Cheps, Twente, NL.
- Clark B (1983) *The Higher Education System. Academic Organization in Cross National Perspective*. University of California Press, Berkeley.
- Christensen T (2011) University governance reforms: potential problems of more autonomy? *Higher Education* 62, 503-517.
- Dill DD, Beerkens M (eds) (2010) *Public Policies for Academic Quality: Analyses of Innovative Policy Instruments*. Springer, Dordrecht.
- Dobbins M, Knill C (2014) *Higher Education Governance and Policy Change in Western Europe*. Palgrave Macmillan, Basingstoke.
- Donina D, Meoli M, Paleari S (2015) Higher Education Reform in Italy: Tightening Regulation Instead of Steering at a Distance, Higher Education Policy. *Higher Education Policy* 28, 215–234.
- Dougherty, K.J. et al. (2014) Performance funding for higher education: forms, origins, impacts, and futures, *The Annals of The American Academy of Political and Social Science* 655, 163-184.

- El-Khawas E (2005) The Push for Accountability: Policy Influences and Actors in American Higher Education. In: Gornitzka A, Kogan M, Amaral A (eds), *Reform and Change in Higher Education*, pp. 287–303. Springer, Dordrecht.
- Enders, J., De Boer, H., Weyer, E. (2013) Regulatory Autonomy and Performance: The Reform of Higher Education Revisited. *Higher Education*, 65, 5–23.
- European Commission (2006) Delivering on the Modernization Agenda for Universities: Education, Research Innovation (COMM 208), Brussels, BE, 10 May.
- European University Association (EUA) (2017), *University Autonomy in Europe III*, Brussels: EUA
- Flannery D, O’Donoghue C (2011) The Life-cycle Impact of Alternative Higher Education Finance Systems in Ireland. *The Economic and Social Review*, 42, 237–270.
- Frolich N, Schmidt E, Rosa M (2010) Funding Systems for Higher Education and Their Impacts on Institutional Strategies and Academia: A Comparative Perspective. *International Journal of Educational Management*, 24, 7–21.
- Helken M, Frolich N and Reymert I (2016) *Steering Approaches in Higher Education*. Oslo: NIFU.
- Horeau C, Ritzen J, Marconi, G (2013) Higher Education and Economic Innovation, A Comparison of European Countries. *IZA Journal of European Labor Studies*, 2, 2–24.
- Howlett M (2011) *Designing Public Policies: Principles and Instruments*. London: Routledge.
- Howlett M, Rayner J (2007) Design Principles for Policy Mixes: Cohesion and Coherence in ‘New Governance Arrangements’. *Policy and Society*, 26:4, 1-18
- Huisman J (eds) (2009), *International Perspectives on the Governance of Higher Education*. Routledge, London.
- Huisman J, Hendriks F (2013) The Netherlands. In *Handbook of comparative higher education law*, edited by C, Russo, 217–238, New York: Rowman and Littlefield.
- 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.
- Johnstone DB (2006) *Financing Higher Education: Cost-Sharing in International Perspective*. Sense Publishers, Rotterdam.
- Jones GA (ed) (2012), *Higher Education in Canada: Different Systems, Different Perspectives*. Routledge, London.

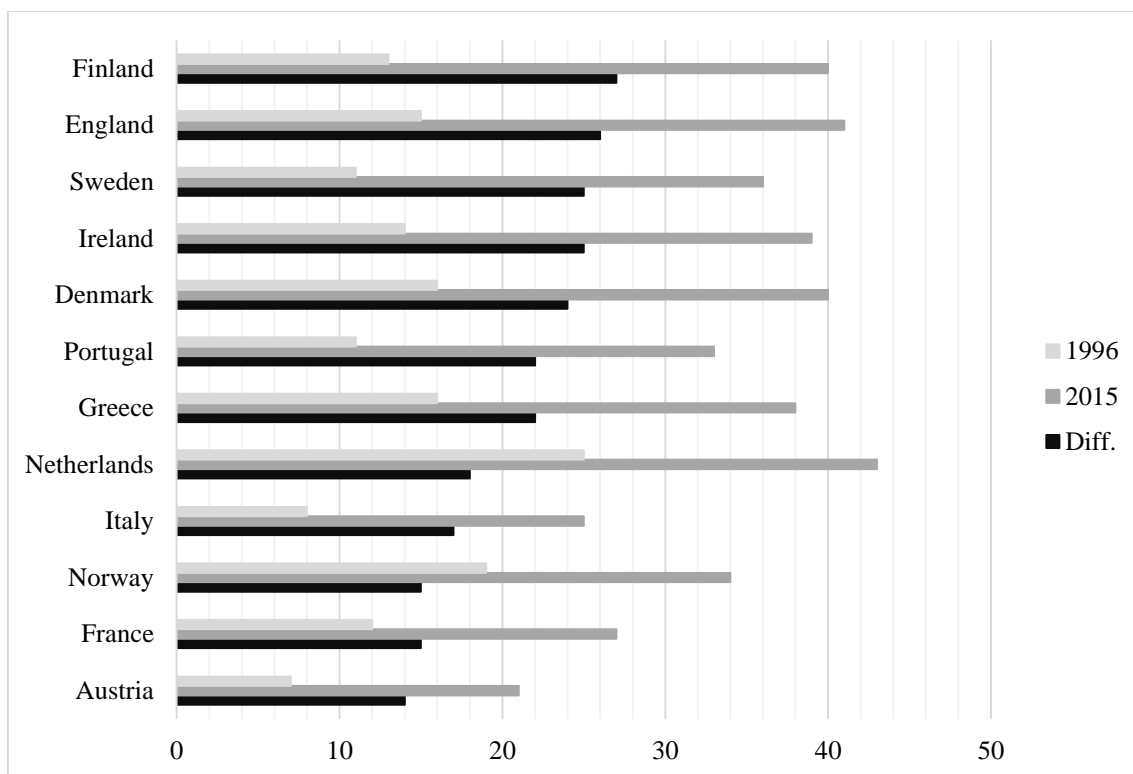
- Knott A, Payne A (2004) The Impact of State Governance Structures on Management and Performance of Public Organizations: A Study of Higher Education Institutions. *Journal of Policy Analysis and Management*, 23, 13–30.
- Lazzaretti L, Tavoletti E (2006) Governance Shifts in Higher Education: A Cross National Comparison. *European Educational Research Journal*, 5, 18–37.
- Loxley A, Seery A, Walsh J (eds.) (2014) *Higher Education in Ireland*. Dordrecht: Springer.
- Maassen P, Gornitzka Å, Fumasoli T (2017) University Reform and Institutional Autonomy: A Framework for Analysing the Living Autonomy. *Higher Education Quarterly*, 71, 239–250.
- McLendon MK, Hearn JC (2009). Viewing Recent U.S. Governance Reform Whole: “Decentralization” in a Distinctive Context. In: Huisman J. (ed) *International Perspectives on the Governance of Higher Education*, pp. 161–181. Routledge, New York.
- Musselin C (2013) How Peer-Review Simultaneously Empowers the Academic Profession and University Managers: Evolution of the Relationships Between the State, The Universities and the Professoriate. *Research Policy* 42, 1165–1173.
- Neave G (2012) *The Evaluative State, Institutional Autonomy and Re-engineering Higher Education in Western Europe*. Palgrave, London.
- Obadić A., Aristovnik A. (2011) Relative efficiency of higher education in Croatia and Slovenia: an international comparison. *Amfiteatru Economic Journal*, 13(30), 362-376.
- OECD (2016) *Education at a glance. 2016*. Oecd, Paris.
- OECD (2008) *Tertiary Education for the Knowledge Society: VOLUME 1: Special Features: Governance, Funding, Quality*. Oecd, Paris.
- OECD (1998) *Education at a glance. 1998*. Oecd, Paris.
- Paradeise C, Reale E, Bleiklie I, Ferlie E (eds) (2009) *University Governance. Western European Comparative Perspectives*. Dordrecht, Springer.
- Peters GB, Capano G, Mukherjee I, Chou MH, , Howlett M, Ravinet P (2018) *Designing for Policy Effectiveness*. Cambridge: Cambridge University Press.
- Pick D (2006) The Re-Framing of Australian Higher Education. *Higher Education Quarterly* 60, 229–241.
- Rabovsky T (2012) Accountability in Higher Education: Exploring Impacts on State Budgets and Institutional Spending Patterns. *Journal of Public Administration Research and Theory* 22, 675–700.

- Ragin CC (2008) *Redesigning Social Inquiry: Fuzzy Sets and Beyond*. The University of Chicago Press, Chicago.
- Ragin CC, Davey S (2017) *fs/QCA [Computer Programme]*, Version 3.0. Irvine, CA: University of California.
- Richardson RC, Martinez MC (2009) *Policy and Performance in American Higher Education: An Examination of Cases Across State Systems*. The John Hopkins University Press, Baltimore.
- Rosa G, Amaral G (eds) (2014) *Quality Assurance in Higher Education*. Palgrave Macmillan, Basingstoke, UK.
- Rutherford A, Rabovski T (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: 185–209.
- Schneider CQ, Wagemann C (2012) *Set-Theoretic Methods for the Social Sciences*, Cambridge University Press, Cambridge.
- Schneider C.Q. & I. Rohlfing (2013) Combining QCA and Process Tracing in Set-Theoretic Multi-Method Research. *Sociological Methods & Research* 42(4): 559-597.
- Schuetze HG, Bruneau W, Grosjean G (eds) (2012) *University Governance and Reform – Policy, Fads, and Experience in International Perspective*. Palgrave Macmillan, New York.
- Shattock ML (2012). *Making Policy in British Higher Education 1945–2011*. Maidenhead: McGraw-Hill and Open University Press.
- Shattock ML (eds) (2014) *International Trends in University Governance*. Routledge, London.
- St. Aubyn MS, Garcia F., Pais J (2009). *Study on the efficiency and effectiveness of public spending on tertiary education* (No. 390). Directorate General Economic and Financial Affairs (DG ECFIN), European Commission.
- Tandberg D, Hillman NH (2014) State Higher Education Performance Funding: Data, Outcomes, and Policy Implications. *Journal of Education Finance* 39: 222–243.
- Tapper T. (2007) *The Governance of British Higher Education*. Springer: Dordrecht.
- Thiem A (2016) Standards of Good Practices and the Methodology of Necessary Conditions in Qualitative Comparative Analysis. *Political Analysis* 24, 478-484.
- Thiem A, Baumgartner M, Bol D (2015) Still Lost in Translation! A Correction of Three Misunderstandings Between Configurational Comparativists and Regression Analysts. *Comparative Political Studies*, online publication first, doi: 10.1177/0010414014565892.

- Usher A (2005) *A Little Knowledge Is a Dangerous Thing: How Perceptions of Costs and Benefits Affect Access to Education*. Education Policy Institute, Toronto, CA.
- Välilä J (2005). Social dynamics of higher education reforms. The case of Finland. In Å. Gornitzka, M. Kogan, & A. Amaral (Eds.), *Reform and change in higher education* (pp. 245-267). Dodrecht: Springer.
- Verhoest K, Peters BG, Bouckaert G, Verschuere B (2004) The Study of Organisational Autonomy: A Conceptual Review. *Public Administration and Development* 24, 101–118.
- Viaene JMA, Zilcha I (2013) Public Funding of Higher Education. *Journal of Public Economics* 108, 78–89.
- Vis B (2012) The Comparative Advantages of fsQCA and Regression Analysis for Moderately Large-N Analysis. *Sociological Methods & Research*, 41, 168–198.
- Volkwein J, Tandberg D (2008) Measuring Up: Examining the Connections Among State Structural Characteristics, Regulatory Practices, and Performance, *Research in Higher Education* 49, 180–197.
- Wagemann C, Schneider CQ (2010) Qualitative Comparative Analysis (QCA) and Fuzzy-Sets: Agenda for a Research Approach and a Data Analysis Technique. *Comparative Sociology* 9: 376–396.
- Whitley R (2008) Universities as Strategic Actors: Limitations and Variations. In: L. Engwall L, Weaire D (eds), *The University in the Market*, pp. 22–37. Portland Press, London, England.
- Whitley R, Gläser J (eds) (2014) *Organizational Transformation and Scientific Change: The Impact of Institutional Restructuring on Universities and Intellectual Innovation*. Emerald Group Publishing, London.
- Williams R, de Rassenfosse G, Jensen P, Marginson S (2013) The Determinants of Quality National Higher Education Systems. *Journal of Higher Education Policy and Management* 35, 599–611.
- Winter-Ebmer R, Wirz A (2002) Public Funding and Enrolment into Higher Education in Europe. *IZA Discussion Papers 503*, Institute for the Study of Labor (IZA).
- Wright S, Orberg J (2008) Autonomy and Control: Danish University Reform in the Context of Modern Government. *Learning and Teaching* 1, 27–57.
- World Bank (2008) *Global Trends in University Governance*. World Bank, Washington.
- Zmas A (2015) Financial crisis and higher education policies in Greece: between intra- and supranational pressures. *Higher Education* 69(3), 495–508.

Tables and Figures

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*



Source: our elaboration of OECD data.

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

Step 1				
<i>Country</i>	<i>University attainment '96</i>	<i>University attainment '15</i>	<i>Difference 2015-1996</i>	<i>Increase 1</i>
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
<i>Mean</i>	<i>13.92</i>		<i>+20.83</i>	
<i>St. deviation</i>	<i>4.68</i>		<i>4.56</i>	
Step 2				
<i>Country</i>	<i>Short-cycle '15</i>	<i>Tertiary attainment '15</i>	<i>Short / Tertiary (%)</i>	<i>Increase 2</i>
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
<i>Mean</i>			<i>15.72</i>	
<i>St. deviation</i>			<i>14.94</i>	
<i>Final outcome</i>				

<i>Country</i>	<i>University attainment 2015-1996</i>	<i>Increase 1</i>	<i>Increase 2</i>	<i>Final outcome</i>
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
<i>Mean</i>	+20.83			+26.26
<i>St. deviation</i>	4.56			4.46

Source: our elaboration of OECD data.

Table 2 *Index of Institutional Autonomy*

<i>Sub-dimensions and indicators</i>	<i>Description</i>	<i>Variation</i>
<i>Financial (F)</i>	<i>Mean of F1 and F2</i>	<i>0 – 1</i>
F1	Free to charge fees to students	Not free at all 0 – 1 Totally free
F2	Free to decide on internal allocation of public funds (including receiving multi-year block grant) and keep surpluses (create endowment)	Not free at all 0 – 1 Totally free
<i>Organizational (O)</i>	<i>Mean of O1 and O2</i>	<i>0 – 1</i>
O1	Free to set internal academic organization (faculties, departments, schools)	Not free at all 0 – 1 Totally free
O2	Free to determine procedures for academic recruitment/promotions/dismissal	Not free at all 0 – 1 Totally free
<i>Teaching (T)</i>	<i>Mean of T1, T2 and T3</i>	<i>0 – 1</i>
T1	Free to select students for bachelor and master's programmes	Not free at all 0 – 1 Totally free
T2	Free to decide on the content of bachelor and master's programmes	Not free at all 0 – 1 Totally free
T3	Free to start new bachelor and master's programmes	Not free at all 0 – 1 Totally free
<i>Institutional</i>	<i>Mean of F, O and T</i>	<i>0 – 1</i>

<i>Autonomy</i>	
-----------------	--

Table 3 *Index of Evaluation*

<i>Evaluation tools</i>	<i>Examples</i>
Accreditation	Establishment of the ‘Danish Qualifications Framework for Higher Education’ (Denmark 2003).
Establishment/reform of an agency for assessment, evaluation and accreditation	Establishment of A3ES (Portugal 2007).
Teaching assessment	Establishment of the ‘ <i>Nuclei di valutazione interna degli atenei</i> ’ (Italy 1999).
Research assessment	Launch of RAE (Research Assessment Exercise) (England 1986-1989-1992-1996-2001-2008).
Contracts between universities and the State	Institution of the instrument of ‘ <i>contractualisation</i> ’, which prescribed that 5%-10% of university budgets be funded by four-year contracts with the State (France 1988).
Performance-/quality-based teachers’ salary	Salaries are based on the salary system for Finnish universities (YPJ), according to which a requirement level is defined for each position and the level of personal performance for each employee (Finland 2006).
Performance-based institutional funding	Formula funding: a percentage of public funding is allocated on the basis of research assessment (Norway 2006).

Index variation: 0 – 1, where ‘0’ means that none of the above-mentioned evaluation tools is present in the country under scrutiny, while ‘1’ means that all the above-mentioned evaluation tools are present in the country under scrutiny.

Table 4 *Autonomy, evaluation, governance and public funding in Western Europe (1988-2008)*

<i>Country</i>	<i>Autonomy</i>		<i>Evaluation</i>		<i>Governance</i>			<i>Funding</i>		
	<i>1988</i>	<i>2008</i>	<i>1988</i>	<i>2008</i>	<i>1988</i>	<i>2008</i>	<i>Mean 88-08</i>	<i>1990</i>	<i>2008</i>	<i>Mean 88-08</i>
Austria	0.15	0.47	0	0.35	0	1	0.33	1.00	1.44	1.32
Denmark	0.32	0.58	0	0.45	0	1	0.29	1.30	2.12	2.19
England	0.79	0.86	0.05	0.25	1	1	1	0.70	0.78	0.86
Finland	0.27	0.38	0	0.30	0	0	0.00	1.20	1.81	1.84
France	0.12	0.29	0.10	0.60	0	0.33	0.03	0.80	1.21	1.06
Greece	0.14	0.22	0	0.40	0	0	0.00	0.80	1.43	1.05
Ireland	0.62	0.52	0	0.30	1	1	1	0.90	1.26	1.09
Italy	0.18	0.47	0	0.20	0	0	0.00	1.00	0.81	0.79
Netherlands	0.32	0.48	0	0.60	0	1	0.57	1.10	1.41	1.32
Norway	0.24	0.50	0	0.60	0	0.67	0.13	1.10	2.01	1.91
Portugal	0.42	0.57	0	0.40	0	0.67	0.06	1.00	0.91	0.95
Sweden	0.17	0.43	0	0.60	0	1	0.81	1.60	1.72	1.81
<i>Mean</i>	<i>0.31</i>	<i>0.48</i>	<i>0.01</i>	<i>0.42</i>	<i>0.17</i>	<i>0.64</i>	<i>0.35</i>	<i>1.04</i>	<i>1.41</i>	<i>1.35</i>

Source: our elaboration of original data.

Notes: With respect to public funding, data from Greece, the Netherlands, Portugal and Sweden in 1990 were missing in OECD reports. As such, they have been replaced by 1995 data. With regard to Greece, the 2008 data were also missing and have been replaced by 2005 data.

Table 5 Calibration of sets (conditions and the outcome) and fuzzy values of countries

	<i>Autonomy</i>	<i>Evaluation</i>	<i>Governance</i>	<i>Funding</i>	<i>Good Teaching Perform.</i>
Austria	0.99	0.59	0.33	0.25	0.08
Denmark	0.96	0.89	0.29	0.98	0.93
England	0.20	0.05	1	0.02	0.95
Finland	0.33	0.33	0	0.88	0.77
France	0.73	0.95	0.03	0.07	0.18
Greece	0.23	0.77	0	0.06	0.76
Ireland	0	0.33	1	0.08	0.98
Italy	0.98	0.05	0	0.01	0.01
Netherlands	0.68	0.99	0.57	0.25	0.77
Norway	0.96	0.99	0.13	0.92	0.91
Portugal	0.62	0.77	0.06	0.04	0.14
Sweden	0.96	0.99	0.81	0.87	0.98
<i>Thresholds</i>					
Full-member (1)	+0.25	+0.50	1	2.0%	+30 p.p.
Maximum indifference (0.5)	+0.13	+0.33	0.5	1.5%	+25 p.p.
Full non-member (0)	0	+0.20	0	1.0%	+20 p.p.

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

Table 6 Analysis of necessary conditions. Outcome (good teaching performance): variation in the (adjusted) % of 25-34 y.o. adults attaining a university degree (1996-2015)

<i>Condition</i>	<i>Outcome</i>		<i>~Outcome</i>	
	<i>Consistency</i>	<i>Coverage</i>	<i>Consistency</i>	<i>Coverage</i>
Institutional autonomy	0.62	0.61	0.92	0.55
~Institutional autonomy	0.53	0.92	0.34	0.36
Evaluation	0.74	0.71	0.70	0.41
~Evaluation	0.38	0.68	0.50	0.54
Governance	0.52	0.92	0.20	0.21
~Governance	0.55	0.53	0.93	0.54
Public funding	0.55	0.92	0.24	0.25
~Public funding	0.55	0.55	0.93	0.55

Table 7 *Truth table*

A	E	G	F	Number	Good Teaching Perform.	Raw consist.	PRI consist.	SYM consist.
0	0	1	0	2 (58%)	1	1	1	1
0	0	0	1	1 (100%)	1	1	1	1
1	1	0	1	2 (41%)	1	0.94	0.91	0.91
1	1	1	1	1 (66%)	1	0.91	0.86	0.86
0	1	0	0	1 (91%)	1	0.83	0.65	0.65
1	1	1	0	1 (75%)	1	0.82	0.68	0.68
1	1	0	0	3 (25%)	0	0.48	0.16	0.16
1	0	0	0	1 (83%)	0	0.33	0	0
1	0	1	1	0 (100%)	-	/	/	/
1	0	1	0	0 (100%)	-	/	/	/
1	0	0	1	0 (100%)	-	/	/	/
0	1	1	1	0 (100%)	-	/	/	/
0	1	1	0	0 (100%)	-	/	/	/
0	1	0	1	0 (100%)	-	/	/	/
0	0	1	1	0 (100%)	-	/	/	/
0	0	0	0	0 (100%)	-	/	/	/

Theoretical assumptions (directional expectations) for intermediate solution: all conditions should contribute to the outcome when they are present.

Table 8a *Intermediate solution: solution terms, consistency, coverage and cases covered*

<i>Solution terms</i>	<i>Raw coverage</i>	<i>Unique coverage</i>	<i>Consistency</i>	<i>Cases covered</i>
F	0.55	0.26	0.92	Den (0.98, 0.93); Fin (0.88, 0.77); Nor (0.92, 0.91); Swe (0.87, 0.98)
G	0.52	0.24	0.92	Eng (1, 0.95); Ire (1, 0.98); Ned (0.57; 0.77); Swe (0.81, 0.98)
E*~A	0.30	0.12	0.87	Gre (0.76, 0.76)

- Intermediate solution coverage (proportion of membership explained by all paths identified): 0.958445
- Intermediate solution consistency ('how closely a perfect subset relation is approximated') (Ragin 2008, 44): 0.895990
- 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

Table 8b *Complex solution: solution terms, consistency, coverage and cases covered*

<i>Solution terms</i>	<i>Raw coverage</i>	<i>Unique coverage</i>	<i>Consistency</i>	<i>Cases covered</i>
A*E*G	0.27	0.03	0.89	Ned (0.57, 0.77); Swe (0.81, 0.98)
A*E*F	0.47	0.19	0.95	Den (0.89, 0.93); Nor (0.92, 0.91); Swe (0.87, 0.98)
E*~A*~G*~F	0.22	0.12	0.83	Gre (0.76, 0.76)
G*~A*~E*~F	0.22	0.19	1	Eng (0.80, 0.95); Ire (0.67, 0.98)
F*~A*~E*~G	0.12	0.05	1	Fin (0.67, 0.77)

- Complex solution coverage (proportion of membership explained by all paths identified): 0.878016
- Complex solution consistency ('how closely a perfect subset relation is approximated') (Ragin 2008, 44): 0.916084
- 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

Figure 2a *Intermediate solution: XY final plot*

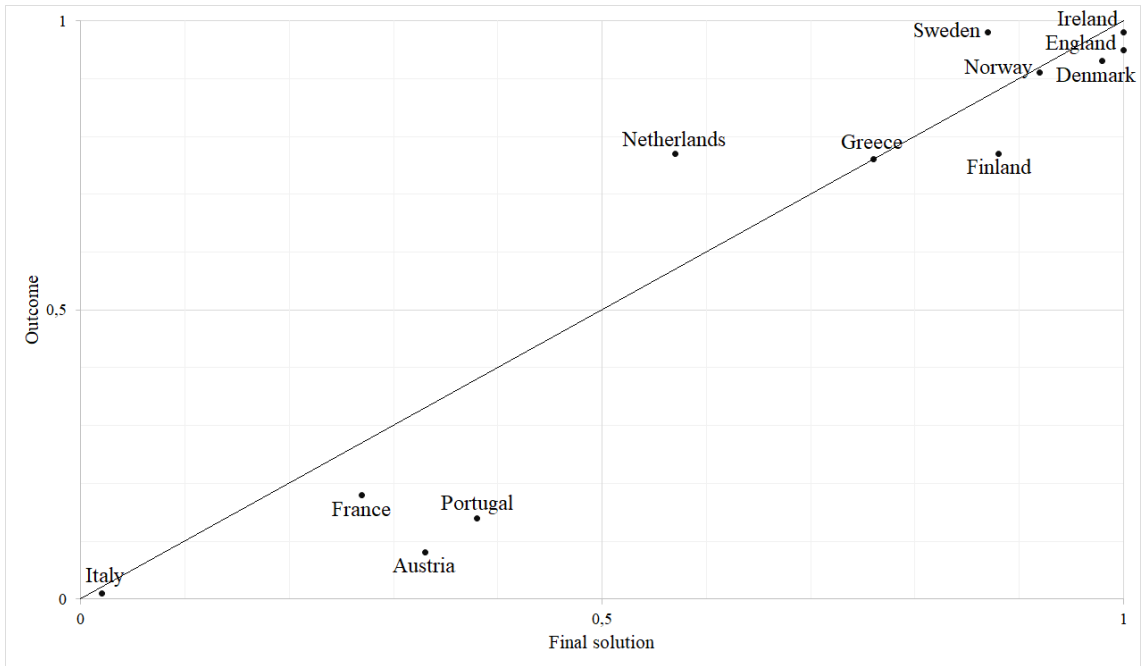
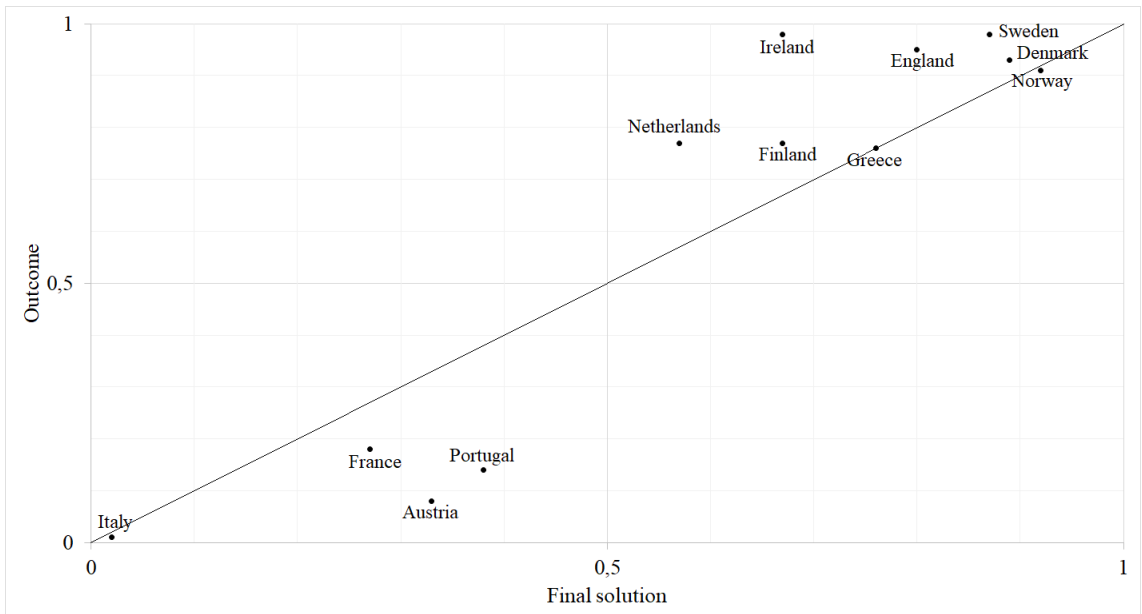


Figure 2b *Complex solution: XY final plot*



Online supplementary material

Appendix A

Correlation between enrollments and teaching performance

Table AA1 *Correlation between enrolment in tertiary education at T_0 and education attainment at T_1 (6-year-time-lag)*

<i>country_year</i>	<i>Enrollment in tertiary education per 100,000 inhabitants</i>	<i>25- to 34-year-old adults with university (BA+MA) degree</i>	<i>country_year</i>
Austria_2009	3685.4	21	Austria_2015
Austria_2008	3418.3	20	Austria_2014
Austria_2007	3143.8	m	Austria_2013
Austria_2006	3061.2	18	Austria_2012
Austria_2005	2968.0	16	Austria_2011
Austria_2004	2910.0	15	Austria_2010
Austria_2003	2817.8	15	Austria_2009
Austria_2002	2757.2	13	Austria_2008
Austria_2001	3276.1	13	Austria_2007
Austria_2000	3244.7	13	Austria_2006
Austria_1999	3148.2	12	Austria_2005
Austria_1998	3084.8	11	Austria_2004
Austria_1997	3002.2	8	Austria_2003
Austria_1996	2987.1	7	Austria_2002
Austria_1995	2934.8	7	Austria_2001
Denmark_2009	4246.6	40	Denmark_2015
Denmark_2008	4198.3	37	Denmark_2014
Denmark_2007	4247.2	m	Denmark_2013
Denmark_2006	4207.1	35	Denmark_2012
Denmark_2005	4287.0	33	Denmark_2011
Denmark_2004	4021.9	31	Denmark_2010
Denmark_2003	3747.9	36	Denmark_2009
Denmark_2002	3654.3	35	Denmark_2008
Denmark_2001	3586.0	32	Denmark_2007
Denmark_2000	3543.5	32	Denmark_2006
Denmark_1999	3571.3	31	Denmark_2005

Denmark_1998	3458.9	27	Denmark_2004
Denmark_1997	m	27	Denmark_2003
Denmark_1996	3330.1	23	Denmark_2002
Denmark_1995	3244.7	22	Denmark_2001
Finland_2009	5555.5	40	Finland_2015
Finland_2008	5826.8	40	Finland_2014
Finland_2007	5845.0	m	Finland_2013
Finland_2006	5866.5	39	Finland_2012
Finland_2005	5832.5	38	Finland_2011
Finland_2004	5735.3	37	Finland_2010
Finland_2003	5594.1	36	Finland_2009
Finland_2002	5457.1	33	Finland_2008
Finland_2001	5389.4	32	Finland_2007
Finland_2000	5219.5	29	Finland_2006
Finland_1999	5090.1	27	Finland_2005
Finland_1998	4852.2	24	Finland_2004
Finland_1997	4405.2	23	Finland_2003
Finland_1996	4174.7	21	Finland_2002
Finland_1995	4013.9	18	Finland_2001
France_2009	3468.7	27	France_2015
France_2008	3473.8	26	France_2014
France_2007	3517.2	m	France_2013
France_2006	3572.8	27	France_2012
France_2005	3571.7	27	France_2011
France_2004	3549.7	26	France_2010
France_2003	3504.8	26	France_2009
France_2002	3377.7	24	France_2008
France_2001	3402.6	24	France_2007
France_2000	3393.6	24	France_2006
France_1999	3404.3	22	France_2005
France_1998	3444.0	22	France_2004
France_1997	3516.5	22	France_2003
France_1996	3579.0	19	France_2002
France_1995	3559.6	18	France_2001
Greece_2009	m	38	Greece_2015
Greece_2008	m	37	Greece_2014
Greece_2007	5415.9	m	Greece_2013

Greece_2006	5883.9	21	Greece_2012
Greece_2005	5841.1	21	Greece_2011
Greece_2004	5403.9	20	Greece_2010
Greece_2003	5090.6	19	Greece_2009
Greece_2002	4806.2	19	Greece_2008
Greece_2001	4352.2	19	Greece_2007
Greece_2000	3855.4	18	Greece_2006
Greece_1999	3554.8	17	Greece_2005
Greece_1998	3445.2	17	Greece_2004
Greece_1997	3363.1	17	Greece_2003
Greece_1996	3069.1	17	Greece_2002
Greece_1995	2785.0	17	Greece_2001
Ireland_2009	4005.8	39	Ireland_2015
Ireland_2008	3984.6	38	Ireland_2014
Ireland_2007	4337.3	m	Ireland_2013
Ireland_2006	4332.7	33	Ireland_2012
Ireland_2005	4438.0	31	Ireland_2011
Ireland_2004	4571.0	30	Ireland_2010
Ireland_2003	4492.9	29	Ireland_2009
Ireland_2002	4442.7	31	Ireland_2008
Ireland_2001	4269.9	30	Ireland_2007
Ireland_2000	4180.9	28	Ireland_2006
Ireland_1999	3989.0	26	Ireland_2005
Ireland_1998	3813.5	26	Ireland_2004
Ireland_1997	3631.0	23	Ireland_2003
Ireland_1996	3492.1	23	Ireland_2002
Ireland_1995	3330.3	20	Ireland_2001
Italy_2009	3382.9	25	Italy_2015
Italy_2008	3394.9	24	Italy_2014
Italy_2007	3438.8	m	Italy_2013
Italy_2006	3443.8	22	Italy_2012
Italy_2005	3435.2	21	Italy_2011
Italy_2004	3404.5	20	Italy_2010
Italy_2003	3298.8	20	Italy_2009
Italy_2002	3216.0	20	Italy_2008
Italy_2001	3159.6	18	Italy_2007
Italy_2000	3097.3	17	Italy_2006

Italy_1999	3151.1	15	Italy_2005
Italy_1998	3278.5	12	Italy_2004
Italy_1997	3317.7	12	Italy_2003
Italy_1996	3109.3	12	Italy_2002
Italy_1995	3136.8	12	Italy_2001
Netherlands_2009	3731.5	43	Netherlands_2015
Netherlands_2008	3645.8	42	Netherlands_2014
Netherlands_2007	3584.5	m	Netherlands_2013
Netherlands_2006	3534.0	40	Netherlands_2012
Netherlands_2005	3459.4	38	Netherlands_2011
Netherlands_2004	3343.3	38	Netherlands_2010
Netherlands_2003	3258.2	38	Netherlands_2009
Netherlands_2002	3214.5	38	Netherlands_2008
Netherlands_2001	3153.3	35	Netherlands_2007
Netherlands_2000	3068.1	34	Netherlands_2006
Netherlands_1999	2972.7	34	Netherlands_2005
Netherlands_1998	2934.6	32	Netherlands_2004
Netherlands_1997	2999.3	25	Netherlands_2003
Netherlands_1996	3163.0	25	Netherlands_2002
Netherlands_1995	3255.0	24	Netherlands_2001
Norway_2009	4539.7	34	Norway_2015
Norway_2008	4457.0	34	Norway_2014
Norway_2007	4563.4	m	Norway_2013
Norway_2006	4600.5	44	Norway_2012
Norway_2005	4626.3	46	Norway_2011
Norway_2004	4659.7	46	Norway_2010
Norway_2003	4656.8	45	Norway_2009
Norway_2002	4343.3	44	Norway_2008
Norway_2001	4209.5	41	Norway_2007
Norway_2000	4251.1	40	Norway_2006
Norway_1999	4197.5	39	Norway_2005
Norway_1998	4122.1	37	Norway_2004
Norway_1997	4199.5	37	Norway_2003
Norway_1996	4112.7	37	Norway_2002
Norway_1995	3967.3	35	Norway_2001
Portugal_2009	3522.1	33	Portugal_2015
Portugal_2008	3563.4	31	Portugal_2014

Portugal_2007	3475.9	m	Portugal_2013
Portugal_2006	3492.7	28	Portugal_2012
Portugal_2005	3634.9	27	Portugal_2011
Portugal_2004	3783.2	25	Portugal_2010
Portugal_2003	3853.1	23	Portugal_2009
Portugal_2002	3827.4	23	Portugal_2008
Portugal_2001	3756.6	21	Portugal_2007
Portugal_2000	3636.2	20	Portugal_2006
Portugal_1999	3485.1	19	Portugal_2005
Portugal_1998	3449.4	12	Portugal_2004
Portugal_1997	m	13	Portugal_2003
Portugal_1996	3157.4	12	Portugal_2002
Portugal_1995	2982.3	11	Portugal_2001
Sweden_2009	4542.2	35	Sweden_2015
Sweden_2008	4410.0	35	Sweden_2014
Sweden_2007	4519.8	m	Sweden_2013
Sweden_2006	4650.6	34	Sweden_2012
Sweden_2005	4725.5	34	Sweden_2011
Sweden_2004	4783.0	34	Sweden_2010
Sweden_2003	4636.7	34	Sweden_2009
Sweden_2002	4296.0	32	Sweden_2008
Sweden_2001	4027.8	31	Sweden_2007
Sweden_2000	3909.7	31	Sweden_2006
Sweden_1999	3780.9	28	Sweden_2005
Sweden_1998	3167.9	26	Sweden_2004
Sweden_1997	3106.6	24	Sweden_2003
Sweden_1996	2951.7	22	Sweden_2002
Sweden_1995	2786.2	20	Sweden_2001
United Kingdom_2009	3881.7	41	United Kingdom_2015
United Kingdom_2008	3776.2	41	United Kingdom_2014
United Kingdom_2007	3863.9	m	United Kingdom_2013
United Kingdom_2006	3851.9	40	United Kingdom_2012
United Kingdom_2005	3799.3	39	United Kingdom_2011
United Kingdom_2004	3755.4	38	United Kingdom_2010
United Kingdom_2003	3842.0	36	United Kingdom_2009
United Kingdom_2002	3778.5	31	United Kingdom_2008
United Kingdom_2001	3499.2	29	United Kingdom_2007

United Kingdom_2000	3438.5	29	United Kingdom_2006
United Kingdom_1999	3547.6	27	United Kingdom_2005
United Kingdom_1998	3316.0	23	United Kingdom_2004
United Kingdom_1997	3246.4	24	United Kingdom_2003
United Kingdom_1996	3135.1	23	United Kingdom_2002
United Kingdom_1995	3131.5	21	United Kingdom_2001

Notes:

- Enrollment in tertiary education per 100,000 inhabitants. Source: World Bank.
- 25- to 34-year-old adults with university (BA+MA) degree. Source: OSCE Education at a glance.
- m = missing value
- Correlation between enrollment and teaching performance (6-year time lag): $r = 0.49$.

Enrolments in tertiary education at T_0 and education attainment at T_1 (with a 6-year time lag) are rather strongly correlated ($r = 0.49$). In other words, the number of graduates is strongly affected by how many students enrol. However, this correlation is only partial, and many more factors could have an impact on the teaching performance of HESs.

In addition, enrollment in tertiary education may have been an omitted condition (Radaelli and Wagemann 2018) in our QCA. However, in Radaelli's and Wagemann's (2018) view, an 'omitted condition problem' more likely arises when: *i*) the frequency of contradictory truth table rows and the distribution of raw consistency values are not ideal; *ii*) empirically contradictory cases (lower right part of the final XY plot) or cases that are not explained (upper left part of the final XY plot) exist; *iii*) a high number of paths results that come very close to or even equal the number of cases. No such issues occur in our analysis. However, "...even in the best scenario [...] this does not liberate us from the doubt that there could yet be any better explanatory factors that – either additionally, or substituting some of them – would improve the explanation" (*ibidem*, p. 11). As such, we cannot exclude a priori that any further omitted condition would have improved our analysis.

In conclusion, in our paper, which is mainly exploratory, we examine how (much) institutional autonomy, evaluation, universities' internal governance and the amount of public funding are associated with good teaching performance. We are aware that our model, like any model, does not explain everything. We aim to shed light on factors that have been scrutinized at length by following a different and innovative methodological approach. Furthermore, we focus on those tools that are in the hands of government when addressing higher education policies.

Appendix B

Robustness check

QCA with a 5-year time lag between conditions and the outcome

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

<i>Condition</i>	<i>Outcome</i>		<i>~Outcome</i>	
	<i>Consistency</i>	<i>Coverage</i>	<i>Consistency</i>	<i>Coverage</i>
Institutional autonomy	0.73	0.65	0.35	0.44
~Institutional autonomy	0.43	0.89	0.35	0.44
Evaluation	0.80	0.65	0.89	0.44
~Evaluation	0.31	0.82	0.29	0.47
Governance	0.62	0.91	0.29	0.26
~Governance	0.49	0.53	0.90	0.59
Public funding	0.56	0.91	0.26	0.25
~Public funding	0.54	0.54	0.91	0.56

Coefficients indicating a necessity relation (>0.9) are shown in bold.

Logical No (~) refers to the absence of a condition.

Here, the analysis is similar to the one conducted in the main text. Again, although no condition (and its non-occurrence, which is indicated with a tilde [~]) is necessary for the outcome, consensual internal governance and low public funding emerge as necessary conditions for the absence of the outcome.

Table AB2 *Analysis of sufficiency (5-year time lag): Truth table*

A	E	G	F	Number	Good Teaching Perform.	Raw consist.	PRI consist.	SYM consist.
0	0	1	0	2 (75%)	1	1	1	1
1	1	0	1	3 (58%)	1	0.93	0.89	0.89
1	1	1	1	1 (83%)	1	0.91	0.86	0.86
1	1	1	0	1 (91%)	1	0.79	0.64	0.64

0	1	0	0	1 (100%)	1	0.78	0.51	0.51
1	1	0	0	4 (33%)	0	0.35	0.08	0.08
1	0	1	1	0 (100%)	/	/	/	/
1	0	1	0	0 (100%)	/	/	/	/
1	0	0	1	0 (100%)	/	/	/	/
1	0	0	0	0 (100%)	/	/	/	/
0	1	1	1	0 (100%)	/	/	/	/
0	1	1	0	0 (100%)	/	/	/	/
0	1	0	1	0 (100%)	/	/	/	/
0	0	1	1	0 (100%)	/	/	/	/
0	0	0	1	0 (100%)	/	/	/	/
0	0	0	0	0 (100%)	/	/	/	/

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

Complex solution: $A^*E^*G + A^*E^*F + \sim A^*E^*G^*\sim F + \sim A^*\sim E^*G^*\sim F$ (coverage 0.89; consistency 0.90).

Parsimonious solution: $\sim E + \sim A + G + F$ (coverage 0.97; consistency 0.87).

Table AB3 Analysis of sufficiency (5-year time lag): intermediate solution

<i>Solution terms</i>	<i>Raw coverage</i>	<i>Unique coverage</i>	<i>Consistency</i>	<i>Cases covered</i>
G	0.62	0.26	0.91	Eng (1, 0.95); Ire (1, 0.98); Ned (0.70, 0.77); Swe (0.95, 0.98)
$E^*\sim A$	0.27	0.09	0.84	Gre (0.65, 0.76)
F^*E	0.55	0.23	0.92	Den (0.89, 0.93); Fin (0.90, 0.77); Nor (0.93, 0.91); Swe (0.88, 0.98)

Solution coverage (proportion of membership explained by all paths identified): 0.966488

Solution consistency ('how closely a perfect subset relation is approximated') (Ragin 2008, 44): 0.880342

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 substantially demonstrates that changing the time lag does not affect results considerably. With regard to the intermediate solution, two solution terms – the presence of verticalized governance and the presence of evaluation tools combined with no substantial increase in institutional autonomy during the time period –

are the same as in the main text, and the cases covered by these solution terms do not change. The third solution term is slightly different: the presence of funding now goes hand in hand with the presence of evaluation. In this case, countries that are characterized by this path are again the same. In conclusion, our results appear to be robust regardless of the time lag used.

Appendix C

Calibration and coding

Calibration of conditions and the outcome

The calibration of conditions and the outcome need to be justified at length, especially with regard to the point of maximum ambiguity.

For the calibration of institutional autonomy, we based our choices on the fact that according to the literature and historical evidence, all Continental European governments have been developing an autonomistic policy for reforming HE (10 out of 12 of our national cases). Furthermore, according to the comparative literature on the diachronic evolution of institutional autonomy (Huisman 2009; Shattock 2014; Capano, Regini and Turri 2016), we needed to take into consideration that the “autonomization” process of 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 deserves more than one decision to find a point of equilibrium. Because institutional autonomy is given by the State, the State can intervene at various times to adjust how the autonomistic policy is working. It can intervene not only by giving more autonomy but also by reducing it. Furthermore, we should consider that in Continental European countries, the level of institutional autonomy was low, if it existed at all (Clark 1983; Capano 2008), whereas in England and Ireland, it was higher. Furthermore, we should be aware that the way institutional autonomy works may vary according to the context and that the composition of the given autonomy (financial, organizational teaching) may be quite different and thus have different effects.

Thus, taking into account the “contextual” dimension of institutional autonomy together with the temporal dimension (and the “give/subtract” dynamics), in calibrating this condition, the full-membership point cannot be too high. On the basis of these theoretical-empirical considerations, it appears reasonable to seek statistical support and to base the calibration starting from the average of variation of all the countries. We considered an increase of +0.25 points (on a scale ranging from 0 to 1) for the coefficient of institutional autonomy during the period under scrutiny the threshold for full membership in the “growing institutional autonomy” set (1/3 more than the average of the variation of all countries)¹⁷. Consequently, we fixed an increase of 0.13 points as the

¹⁷ When calibrating conditions (and the outcome), QCA scholars suggest not building on arithmetical reasons. However, when theoretical arguments are not available or clearly decisive and the empirical evidence is not sufficient to develop “empirical argumentation”, it is generally accepted to recur to means and medians (Reutter 2017, 10-12).

point of maximum ambiguity because this is a reasonable threshold between a real attempt to develop an autonomistic policy and a symbolic or mechanic trend (i.e., due to passive reception of the common template). Obviously, this choice may seem to penalize England because its grade for institutional autonomy was high in 1988. On the other hand, it rewards and penalizes in a reasonable manner those countries that started from similar situations but have performed differently. For example, Greece and Austria started from a very low level of institutional autonomy, but Austria gave its universities much more autonomy in the analyzed period, and Finland (which started with 0.27) performed significantly lower than Portugal (which started from 0.42). Finally, no increase in the degree of institutional autonomy constitutes the value for full non-membership, which is reasonable given that our condition is labelled *growing* institutional autonomy.

For the variation in the number and relevance of evaluation tools that governments imposed on universities, the choice was not simple because all of the selected countries adopted evaluation tools during the analyzed time period (accreditation, evaluation of research and teaching, performance funding). This choice was made according to the theoretical assumption that the introduction of these evaluative tools is based on the will of the decision makers, who have decided to delegate some autonomy to universities, to monitor and check whether those universities are behaving as expected. This 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 they can be capable of finding “opportunistic” strategies to deal with the periodic evaluation of research (Lane and Kivisto 2008). To avoid this risk, decision makers are expected not only to intervene often to redesign these tools (Schwarz and Westerheijden 2007; Joao Rosa and Amaral 2014) but also to give them specific intensity (which we have assessed in terms of the “quality” of the instrument adoption or change). Furthermore, we were deeply aware that evaluation has been adopted in an extensive way in all countries and that, according to the existing literature (Neave 2012), evaluation has become a pillar of any governmental strategy in steering HE policy, becoming a kind of pervasive instrument. Thus, we decided to be very selective and set an increase of 0.5 points (again, on a scale ranging from 0 to 1) as the threshold for full membership in the set of “severe evaluation” and thus to reward those countries that have shown significant commitment not only in introducing specific evaluative tools in a consistent way but also in designing them in a coherent and impacting way. Accordingly, an increase of 0.20 was fixed as the threshold for full non-membership because under this threshold, the adoption of evaluation can be considered marginal and lacking any clear impact on the steering of the HES. Finally, an increase of 0.33 points was set as the point of maximum ambiguity

to reward those countries that have a value that is sufficiently close to the average to be considered positive, although marginally.

In this manner, we have drastically penalized, as a full non-member, Italy, which developed its real evaluative turn from 2010 (with the so-called “Gelmini Law”), after the period that we consider for our QCA analysis (1988-2008). We have consistently rewarded those countries that have drastically invested in evaluation, such as the Netherlands, Norway, Sweden and France. Eight countries overcome the point of maximum ambiguity and are thus characterized by severe evaluation. This can be considered reasonable to understand the collective trend towards increasing evaluation in HE in recent years.

Calibrating governance is apparently more straightforward, but it also needs a theoretical foundation. For this dimension, we have decided to be very selective. We fixed 1 for the full membership in the set of verticalized governance and 0.5 as the point of maximum indifference, following the most relevant literature on university governance. According to this literature, we know that all countries have pursued a strategy of increasing the verticalization of internal governance (except the two countries in which it was already present at the beginning of the temporal range considered, 1988). Furthermore, according to this literature, we can argue that the longer universities’ internal governance is verticalized, the more likely we are to see effects in terms of outcome (Huisman 2009; Shattock 2014; Bleiklie, Enders, and Lepori 2017). In fact, the impact in terms of outputs and outcomes of the shift from a democratic internal governance to a verticalized one cannot be expected to be immediate, at least in aggregate/systemic effects, because universities need time to adapt to the new “corporate” logic and to learn how to behave by dealing with internal resistance. This is why we were quite selective and have decided to consider 1 as the point of full membership, meaning the presence of verticalized governance for all the years of the analyzed temporal range. Thus, only two countries are completely rewarded for this (Ireland and England). At the same time, we have considered it reasonable to establish 0.5 as the maximum indifference point to underline the need for at least 11 years of verticalized internal governance out of 21 (1988-2008) to allow the possible aggregate effects of the shift to verticalized governance. In other words, to be considered more in than out of the set of “verticalized governance”, a country should experience more years with verticalized governance than years with consensual governance, which appears to be perfectly reasonable.

The calibration of the condition “generous public funding” is particularly challenging because there is not enough theoretical-empirical support from the existing literature. In fact, as we have shown in the state of the art in the main text of the paper, there is no conclusive evidence of the relationship between the amount of public funding and systemic performance, but there are also no clear indications with respect to the

“fair”, “coherent” amount of public funding needed in a system. This is particularly true when the focus is on the tools/procedures by which the public funding is allocated. At the same time, the percentage of public funding devoted to higher education is also an indicator of the social/political relevance of the field, and this dimension must be taken into consideration. This is because, given the different negative and positive trends in public funding among the analyzed countries in the analyzed period, we decided to reward those countries that have significantly higher funding for HE with respect to the others. This is because we chose 2.0% as the threshold of the GDP for full membership in the set of “generous public funding”. In this way, we have rewarded only one country, Denmark, while by considering 1.5% as the point of maximum ambiguity and 1% as the full-non membership point (the mean of the selected countries in 1990), we have built on the large gap that exists between the lowest of the four Scandinavian countries, Sweden (1.81%), and the highest of the others (Austria and the Netherlands, with 1.32%). The gap is so significant that establishing the ambiguity point in the middle seems to be reasonable and the only way to reward the top performers.

The choice of calibrating our outcome – good teaching performance – was also not simple because no literature addresses this issue. We have shown above in Appendix A that there is a significant, although not definitive, correlation between the enrollments and the number of graduates. Thus, we are working on a very delicate and sensitive issue. We decided that the influence of the trend of increasing enrollment could be weighted as 1 percentage point per year, thus fixing the full non-membership threshold at +20 percentage points (meaning an inertial increase due to increasing enrollment). Regarding the full membership threshold, we have assumed that the countries that were capable of improving their performance after an average of at least 1.5 percentage points per year could be considered very effective and capable of an average annual performance of 50% higher than the inertial trend. Finally, the point of maximum ambiguity is difficult to maintain because it affects the final assessment of the analyzed countries. In a nutshell, the selection procedure places 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), on the one hand, and France (22.50), on the other, appeared to be a “natural” place to put the maximum ambiguity point (25.00) and, in turn, to differentiate between countries showing a good teaching performance, on a side, and countries that do not, on the other.

Coding scheme for assessing (and measuring) changes in institutional autonomy and evaluation

In coding changes in institutional autonomy and the evaluation tools that governments imposed on universities, we followed two general rules: first, all the changes have been assessed according to the specific context and legacy respect to which they have been introduced (because apparently similar changes may be weighted in different ways in different countries); second, we used the qualitative classification of change that follows:

Table AC1 *How to code changes in institutional autonomy and evaluation*

<i>Kind of change</i>	<i>Definition</i>	<i>Coding</i>
Radical negative	Highly innovative negative (i.e., less institutional autonomy or less evaluation) change that deeply transforms the considered dimension with deep systemic effects expected.	-0.40
Significant negative	A change that has a consistent negative (i.e., less institutional autonomy or less evaluation) impact on the considered dimension with some systemic effects expected.	-0.30
Moderate negative	A change that is not a simple adjustment of the actual policy but whose negative (i.e., less institutional autonomy or less evaluation) impact on the considered dimension is expected to be limited.	-0.20
Marginal negative	A change that is expected to have a borderline negative (i.e., less institutional autonomy or less evaluation) impact on the considered dimension.	-0.10
No change	When no actual change on the considered dimension occurs.	0
Marginal positive	A change that is expected to have a borderline positive (i.e., more institutional autonomy or more evaluation) impact on the considered dimension.	+0.10
Moderate positive	A change that is not a simple adjustment of the actual policy but whose positive (i.e., more institutional autonomy or more evaluation) impact on the considered dimension is expected to be limited.	+0.20
Significant positive	A change that has a consistent positive (i.e., more institutional autonomy or more evaluation) impact on the considered dimension with some systemic effects expected.	+0.30
Radical positive	Highly innovative positive (i.e., more institutional autonomy or more evaluation) change that deeply transforms the considered dimension with deep systemic effects expected.	+0.40

However, not all legislative issues were separately coded by all coders in the same way initially. Approximately 15% of all legislative issues thus represented ‘contradictory

cases' for coding. When this occurred – i.e., when unanimity did not exist among the coders on how to evaluate such changes – we decided to attribute an intermediate coefficient between the two categories that the coders indicated. For example, when a legislative measure was considered to represent a 'radical negative' change by some coders but only a 'significant negative' change by others, we coded that same measure with a coefficient equal to -0.35; when a legislative measure was considered to represent a 'significant negative' change by some coders but only a 'moderate negative' change by others, we coded that same measure with a coefficient equal to -0.25; etc.

Examples of coding: Index of Institutional Autonomy

Table AC2 *Examples of coding on institutional autonomy: countries, regulation and issues*

<i>Country</i>	<i>Piece of regulation</i>	<i>Issue</i>	<i>Dimension</i>	<i>Coding</i>	<i>Before regulation (0-1)</i>	<i>After regulation (0-1)</i>
Austria	Universities Act 2002	Federal government may, by decree, direct one or more universities to establish a program of study if this is necessary for overriding educational or scientific policy reasons and no performance agreement to this effect is reached.	T3 Free to start new bachelor and master programs	-0.05	0.60	0.55
Denmark	Act on Universities 2003	Universities are transformed to publicly financed self-owning institutions and may therefore decide on their use of public grants (and other possible sources of income), within the regulations from the Ministry.	F2 Free to decide on internal allocation of public funds and keep surpluses	+0.40	0.30	0.70

England	Higher Education Act 2004	Universities may decide to change students' fees subject to a plan that has to be approved by the State (the fees established in the plan may not exceed the fee cap set by regulations).	F1 Free to charge fees to students	+0.20	0.20	0.40
Finland	Amendment 1504/2007	Amendment to the regulation of tuition fee free education: right to provide fee-based degree education for non-EU/EEA students if the fee is paid by a third party such as a foreign state.	F1 Free to charge fees to students	+0.05	0.00	0.05
France	Law 1199/2007	Local ad hoc recruitment committees replace nationally ruled "Commissions de spécialistes": choices of positions to be offered, employment status and recruitment are in the hands of the president and executive board.	O2 Free to determine procedures for academic recruitment / promotions / dismissal	+0.40	0.10	0.50
Greece	Law 2083/1992	Departments should draw up and forward to the Ministry of Education a four-year planning scheme that includes their planning for academic appointments.	O1 Free to set internal academic organization (faculties, departments, schools)	-0.10	0.30	0.20
Ireland	Law on Budget 1995	From 1996, tuition fees for first-time full-time	F1	-0.40	0.40	0.00

		undergraduate students were abolished in publicly funded institutions.	Free to charge fees to students			
Italy	Ministerial decree 509/1999	Establishment of the bachelor (three years) + master (two years) teaching programs	T3 Free to start new bachelor and master programs	+0.30	0.30	0.60
Netherlands	Stb. 303/2003	Gives institutions the possibility to offer flexible programs with an international orientation. Institutions have the choice to offer bachelors programs directly connected to master programs.	T2 Free to decide on the content of bachelor and master programs	+0.15	0.50	0.65
Norway	Quality Reform of HE 2001	Higher education institutions can now decide on the number of study places to offer and redistribute study places from one area of study to another.	T1 Free to decide on number of study places for bachelor and master's programs	+0.10	0.20	0.30
Portugal	Law 62/2007	Institutions are free to manage their own funds, including those given by the government, as long as they make their financial records publicly available.	F2 Free to decide on internal allocation of public funds and keep surpluses	+0.40	0.10	0.50
Sweden	Open University Act 2001	Each university and college establishes local action plans on how to broaden the recruitment of students.	T1 Free to decide on number of study places for bachelor and master's programs	+0.05	0.60	0.65

Examples of coding: Index of Evaluation

Table AC3 Examples of coding on evaluation tools: countries, regulation and issues

Country	Piece of regulation	Issue	Coding	Before regulation (0-1)	After regulation (0-1)
Austria	Universities of Applied Sciences Studies Act 1993	Establishment of the “Fachhochschule council”, which is in charge of accreditation of the programs of UoAS. It is also in charge of quality assurance, continuous evaluation, counselling to the Federal Minister in matters of the UoAS, yearly report to the Federal minister.	+0.20	0.05	0.25
Denmark	Act 403/2003	Universities must systematically develop and improve the quality of their teaching, education, and research.	+0.10	0.25	0.35
England	Further and Higher Education Act 1992	The distinction between universities and polytechnics is abolished, and the Universities Funding Council is replaced by region-wise funding councils such as the HEFCE, which is in charge of the institution of the Research Assessment Exercise in 1992.	+0.10	0.10	0.20
Finland	Formula funding 2001-2003	Core funding 95%: salaries and facilities 19% (input criteria); master’s degrees 46% (targets 2/3, realized 1/3) (output criteria); doctoral degrees 30% (targets 2/3, realized 1/3) (output criteria). Funding of specific operations (earmarked) 5%: graduate schools; open university activities.	+0.10	0.00	0.10
France	Law 450/2006	This law created the <i>Agence d’évaluation de la recherche et l’enseignement supérieur</i> (AERES – the research and higher-education assessment agency), assigned to assess the following: <i>i</i>) higher-education and research institutes; <i>ii</i>) scientific co-	+0.20	0.30	0.50

		operation foundations; <i>iii</i>) higher-education training courses and degrees.			
Greece	Law 3374/2005	This law introduced a quality assurance system in HE. The Hellenic Quality Assurance Agency in Higher Education (HQA) was established as an independent and specialized agent operating under the supervision of the Ministry of Education.	+0.20	0.05	0.25
Ireland	Qualifications Act 1999	This Act established a National Qualifications Authority (NQA) responsible for establishing and maintaining the NQF.	+0.10	0.10	0.20
Italy	Law 370/1999	Establishment of the ' <i>Nuclei di valutazione interna degli atenei</i> ', which are collegial bodies assessing teaching quality.	+0.10	0.10	0.20
Netherlands	Stb. 387/2002	This Act stipulated the mandate of the Inspectorate to oversee and report upon the quality of (higher) education. Its core business is to assess quality (in the different education sectors). The Act was meant to (further) secure the autonomy of the Inspectorate.	+0.10	0.20	0.30
Norway	Act 2006	The revised regulation specifies new requirements for the number of academic staff involved in education that must have doctoral degrees or similar (now at least 50 % of the staff in master's programs).	+0.10	0.45	0.55
Portugal	Decree-law 369/2007	Created a new independent agency – the A3ES – responsible for HE evaluation and accreditation.	+0.20	0.10	0.30
Sweden	Law 169/1992	Each university and college is given a three-year training mission that includes a framework for the operation, payment for services rendered and requirements for quality assurance systems.	+0.10	0.00	0.10

References

- Bleiklie I, Enders J, Lepori B (eds.) (2017) *Managing Universities: Policy and Organizational Change from a Western European Perspective*. London: Palgrave-MacMillan.
- Capano G (2008) Looking for Serendipity: The Problematical Reform of Government within Italy's Universities. *Higher Education*, 55, 481–504.
- Capano G, Regini M, Turri M (2016) *Changing Governance in Universities. Italian Higher Education in Comparative Perspective*. London: Palgrave MacMillan.
- Clark B (1983) *The Higher Education System. Academic Organization in Cross National Perspective*. University of California Press, Berkeley.
- Huisman J (eds) (2009), *International Perspectives on the Governance of Higher Education*. Routledge, London.
- Joao Rosa M, Amaral A (eds.) (2014) *Quality Assurance in Higher Education*. London: Palgrave.
- Lane JE, Kivisto J (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 TM (1984) The new economics of organization. *American Journal of Political Science* 28(4): 739–777.
- Neave G (2012) *The Evaluative State, Institutional Autonomy and Re-engineering Higher Education in Western Europe*. Palgrave, London.
- Radaelli C, Wagemann C (2018) What did I leave out? Omitted variables in regression and qualitative comparative analysis. *European Political Science*, online first 18 January, <https://doi.org/10.1057/s41304-017-0142-7>.
- Reutter W (2017), The Changeableness of Subnational Constitutions: A Qualitative Comparative Analysis. *Government & Opposition*, online publication first, doi: 10.1017/gov.2016.45.
- Schwarz S, Westerheijden D (eds.) (2007) *Accreditation and evaluation in the European higher education area*. Dordrecht: Springer.
- Shattock ML (ed.) (2014) *International Trends in University Governance*. London: Routledge.