Science and Public Policy, 46(6), 2019, 834–852
doi: 10.1093/scipol/scz033
Advance Access Publication Date: 5 August 2019
Article

OXFORD

Analysing the determinants of higher education systems' performance–a structural equation modelling approach

Tommaso Agasisti† and Alice Bertoletti*,†

School of Management, Politecnico di Milano, Milan 20156, Italy

*Corresponding author. Email: alice.bertoletti@polimi.it
†The authors wish it to be known that in their opinion both the authors should be regarded as joint first authors.

Abstract

This article deals with the relationships between the performance of higher education systems (HESs) and their potential determinants. The research employed data relating to twenty-nine European countries, within a time span of 15 years. The empirical work is based on a conceptual framework that has been tested through Structural Equation Modelling. The data was selected from a novel data set that we have built by gathering HES indicators from the major international databases. Among the more noteworthy results, we found no statistically significant correlation between graduation rates and research performance. Additionally, we found a moderate effect of public expenditure on HES performance. The article addresses the issue of the lack of a comprehensive set of systemic indicators for HES performance, providing a holistic picture of the relationship between HES determinants and performance, which, in turn, could help to provide an adequate evaluation of the policies in the field of higher education.

Key words: higher education system; performance evaluation; structural equation modelling

1. Introduction

The role of higher education (HE) is back with a vengeance in the agenda of European governments, and from the late 1990s onwards European countries have started to work on reforms of HE (Dobbin and Knill 2009). Over the past decade, these countries have also been facing increasing pressure to evaluate the economic, social, and cultural paybacks of HE in order to justify public and private investment in HE systems (HESs) (OECD 2017).

Against this backdrop, there is the increasing need to carry out an exhaustive assessment of HE policies and reforms where all the impacts of HE are taken into consideration. A fundamental requirement for achieving this purpose is to implement a full and comprehensive model for evaluating HES performance. The need for performance indicators (PIs) for HESs is also stressed in a report published by the International Institute for Educational Planning, a UNESCO affiliate:

> Many countries [. . .] perceive the need for an indicator system to improve communication on the progress of their higher education systems to the public at large and funding organisations, as well as to monitor the implementation of their public higher education policies (Martin et al. 2011: 9).

However, after more than three decades of HE reforms, there is still no adequate model to evaluate HES performance. This lack can be explained through three main gaps extant in literature. First, performance in HE has so far reached only a fragmented representation in literature (Cunha and Rocha 2012). Secondly, there is very little research focused on systemic performance in HE, partially caused by the scarcity of available data at country level (Hanushek and Woessmann 2011). Thirdly, although the outcome of HE should be considered as the combination of outputs from teaching, research, and the third-mission (Martin 2012), there are few papers that analyse these three dimensions jointly. Therefore, the existing literature cannot provide an overall picture of systemic (i.e. country level) HE performance, while this is, instead, essential for evaluating HE policies.

The aim of this article is to fill these gaps by studying the performance of HESs and their respective determinants, providing the model most suited to representing them in a comprehensive way (taking in the combined effect of teaching, research, and third-mission activity) explicitly from a long-term perspective.

The main purpose of this work is to describe the relationship between HE performance and its determinants, adopting a diachronic perspective. The study, therefore, examines the following two research questions:

a. Which are the main factors associated with HES performance, and how are they inter-related?
b. How has the relationship between HES performance measures (and their determinants) varied over time between 2000 and 2014?

This study addresses these research questions by expressly adopting a descriptive intention. For this reason, the article is not structured with the purpose of finding the causal relationships between performance and its determinant factors, but rather to provide a diachronic description of these relationships within the European context. In accordance with its descriptive intention, the article answers the research questions by developing a conceptual framework and testing it empirically through a Structural Equation Modelling (SEM) approach. The empirical analyses were conducted on a sample of twenty-nine European countries over a time span of 15 years. As a by-product, the research contributes to the literature with a novel data set of HES indicators collected from the major international agencies (OECD, Eurostat, and World Bank) and integrated with data from specialist data sets containing information about research and third-mission performance (InCites, SciVal, SCImago Journal & Country Rank, the European Innovation Scoreboard, and Patstat).

In this research, the main unit of analysis corresponds to the HES, defined as a group of universities or academic institutions that operate within a given administrative territory (Filippakou et al., 2012). Here, by this, we refer to the national territory of each European country. We consider the HES as comprising universities- and vocational-type institutions, and assess the determinants of systems' performance all together without separating the contribution of the two sub-sectors. In this perspective, the analyses presented here refer to indicators intended to capture phenomena that are 'system level' in a broad sense.

The article is structured as follows. Section 2 contains a review of the key literature covering the analyses for HES performance and the indicators with which they can be measured in Europe and worldwide. The main aim of this part is to provide a state-of-art overview of the theoretical and empirical research in the field. In Section 3, we develop our conceptual framework on the basis of the previous review. Section 4 describes how the data set was collected and created, and describes the selected variables that define the empirical model. In Section 5, we present the details of our methodological approach based on SEM. Section 6 contains the main findings, which are then discussed, along with their contributions to research and the policy implications, in Section 7.

2. Background and literature review

There is a long history behind the issue of evaluating performance in HE:

> Students are educated, research is profound and some professors do some University or public service. But at the end of each year how can we measure the achievement of our most cherished goals? (Richman and Farmer 1974).

Since this question was first asked, PIs for HE have played a wide range of roles in the academic literature (Cave et al. 1995).

This section focuses on a specific branch of this research field, which studies the performance of HESs, that is, where the focus is on countries in their entirety and not on single universities or students. We have summarised the indicators for HES performance and the respective determinant factors that have been identified through the main studies on this topic given in Table 1.

The overview of literature shows that studies evaluating systemic performance in HE are numerically few. This shortage is due, on the one hand, to the complexity of HESs and the lack of a clear single 'objective' in HE (Phillimore 1989) and, one the other hand, to the problem of availability of HE data at system level. However, a system-level approach is essential to provide cross-country comparisons, since the much larger variations between different countries (which do not exist within the individual countries) allow a better understanding of the factors that are statistically associated with higher performance (Hanushek and Woessmann 2011).

Taking these limitations into account, some authors have proposed a method for evaluating HESs by defining PIs as an aggregation of the scores achieved by universities in international rankings. Within this branch of studies, Docampo (2011) has used the Shanghai ranking to assess the research performance of university systems in thirty-two OECD countries. Using the Principal Component Analysis approach, the author has proposed a single composite index for measuring research performance. However, these rankings are considered often inappropriate, since they represent only a certain type of performance. HE rankings can lead to dangerous biases, favouring universities with high performance in research, which is more readily measurable. Moreover, since ranking scores are obtained by data which have been transformed and combined through non-transparent processes, the use of rankings can generate problems in the interpretation of results (Williams and de Rassenfosse 2016).

Williams et al. (2013) have provided a different methodology for measuring the performance of HESs. Their paper examines the determinants of HES performance, measured through a set of indicators and grouped into three different types of output: teaching, research, and training performance. The aim of their work was to analyse the determinants of HESs, taking into account three types of determinant factors: resources, environment, and connectivity. The authors measure the effect of these determinants over the overall output of HESs through a simple regression model. The value of the overall output is calculated as the weighted arithmetic mean of the output indicators. It follows that there still is a problem with the subjectivity of the weights, even though the criteria for aggregation are more transparent than in the models using ranking scores.

The problem of assigning arbitrary weights has been debated by Saisana et al. (2011), who found that the inferences of the rankings for HE institutions lack robustness. Addressing this issue, Hoareau et al. (2013) developed a 'non-arbitrary' method based on a Factory Analysis used at the preliminary stage of the model. This technique is used to reduce multi-collinearity, which is typically found between ranking indicators. The authors have applied this method in a study on the impact of HESs on the economic development and level of innovation in thirty-two European countries. In particular, the work investigates the impact of two types of performance of HES, 'Research productivity and attractiveness' and 'Graduate employment and graduation'.

Measures of HE performance (outputs) and their determinants (inputs) have been provided extensively within the field of studies on HE efficiency based on production frontier models. However, there are still very few contributions in this field that measure the HE efficiency at system level. Obadić and Aristovnik (2011) used Data Envelopment Analysis (DEA) to study the efficiency of government spending on HESs in a sample of new European Union (EU) member

Table 1. List of previous studies on performance in higher education systems.

References	Method	Unit of analysis	Determinants factors	Teaching PIs	Research PIs
St. Aubyn et al. (2009)	Two-stage DEA and SFA	Twenty-eight EU countries plus Japan and USA	Academic staff; students in postgraduate diploma (PGD) institutions; Total expenditure (exp.) on PGD institutions in % of GDP; total exp. on PGD institutions; total public exp. on HE; Total public exp. on HE institutions; PISA data (institutional and environmental); supply of HE	Graduates in PGD institutions; recruiter survey ranking THES—QS; peer survey ranking THES—QS	Published articles; number of citations
Hien (2010)	Qualitative comparison	Eleven East and Southeast Asian countries			Total number of peer-refereed international publications; mean score of no. of citations; contribution of domestic authors to PRIP production
Docampo (2011)	PCA	Thirty-two OECD countries			Average score of a university system on the indicator of ARWU measures: alumni; award; HiCi; N&S; PUB
Obadić and Aristovnik (2011)	DEA	Thirty-seven EU and OECD countries	Expenditure per student, tertiary education (% of GDP per capita)	Enrolment rate in HE; labour force with an HE; unemployed population with a tertiary education	
Agasisti (2011)	DEA	Eighteen European countries	**Input:** Exp. on HE institutions (from public and private sources); net entry rates to HE; ratio of students to teaching staff in HE Socio-economic and organisational factors: GDP per capita; Exp. per student; % of students in public universities; % of public funding; average years in education	Percentage of population with HE (25- to 34-year old); graduation rates; employment rate (25- to 34-year old); % of foreign students enrolled in HE	
Williams et al. (2013)	Regression model	Forty-eight countries world-wide	**Resources:** Government exp. on HE; total exp. on HE; total exp. per full-time equivalent student; Exp. on R&D in HE institutions **Environment**: Proportion of female students and female academic staff **Connectivity:** % of international students; % of articles with international collaborators	Shanghai ranking scores for a nation's best three universities; tertiary ER; % of population over 24 years old, with HE; weighted Shanghai ranking scores for universities; unemployment rate of tertiary-educated population compared to school leavers (training)	No. of articles produced by HE institutions; total articles per head of population; average impact of articles as measured by the Karolinska Institute; no. of researchers (training)
Hoareau et al. (2013)	Factor analysis and two-stage regressions	Thirty-two European countries	Expenditure on HE financial aid; expenditure per HE student; other policy features (qualitative data)	**Access:** No. of enrolled students; no. of students enrolled on a non-university alternative route. **Teaching:** Graduation rate (ISCED 5 and 6); employment rate among 18- to 34-year old (recent graduates)	Publications among the most cited 10%; no. of universities in the ARWU top 500 world ranking (per million inhabitants); no. of incoming Marie Curie fellows per million inhabitants; no. of ERC starting grants per million inhabitants; private–public co-publications
Hardeman and Van Roy (2013)[a]	Cluster analysis, multivariate analysis, PCA	Thirty-four European countries			Widely cited publications per total number of publications; PCT patent applications per million inhabitants; top universities and public research institutes per total R&D expenditure; ERC grants received per public R&D expenditure
Triventi (2014)	cluster analysis	Sixteen OECD HES	**Tracking:** % of time in primary and secondary school in a tracking regime; % of students in vocational upper secondary education. **Resources:** % of exp. for HE on total exp. in education; % of exp. for HE on GDP; student–teacher ratio. **Differentiation:** ER in non-public institutions; % of private resources on the total in HE; % of students enrolled in ISCED 5B. **Affordability:** Educational costs; total costs; out-of-pocket costs	Percentage of graduates in HE; % of graduates in ISCED 5A; expected years in HE; education equity index; relative wages of ISCED 5A graduates compared to ISCED 3A (25- to 64-year old); % of those employed among ISCED 5A graduates (25- to 64-year old)	

[a]This publication is a reference report by the Joint Research Centre of the European Commission.
Source: Authors' elaboration of academic works in literature.

countries, with a focus on teaching performance. The authors selected a single input, namely the expenditure per student in tertiary education, whereas a range of different HE outputs have been used to measure teaching performance. The focus on teaching performance has been also taken up by Agasisti (2011), who conducted a cross-country comparison of the efficiency of 16 HESs. The author, performing a DEA analysis, included in the empirical model a more comprehensive set of indicators, taking into consideration various features relating to HE (such as foreign students, students in public institutions, and public subsidies) and the country (such as employment and pro-capita GDP). Lastly, the effectiveness of HESs for both teaching and research performances has been assessed in the work of St. Aubyn et al. (2009). The authors have used two different methods, DEA and stochastic frontier analysis (SFA), to evaluate the efficiency of HESs with regard to teaching and the research output. Their PI framework has been developed over three dimensions: input indicators (e.g. academic staff and students enrolled), financial resources (e.g. expenditure in universities), and output indicators (e.g. graduates, university ranking, and number of publications).

The brief overview of literature on the research into HES performance has revealed two main gaps. First, the studies reviewed above focus mainly on one or other aspect of HE performance, overlooking the overall effect and without a holistic approach that includes teaching, research, and third-mission performance. However, a multidimensional output—one that covers research, teaching, and third-mission performance— is needed to elaborate a complete and proper evaluation of the determinants of HES performance: 'Given the importance of higher education, a nation needs a comprehensive set of indicators in order to evaluate the quality and worth of its higher education system' (Williams et al. 2013: 1). In particular, it is clear that there is a lack of research that takes third-mission performance into account, especially at system level. This shortcoming could depend on the complexity in identifying and measuring this area. However, third-mission performance is an important part of HE functions, reflecting as it does the interaction between universities and the rest of society. Second, there is a lack of research covering the evaluation of relationships between HES performance and its respective determinants. A structural understanding of how the different factors interrelate with each other and how they influence the performance of different HES operations has, in fact, not yet been provided. Moreover, there is still limited research investigating how the performance of different HES activities can influence one another.

In answering the research questions, this article seeks to fill these two gaps by analysing the issue from a diachronic perspective.

3. Conceptual framework

The article offers a conceptual framework to represent the interaction between the performance of HESs and their associated determinants. This framework was derived from the theoretical definition of HES, which in literature is identified as a multi-inputs and multi-outputs process that interacts with the context in which it operates (Bonaccorsi and Daraio 2007). As mentioned in literature, reference is made to three types of functions in HE, which are closely connected to the goals and missions of HESs: teaching, research, and the third-mission (Martin 2012). Based on this theoretical background, we developed the conceptual framework set out in Fig. 1.

The framework comprises seven factors indicating the key aspects of interest for analysing HESs. However, these factors cannot be measured directly and, for this reason, can be modelled as latent variables. Latent variables are factors of interest that, being intangible, cannot be observed through data. These latent variables, however, affect other related variables that can instead be measured. Taking these tangible variables into account, latent variables can be included and estimated through an empirical model. In this section, we describe the latent variables that compose the framework, while the selection of the specific indicators that measure them is presented in the next section.

As shown in Fig. 1, the framework is designed around two classes of factors: *determinants* and *performance*. Determinants are the factors that affect HES performance through a series of direct and indirect connections. In particular, our conceptual framework comprises four determinants. The economic context in which HESs operate is included through the factors of *economic development*. As shown in literature, since every HES is intrinsically part of a country, the economic factors of that country are important determinants of HES performance (see Agasisti 2011). The economic development of a country can influence its HES and performance directly or indirectly, for instance, by influencing the level of financial resources dedicated to HE. *Human capital stock* has been included to capture the general level of education of the country, and this factor enters the HES as an input. Linked to this is the idea that the performance of HES also depends on the initial conditions in the form of level of education in the country. For example, it is reasonable to think that countries where students achieve a high standard of learning at secondary level are more likely to have a good graduation rate. Certainly, this factor is also the expression of previous policies in education, being the outcome of the results obtained by HESs in the past. In our study, by including this factor, our intention is to represent the educational level of the nation from both quantitative (e.g. the educational level achieved by the population) and qualitative (e.g. the skills and the knowledge acquired by students through education) perspectives. Similar factors have been included by Williams et al. (2013), who have controlled for the HE attainment among the adult population; and by St. Aubyn et al. (2009), who worked with the Programme for International Student Assessment (PISA) standardised results.[1] The *HES Resources* consist of physical and financial resources used in the HES. As shown in the literature review, the resources used by HESs are the primary determinants studied in literature (see Hoareau et al. 2013; Williams et al. 2013). *Access to HES* measures how well the HES can attract students from inside and outside the country and from a diversity of backgrounds. In our study, access to HE is considered as an input to the framework, but it does have two facets to its meaning. On the one hand, as input to the framework, access to HE is a measure of the HES' capacity (e.g. the number of students enrolled) and so influences its performance (e.g. the resulting number of graduates; see Agasisti 2011; Docampo and Cram 2017). On the other hand, access to HE is also a measure of the system's equality (again, the number of students in university), and for this reason can be interpreted as a performance in itself (see Lucas 2001). According to the perspective adopted in this work, we consider this element as a mediating factor between resources and academic results.

The descriptions of the determinants highlight that these factors belong to different conceptual levels. Economic development and human capital (HC) stock are context variables, HE resources are a determinant variable specifically associated to HESs, and, lastly,

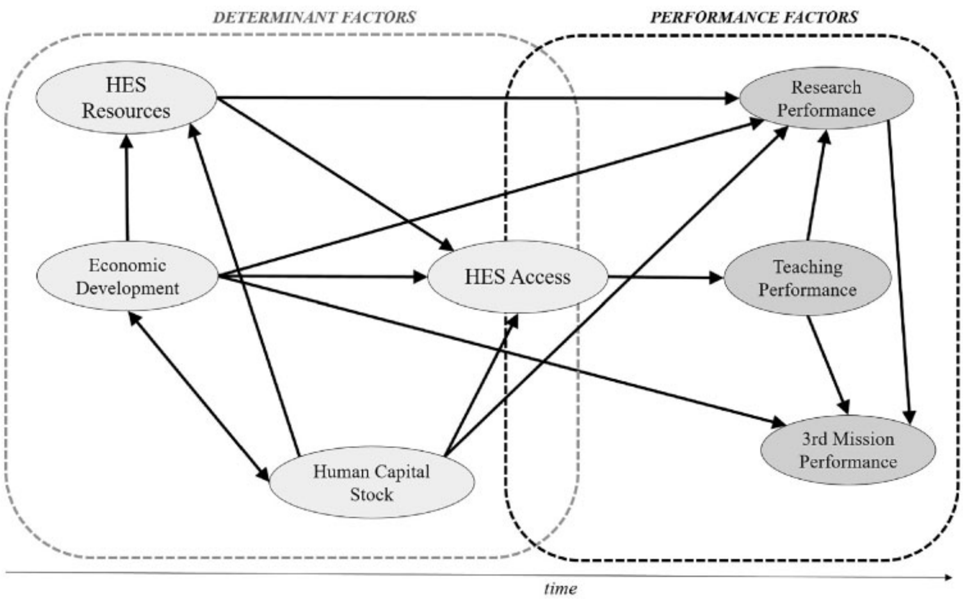

Figure 1. The conceptual framework/note: the figure shows the latent factors that indicate the key factors of interest. The determinant factors are shown in light grey and the performance factors in dark grey. The arrows show the direct effects and the double arrow represents the correlation between variables. *Source*: Authors.

HE access is a mediating factor between determinants and performance variables. The empirical approach (SEM) adopted in this article is particularly suited to analysing these connections, since it allows to estimate the relationships between different determinant factors of the framework (see the practical details presented in Section 5).

The three performance factors in the conceptual framework have been defined on the basis of the three main goals of HESs mentioned previously: teaching, research, and the third-mission. *Teaching performance* consists of the outputs linked to teaching and learning goals for HESs. This factor is a measure of the results achieved by students in education. In literature, performance in this area relates mainly to the number of graduates or graduation rate (see Agasisti 2011; St. Aubyn et al. 2009). *Research performance* refers to the research output of academic staff and HE researchers. The existent papers focus mainly on research productivity, which is measured in terms of publications and their academic impact (see Hien 2010; Hoareau et al. 2013). The factor of *third-mission performance* is meant to capture the contribution of universities to society, including the benefits outside the academic environment that arise from HE (see the definition given by the Higher Education Funding Council for England 2011). Teaching activities can contribute to third-mission by means of HC formation, while research produces knowledge that can be transferred to society (Veugelers 2016). Therefore, third-mission performance can be considered to arise from the other two HE missions and we can suppose that it is affected by teaching and research performance as represented in the conceptual framework. Nevertheless, third-mission also plays an important role per se, which is connected, for instance, to the widely discussed concept of 'entrepreneurial university' (Veugelers 2016). The set of activities that we can ascribe to third-mission is indeed multifaceted and complex (Molas-Gallart et al. 2002). Third-mission encompasses learning activities tailored to the industry and community needs and it includes public lectures and activities fostering continuing education. We can consider third-mission activities also the collaborations between researchers and industries, activities involving external research contracts and producing innovations with commercial values. Lastly, third-mission activities involve scientific dissemination and social networking aimed at fostering the diffusion of knowledge and skills from universities to the rest of society (see de Rassenfosse and Williams 2015; Molas-Gallart et al. 2002). In the article, we choose to focus on the university–industry connection. On the one hand, this perspective plays an important role in literature, especially in the research field related to 'entrepreneurial university'; on the other hand, the choice is linked to tangible measures and available data.

The value of framework does not merely lie in the identification and partitioning of the seven performance factors but also in the relationships that are established between each other (represented by arrows in Fig. 1). These relationships should be considered as theoretical hypotheses, which have been tested through empirical estimations. The two research questions presented in Section 1 have been addressed by analysing the framework from two different perspectives. The first research question intends to identify the relationships between factors, while the second research question examines how these relationships vary over time.

It is important to preliminarily note that the article does not take into consideration policy instruments or decisions and their link with performance. Policies can be seen as exogenous factors; however, they are not investigated through the framework. The future work will include information about policies and reforms at national level in the framework and therefore estimate their specific impact on performance. More specifically, we will take advantage of a novel data set developed in a funded project,[2] where details about the reforms implemented in the past 20 years is reported. The aim is to exploit the variation in the timing and scope of the various policies and investigate whether these variations are related to the trends in PIs at system level—considering an appropriate time lag for such policy changes to take effect.

Table 2. Description of data sources.

Name of database/ source	Description of the database/source
Education at a glance, OECD indicators	Databank collecting data from 1995 (1998 edition). It provides information on the state of education in all 35 OECD countries plus several others that use the same protocol to present their information. *Education at a Glance* contains indicators for several domains: the output of educational institutions; the impact of learning across countries; financial and human resources invested in education; access, participation and progression in education; and the learning environment and organisation of schools. Note that not all 29 EU countries analysed in the paper are also OECD countries. [9]
Eurostat	Source providing high quality statistics for European countries. The data set used here collected information from the 'Education and Training', 'Economic and Finance' and 'Labour Market' sections. The indicators relate to different time spans, in general, educational information relates to a time span of 12 years, from 2001 to 2012.
World Bank	Source providing free and open access to global development data by country. The section on 'Education and Statistics' provides a wide set of indicators on HES features, HES resources and teaching outputs. Other sections of the World Bank database have been used for information about country's characteristics. The time span of available data usually covers the period considered in the data set, with the exception of 2015, which is still not provided for all the indicators—at the time in which we conducted the data collection (October 2017).
Scimago journal & Country rank	Source including the journal's and country's scientific indicators elaborated from the information contained in the Scopus database. This databank was used for data on research outputs.
InCites	This source publishes data and statistics on research performance using data extracted from the Web of Science Core Collection.
PATSTAT	EPO Worldwide Patent Statistical Database contains bibliographical and legal status data relating to more than 100 million patent documents from leading industrialised and developing countries.

Source: Authors.

4. The data set and path diagram

4.1 The data set

This research contributes to the literature on HESs by building a novel data set of HES indicators collected from the major international databanks (OECD, Eurostat, and the World Bank) with further information from specialist data sets containing data on research and third-mission performance (InCites, SciVal, SCImago Journal & Country Rank, European Innovation Scoreboard, and Patstat database). This data set will act as the first step in filling the lack of systemic data on HE performance, revealed in literature and at institutional level.

Based on the diachronic perspective adopted in the article, the data covers a long time span, from 1995 to 2015. The data set focuses on European HESs, with data collected for twenty-nine European countries, namely all twenty-eight EU members with the exclusion of Luxemburg plus Norway and Switzerland.[3] The result is a panel data set of 150 variables (see Supplementary Annex, Section A1) with indicators coming from the eight different sources of data described in Table 2.

The process of collecting data has highlighted the difficulty in obtaining a complete set of data for each country, an issue where the reasons lay in two main causes. In the first place, the three main databases (Eurostat, OECD, and the World Bank) cover the years and countries examined in the study only partially. In the second place, although the same information could be obtained from indicators in different data sources, since the definitions used vary, data cannot be easily combined. As explained in the following sections, this problem also affected our empirical analysis, limiting the choice of indicators to use in the model. In order to reduce the effect of missing data, we included in our data set other indicators describing the same latent factor, coming from several different data sources (among the eight we have used). By using this strategy, it was possible to identify the indicator with the greatest amount of data available among all the various data sets examined. The process of collecting data has also shed light on the lack of availability of some information at system level, especially that regarding third-mission performance. For instance, even though national databases of university spin-offs are available for some European countries, an international data source of HE spin-offs does not exist. More generally, systemic data for third-mission activities are scarcely provided and they are mainly limited to tangible indicators. Our effort was to include the most widely available and updated indicators, as described in the following section.

4.2 Variable selection and path diagram

The variables in the data set have been used to describe the key factors of interest represented in the conceptual framework. Since these factors are latent variables, they cannot be described directly through observable data. Based on the SEM approach, we selected a set of indicators that can be affected by latent factors from among the pool of observable variables in the data set.

As the first step, the selection process has been driven by the evidence in literature and influenced by the availability of the data. The result is illustrated in the path diagram of Fig. 2, where we have linked each latent variable to the selected indicators that measure them. The descriptions of the specific indicators chosen for the analysis are given in Table 3, and are commented in the following Section 4.3.

In order to address our second research question, we have tested this path diagram empirically at three different moments in time. Comparing the results for the three estimations, we were able to examine the evolution of the relationships between the determinants and the HES performance over the 15 years of the study. As shown in the path model, there is a time lag between the determinant factors and the performance factors. In reality, the outputs at time t cannot be compared with the contemporaneous inputs at time t, and must instead be related to the inputs at time $(t - n)$, which are more likely to be those that actually influenced them (see Guerrero et al. 2015).

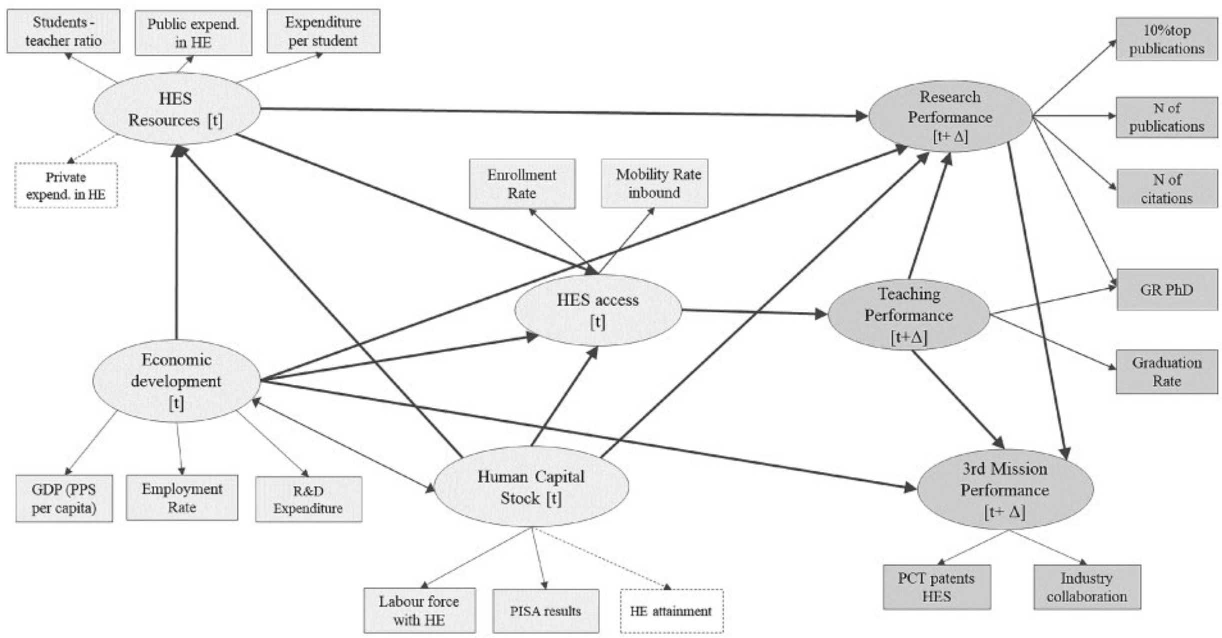

Figure 2. Path diagram: Conceptual framework, including the indicator of the latent factors/*Note*: The figure shows the path diagram of analysis, showing the indicators (the boxes) selected for measuring the latent factors (the circles) of the conceptual framework (Fig. 1). According to path diagram conventions, the paths (the lines) represent the direct effects of a variable on another. The double arrow represents a correlation effect between two variables. In the figure, the paths composing the structural model (relations between latent variables) are shown by the thick lines; and the paths composing the measurement models are shown by the thin lines. The indicators represented by dotted lines are the variables omitted due to problems of missing data. The figure also indicates the time lag between determinants (period *t*) and performance (period $t+\Delta$). *Source*: Authors.

According to this strategy, we have extracted the data for the selected variables in the data set for 4 years: 2000, 2005, 2010, and 2014.[4] Based on this cross-sectional data, we have developed and tested the following three models with a time lag of approximately 5 years: (M1) composed of the determinants for the year 2000 and the performance in 2005; (M2) composed of the determinants for 2005 and the performance in 2010; (M3) composed of the determinants for 2010 and the performance in 2014.

During the preliminary phase of the empirical analysis, the selection of indicators presented here has been further refined to meet the requirements of the empirical approach (see Section 5.2).

4.3 The path diagram: variable description

This section presents the description of the main indicators used in the empirical analysis (see Section 6). Table 4 shows the variability of the data between countries, giving the means of the values over the 4 years of the analysis.

In terms of the economic indicators, we can see a significant heterogeneity between the twenty-nine countries in the study. Norway is the richest country in terms of GDP, with 41,375 PPS[5] per capita, whereas Bulgaria is in last place with a GDP of 9,625 PPS per capita. The variation in GDP partially reflects data on employment rate, which is highest in Norway with 68.03 per cent of people employed among the adult population. The opposite is true for Italy and Croatia, which have the lowest employment rates, at 43.80 per cent and 43.95 per cent, respectively. We have included the results of PISA standardised tests for 15-year-old students[6] among the indicators of HC stock in order to observe the average level of pre-tertiary school performance for pupils in a given country. The values are different from one country to another, ranging from 426.35 in Romania to 539.78 in Finland (scale of values with a mean of 500 and a standard deviation of 100 points). The HC stock is also expressed as the percentage of labour force with tertiary education. In this case, the data shows a high level of heterogeneity, varying from 13.57 per cent in Romania to 36.9 per cent in Belgium. Analogous to the indicator of HES resources, the path diagram includes the indicator of public expenditure on HE, expressed as a percentage of the GDP of the country. However, the information is only partial, since private expenditure has been excluded due to the high level of missing data. By taking into consideration, whenever possible, information on private expenditure, two different situations can be identified. Countries such as Finland and Denmark, which have the highest levels of public expenditure (equal to 1.68 per cent and 1.59 per cent of GDP, respectively), also have very low levels of private expenditure (around 0.07 per cent of GDP). Cyprus and the UK, on the contrary, have very high rates of private expenditure (0.7 per cent and 0.66 per cent of GDP, respectively) but also a low level of public expenditure (0.65 per cent and 0.7 per cent of GDP, respectively). This circumstance can explain the difference in relative values between GDP per capita and public expenditure on HE in the case of the UK.

Higher education enrolment rates (ERs) are particularly low in Cyprus and Malta (with 35.73 per cent and 38.55 per cent, respectively), while they are highest in Finland and Greece, with 89.26 per cent and 87.88 per cent, respectively. Comparing data on HE enrolment numbers with the number of graduates among the population, the capacity for retention clearly varies substantially from one country to another. For example, despite Ireland and Italy having similar enrolment rates (around 60 per cent), Ireland, among all the countries in the study, has the best graduation rate (61.71 per cent) whereas in Italy the rate of graduates is low (26.91 per cent). These values confirm the fact that Italy has one of the highest dropout rates among OECD countries (di Pietro 2004). The indicators of

Table 3. Description of the selected indicators of latent factors.

Latent factor	Variables	Names	Definitions	Data sources	References
Economic development	GDP (PPS per capita)	GDP_pc	Gross domestic product at market prices, at current prices and PPS per capita	Eurostat	Agasisti (2011)
	Employment rate	Employ	Total national employment, divided by number of adult population (20–64-year old)	Eurostat	Agasisti (2011); Hoareau et al. (2013)
	R&D Expenditure	R&D_exp	GERD (Gross domestic expenditure on R&D) as a percentage of GDP. Gross domestic spending on R&D is defined as the total expenditure on R&D carried out by all resident companies, research institutes, universities, government laboratories, etc. in a country	Education at a glance/ OECD	Williams et al. (2013)
Human capital (HC) stock	Labour force with tertiary education	L_forceHE	Labour force with tertiary education indicates the share of total labour force that attained or completed tertiary education (the highest level of education). The labour force includes people who are currently employed and people who are unemployed but seeking work, as well as first-time job-seekers	World Bank	Obadić and Aristovnik (2011)
	PISA results	PISA	The variable is the mean value of PISA results at country level for science, maths, and reading literacy. Being a triennial initiative (last held in 2015), we used the 2015 results for the values of 2014, the results of 2009 for the values of 2010, those of 2006 for 2005, and those of 2000 for 2000	World Bank	St. Aubyn et al. (2009)
HES access	Enrolment rate	ER	Total enrolment in tertiary education (ISCED 5 to 8), regardless of age, expressed as a percentage of the total population in the 5-year age group following on from secondary school	World Bank	Agasisti (2011); Hoareau et al. (2013); Obadić and Aristovnik (2011); Williams et al. (2013);
	Mobility rate inbound	MR_in	Number of students from abroad studying in a given country, the indicator is expressed as a percentage of the total tertiary enrolment in that country	World Bank	Agasisti (2011); Williams et al. (2013)
HES resource	Expenditure per student (% GDP)	exp_stud	Average total (current, capital and transfers) general government expenditure per student in HE, expressed as a percentage of GDP per capita Divide total government expenditure on HE by total enrolment in that the same level, divide again by GDP per capita, and multiply by 100	World Bank	Agasisti (2011); Obadić and Aristovnik (2011); Williams et al. (2013)
	Public expenditure on HE	pub_exp	Public spending on tertiary education as a % of GDP	World Bank	Agasisti (2011); St. Aubyn et al. (2009); Williams et al. (2013)
	Student–teacher ratio	ST_ratio	Average number of pupils per teacher at tertiary education (based on headcounts). When feasible, the number of part-time teachers is converted to ‘full-time equivalent’ teachers; a double-shift teacher is counted twice, etc.	World Bank	Agasisti (2011); Triventi (2014)
Research performance	No. of publications	Publications	Number of publications by country in the Scopus database, divided by the total population (World Bank data), expressed in thousands of inhabitants	SJR – Scopus	St. Aubyn et al. (2009); Williams et al. (2013)
	No. of citations	Citations	Number of citations by country in the Scopus database, for SJR (SCImago Journal & country rank)	SJR – Scopus	St. Aubyn et al. (2009); Hien (2010)
	10% Top publications	10% top	Percentage of publications in the top 10% based on citations by category, year, and document type. The data refer to documents published in Web of Science database	InCites	Hardeman and Van Roy (2013); Hoareau et al. (2013)

(continued)

Table 3. (continued)

Latent factor	Variables	Names	Definitions	Data sources	References
Research/teaching performance	Graduation rate PhD	GR_phd	The share of the total number of PhD graduates per 100,000 inhabitants. PhD graduates (ISCED 8) are referred to the reference academic year, in public and private tertiary education institutions	World Bank	
	Graduation rate (GR)	GR	Total number of students successfully completing tertiary education programmes (ISCED 5 to 8) in public and private tertiary education institutions per 100,000 inhabitants. The number is calculated as the sum of graduates in the reference year and the preceding 4 years	World Bank	Agasisti (2011); Hoareau et al. (2013); Triventi (2014)
Third-mission performance	PCT patents HES	PCTpat_HE	Number of PCT patent applications by universities, divided by the country's population in millions of people (World Bank data). PCT (Patent Cooperation Treaty) applications identify the applications with international patent law. The indicator is built as aggregation of the number of applications in the reference year and the number of applications in the respective following two years (e.g. the value of the indicators in 2005 is the sum of the number of applications in 2005, 2006 and 2007, divided by the country s population in the reference year).	PATSTAT	de Rassenfosse and Williams (2015); Hardeman and Van Roy (2013)
	Industry collaboration	industry_col	Percentage of publications that have co-authors from the industry. The data refer to documents published in Web of Science database.	InCites	
Omitted variable (due to the high number of missing data)					
HC stock	HE attainment	HE_attain	Share of the population successfully completing tertiary education, over the population of over 24-year old	World Bank	Williams et al. (2013)
HES resource	Private expenditure in HE	private_exp	Expenditure on tertiary educational institutions from private funds, expressed as a percentage of GDP	Education at a glance and Eurostat*	Agasisti (2011); Triventi (2014)

Note: The table describes the indicators selected for measuring the latent factors of the conceptual framework. The first column contains the names of the latent factors; the second column contains the names of the indicators; the third column shows the names of the variables in the data set; and the last three columns give the definitions of the variables and their relative sources.

Source: Authors.

research performance correspond to the number of citations in Scopus and the proportion of publications in the top 10 per cent of those being referred to. Both indicators show a similar pattern, with high values for Switzerland, Denmark, and The Netherlands, and lower values for Romania, Bulgaria, and Croatia. The variability among different countries is high, with a difference between minimum and maximum values of around 18 percentage points for citations and 11 percentage points for publications in the top 10 per cent. Lastly, third-mission performance is measured by two indicators, the Patent Cooperation Treaty (PCT) patents applications from HE institutions and industry collaboration. As said, the scarcity of data availability on third-mission activities has limited the choice among possible indicators. Nevertheless, the choice of these specific indicators well represents the university–industry connection. Patent applications from universities provide a measure of HE innovations having a potential commercial value. Since no information is available on the actual use of patents or on the quality of the invention, we take into account only PCT patent applications, which are considered inventions with high market potential (as widely accepted in literature, see de Rassenfosse et al. 2014; van Zeebroeck and van Pottelsberghe de la Potterie 2011). Table 4 depicts a situation that is heterogeneous among countries, going from a very limited application of PCT patents (such as in Romania and Bulgaria) to a maximum rate that exceeds 50 applications per million people, in the case of Switzerland. *Industry collaboration* is measured by the percentage of publications with co-authors from industry in the Web of Science database. Switzerland is also the country with the highest industry collaboration, with an average value of 6.58 per cent of papers co-authored with the industry sector, whereas the lowest values are in Romania, Slovakia, and Lithuania.

The variability in data can also be represented in terms of variation over the 4-year period of the analysis. These statistics are given for the main indicators used in the empirical analysis given in Figs 3, 4, and 5.

Figure 3 shows the variations in terms of GDP per capita and of public expenditure on HE. The GDP per capita has increased almost constantly in the twenty-nine countries studied during the past 15 years. Public expenditure on HE reveals a more heterogeneous situation. Countries reporting low levels of public expenditure in 2000, such as Greece, Malta, and the UK, have greatly increased their public funding over the last 15 years. On the other hand, countries with generally high rates of expenditure, such as Denmark, Finland, and Sweden, have maintained similar values over this period.

Figure 4 shows that there has been a general increase in the data on enrolment numbers and graduation rates from 2000 to 2014.

Table 4. Descriptive statistics for the variables used in the empirical analysis—average values over the years.

Country	GDP_pc	Employ	PISA	L_forceHE	ER	pub_exp	GR	citations	10% top	PCTpat_HE	industry_col
Austria	30,750	56.53	495.24	21.03	63.27	1.32	25.03	18.29	13.13	19.63	3.61
Belgium	29,125	49.05	507.48	36.9	65	1.15	39.97	20.11	14.52	39.08	5.62
Bulgaria	9,625	45.65	431.11	24.93	54.39	0.5	32.97	8.66	6.5	0.14	1.57
Croatia	13,400	43.95	476.24	19.93	50.34	0.67	26.67	7.51	6.36	0.92	1.27
Cyprus	22,525	57.7	437.53	34.93	38.55	0.65	19.28	13.93	11.49	1.47	1.21
Czech Republic	19,375	54.85	495.86	16.33	51.72	0.74	32.81	10.39	8.81	5.79	1.66
Denmark	30,525	60.7	500.47	29.58	73.2	1.59	42.78	23.92	16.3	45.20	5.72
Estonia	14,900	53.03	517.83	34.8	65.89	0.95	35.74	15.53	11.04	13.74	2.06
Finland	27,675	55.13	539.78	35.7	89.26	1.68	39.62	19.75	13.2	14.68	3.24
France	26,500	50.85	498.21	30.03	57.85	1.07	47.47	16.89	12.4	21.69	4.42
Germany	29,175	54.35	502.52	25.38	56.57	0.96	19.59	17.3	12.73	18.54	4.00
Greece	19,925	44.7	464.01	24.95	87.88	1.07	23.55	13.67	9.98	0.82	1.85
Hungary	15,000	46.78	487.62	21.08	53.65	0.8	30.83	12.82	8.86	3.08	3.08
Ireland	32,875	55.3	507.34	33.83	59.18	1.01	61.73	17.03	11.22	40.84	2.48
Italy	25,550	43.8	478.38	15.58	60.77	0.64	26.91	16.75	11.86	6.59	2.69
Latvia	12,425	50	479.7	25.3	68.25	0.71	46.01	9.88	6.89	4.36	1.04
Lithuania	13,950	50.4	478.57	36.7	71.21	1.09	53.7	9.2	7.16	2.91	1.13
Malta	20,250	47.6	459.39	16.73	35.73	0.97	31.16	10.64	8.62	5.41	1.27
The Netherlands	32,375	61.35	515.83	29.73	64.1	1.08	33.61	23.14	15.97	22.51	4.63
Norway	41,375	68.03	498.31	35.25	74.39	1.38	40.28	18.99	13.81	15.56	5.15
Poland	13,900	48.48	495.67	22.03	64.46	1.05	58.26	9.19	7.29	2.42	1.44
Portugal	19,450	55.23	479.66	15.48	58.71	0.84	33.03	14.08	10.98	9.06	1.53
Romania	10,475	53.05	426.35	13.58	47.64	0.75	39.95	6.95	5.96	0.05	0.95
Slovakia	16,075	50.05	477.76	16.08	44.66	0.7	41.92	8	6.45	1.48	1.17
Slovenia	20,025	53.95	504.67	23.38	76.42	0.94	36.99	10.47	8.59	5.36	1.35
Spain	22,875	47.43	484.64	32.28	73.21	0.89	34.02	15.39	10.98	17.10	2.38
Sweden	30,100	67.83	502.15	31.63	71.53	1.38	29.55	21.61	14.31	1.01	5.44
Switzerland	37,625	64.5	510.79	30.35	48.36	1.15	43.03	24.62	16.91	50.27	6.58
UK	26,850	58.53	500.59	32.83	58.27	0.7	51.03	19.85	13.23	25.69	3.14
No. of observations	116	116	104	116	114	109	116	116	116	116	116

Note: This table shows the average values for the 4 years of analysis (2000, 2005, 2010, and 2014). The statistics are reported for the 29 countries studied. The last row indicates the number of observed data items for each variable (maximum of observations per column = 116).

Source: Authors.

Regarding the values for enrolment rates, Greece saw the largest increase over time, with a difference of around the 58 percentage points between 2000 and 2014. The situation is different in the UK, which recorded stable values over the 15 years under analysis. In terms of graduates, Greece does not conform to the general increased trend in enrolment, as the rates here are low and have only increased slightly over the years. The country with the largest increase over time is Poland, with a difference of about 60 percentage points.

Lastly, Fig. 5 reveals how the research quality has increased significantly in countries with initial lower results. Differently, the statistics on PCT patent applications by universities show a communal situation for all countries in 2000, with low numbers of applications. However, some countries, such as Denmark, Switzerland, Ireland, and Belgium, have largely improved over time, achieving a substantial number of applications in 2014.

The results discussed above describe a situation that varies for the different European HESs. While, on the one hand, there was a high variability among the indicators over time, thereby justifying the importance of analysing the evolution of the models over the last 15 years, on the other hand, the descriptive statistics shed light on the heterogeneity typical of European HES, a fact that must be considered during the analysis of the empirical results.

4.4 Institutional diversity of higher education systems

In the previous section, we have described the countries of the sample by means of statistics on the variables used in the empirical analysis. Nevertheless, the HESs we analyse are also characterised by a structural (institutional) diversity, namely the variety of services or programmes offered by different systems of HE (Santoalha et al. 2018). Therefore, as a primary and major factor of structural diversity, we have examined the distinction between binary and unitary systems.

Table 5 reports the rate of students enrolled in HE institutions that are specifically focused on professional education and offer vocational programmes. The indicator is built on the data available on the European Tertiary Education Register (ETER) database, in which these types of institutions are classified as 'University of Applied Sciences/college' (UAS). The indicator has been reported for each available year on ETER database, corresponding to the period from 2011 to 2015. The time trend shows stable values over time, suggesting that no structural change has occurred during the specific time span. Ten countries of the sample show a null value in all years, indicating that their HESs do not include any vocational program. On the contrary, the Netherlands, Belgium, and Norway are the countries with the highest rate of students enrolled in UAS institutions, with a rate over 50 per cent.

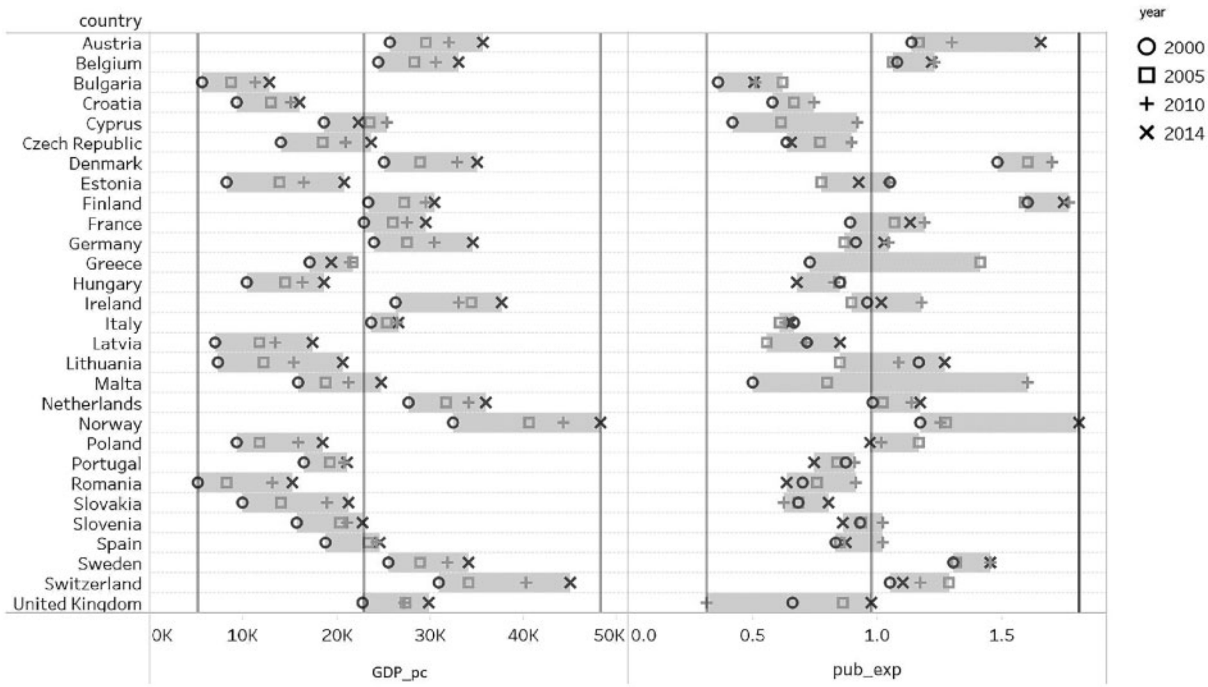

Figure 3. GDP (PPS per capita) and public expenditure on HE, per country and year. *Note*: The figure displays the values of GDP (PPS per capita) and public expenditure on HE (see the definitions in Table 3) for each country among those studied. The chart shows the values for the 4 years selected using indicators marked in different colours and shapes. The missing values are omitted; the number of observations for each variable is given in Table 4. *Source*: Authors.

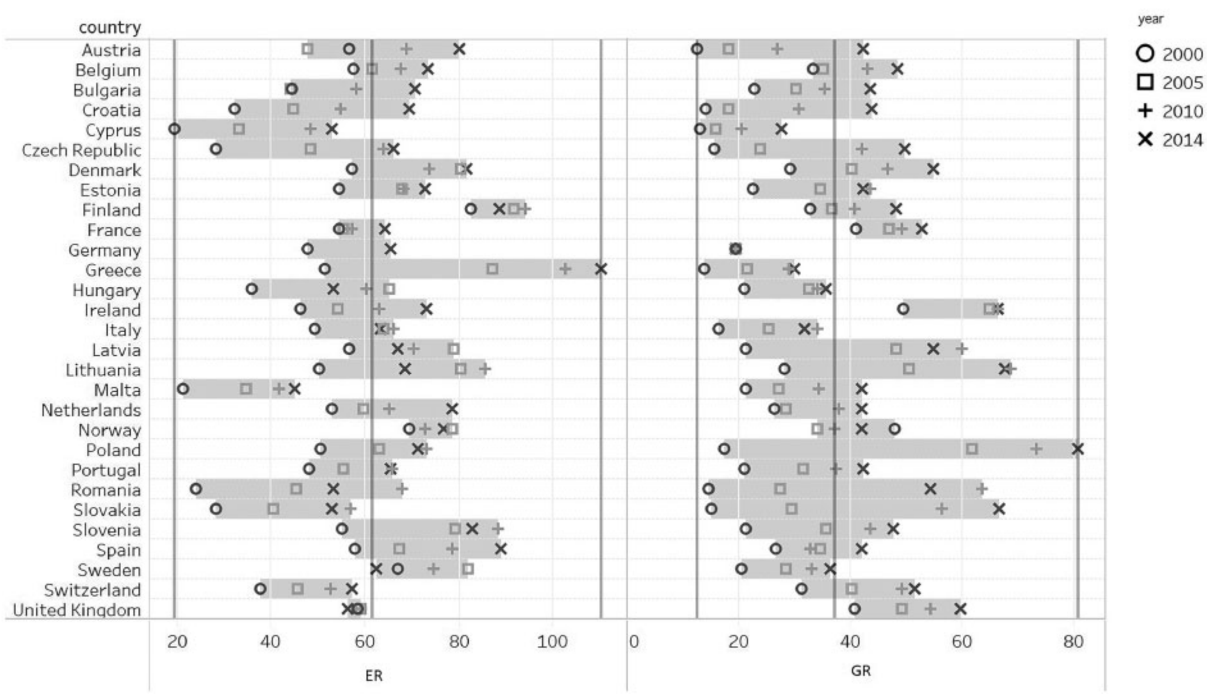

Figure 4. Enrolment and graduation rates per country and year. *Note*: The figure displays the values of enrolment rates and graduation rates (see definitions in Table 3) for each country among those studied. The charts show the values for the 4 years selected using indicators of different colours and shapes. Missing values are omitted; the number of observations for each variable is given in Table 4. *Source*: Authors.

Despite there are few studies analysing empirically the effect of structural diversity, the evidences in literature highlight that this feature could matter in terms of empirical results (Santoalha et al. 2018). For this reason, in Section 6.2 we have performed a robustness check analysis to understand whether HE with binary or unitary systems has different levels of performance and different factors determining it. To this end, in the last column of Table 5, the countries of the sample have been divided in two groups: (1)

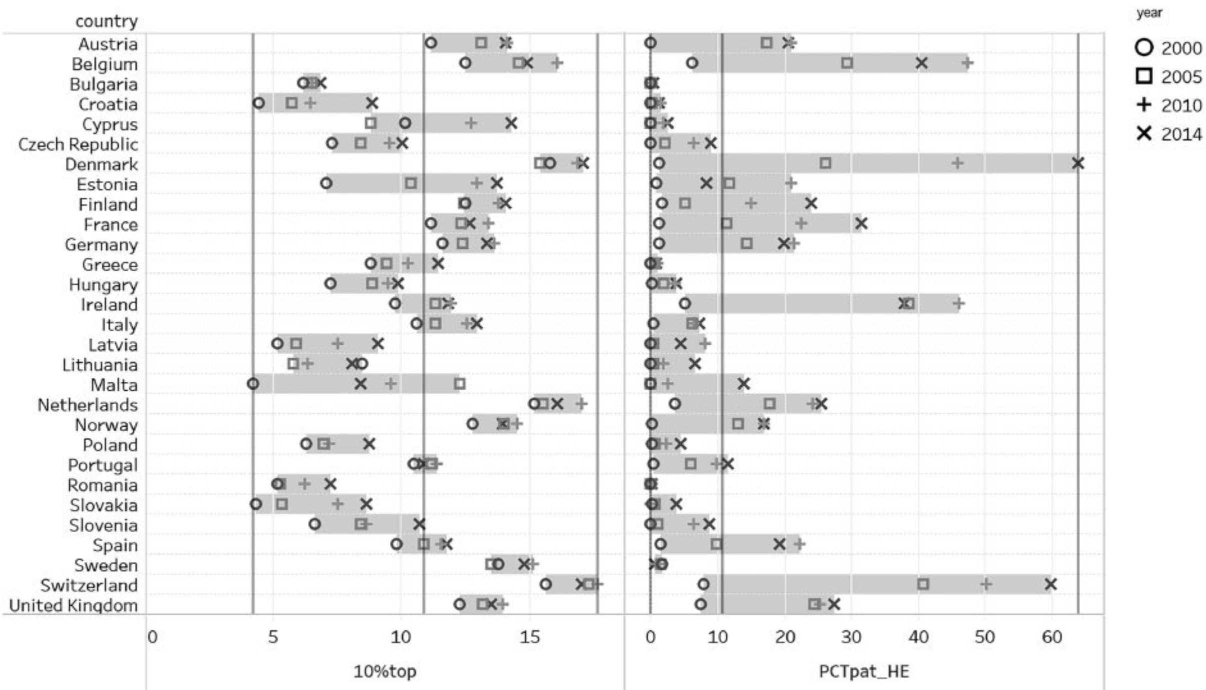

Figure 5. Research quality and PCT patent applications from HE, per country and year. *Note*: The figure displays the values of 10 per cent top publications and number of PCT patents HES (see definition in Table 3) for each country among those studied. The charts show the values for the 4 years selected using indicators of different colours and shapes. The number of observations for each variable is given in Table 4. *Source*: Authors.

the binary systems, which have an average rate of students enrolled in a UAS higher that 15 per cent; and (2) the unitary systems, which have an average rate of students enrolled in a UAS lower than 15 per cent.

We are aware that this unitary/binary classification, based on the number of students in UAS, is not able to capture the singularity of some HESs. This is the case with France, which despite it does not include institutions identified as UAS is not easily classifiable in one category (see Kyvik 2004). In fact, French HES includes, beside the traditional universities, different types of HE institutions, such as 'Grandes Ecoles'—which provide elite education—or the 'Sections de techniciens supérieurs'—which offer a technical post-diploma education. For this reason, French HES has been defined in literature as a 'fragmented' system (Jallade 1992). From an empirical point of view, this issue has been addressed by performing a robustness check analysis, discussed in Section 6.2.

Lastly, HE systems are structurally different also in terms of the location of the research activity. In particular, both France and Germany have a significant part of the research activity that is external to the universities and located in research centres. The diversity in the organization of research activities may be relevant also in relation to the distinction between unitary and binary HESs. For this reason, a robustness check analysis dealing with this issue has been presented in Section 6.2.

5. Methodology: SEM

5.1 The SEM approach

The relationships between HES performance and their respective determinants were estimated through SEM. The technique is particularly suited for the purpose of research for three reasons. First, it is a statistical technique that can model latent variables in an explicit setting (Kline 2015). Relying on this property, we have estimated the relationships between the factors of the conceptual framework, although they cannot be observed directly. Second, SEM can be used to model networks of effects occurring between the system's components, and this property is necessary in terms of carrying out the testing of the conceptual framework, since some factors are both the dependent and the independent variables. Moreover, the method allows the impact of the various determinants on all the possible outputs to be tested simultaneously. Therefore, this 'structural' property makes it possible to test the functional and simultaneous relationships between the dependent variables. As shown in the following equation, responses (y-variables) can depend on other responses (other y-variables):

$$Y_1 = f(X_i, \ldots, X_n; Y_k, \ldots, Y_m). \quad (1)$$

Because of this feature, SEM is very flexible and can lead to complex hypotheses being set out. Third, SEM is a powerful method for representing complex systems and, in particular, it can be used to describe the multi-dimensionality of HESs. Indeed, SEM is based on path diagrams, which provide a simple visual representation of models and empirical estimates.

The path diagram used in the SEM analysis is also the representation of the set of equations that model the assumptions expressed in the theoretical framework. The equational form of SEM comprises two main models: the *measurement model* and the *structural model*. The former represents the relationships between the latent variable and its reflective indicators, that is, the observable variables measuring it. According to Jöreskog and Sörbom (1993), the *measurement model* can be written in terms of two matrix equations:

$$Y = \Lambda_y \eta + \varepsilon, (2), \text{ for the latent dependent variables,}$$
$$X = \Lambda_x \xi + \delta, (3), \text{ for the latent independent variables,}$$

where the vector Y $(p \times 1)$ and the vector X $(q \times 1)$ denote the observed variables. The vector Y measures the latent dependent

Table 5. Description of institutional diversity between countries.

	% students UAS 2011	% students UAS 2012	% students UAS 2013	% students UAS 2014	% students UAS 2015	Average	Binary System
Austria	12.71	13.04	13.73	14.27	14.82	13.44	0
Belgium	58.98	58.84	49.46	49.90	50.35	54.30	1
Bulgaria	2.64	2.50	2.65	2.81	2.65	2.65	0
Croatia	12.39	12.31	14.54	14.26	13.70	13.37	0
Cyprus	20.84	17.07	12.65	11.99	16.01	15.64	1
Czech Republic	0.00	0.00	0.00	0.00	0.00	0.00	0
Denmark	39.86	43.83	43.54	0.00	39.86	31.81	1
Estonia	25.66	25.02	24.57	20.66	21.28	23.98	1
Finland	48.45	48.51	48.60	49.09	50.33	48.66	1
France	0.00	0.00	0.00	0.00	0.00	0.00	0
Germany	32.29	33.00	33.28	34.47	34.86	33.26	1
Greece	33.00	34.64	34.33	32.68	n.a.	33.66	1
Hungary	25.77	n.a.	24.28	20.53	n.a.	23.53	1
Ireland	42.44	42.06	39.65	39.69	39.25	40.96	1
Italy	0.00	0.00	0.00	0.00	0.00	0.00	0
Latvia	10.90	11.49	11.67	11.37	9.98	11.35	0
Lithuania	27.77	28.26	29.49	29.93	27.10	28.86	1
Malta	0.00	0.00	0.00	0.00	0.00	0.00	0
The Netherlands	63.01	62.41	63.15	63.32	62.01	62.97	1
Norway	50.80	51.06	50.73	49.92	49.76	50.63	1
Poland	0.00	0.00	0.00	0.00	0.00	0.00	0
Portugal	34.32	33.15	45.18	45.52	46.91	39.55	1
Romania	0.00	0.00	0.00	0.00	0.00	0.00	0
Slovakia	8.70	8.07	7.40	6.95	6.07	7.78	0
Slovenia	0.00	0.00	0.00	0.00	0.00	0.00	0
Spain	0.00	0.00	0.00	0.00	0.00	0.00	0
Sweden	0.00	0.00	0.00	0.00	0.00	0.00	0
Switzerland	35.76	36.33	36.47	36.72	37.33	36.32	1
UK	0.00	0.00	0.00	0.00	0.00	0.00	0

Note: The table displays the rate of students enrolled in UAS institutions between 2011 and 2015 together with the average value over the time span. In the last column, the table reports the classification of HES which is equal to 1 if the HES is considered binary (average rate >15%), and it is equal to 0 if the HES is considered unitary (average rate <15%).

Source: Authors' elaboration of ETER data.

variables h ($m \times 1$), and the vector X measures the latent independent variables x ($n \times 1$). The relationships between the latent variables and the observed variables are identified by the ($p \times m$) matrices Λy and Λx. The measurement errors are the $p \times 1$ vector ε and the $q \times 1$ vector δ.

The other SEM model, the *structural model*, describes relations between the latent variables and it can be expressed in terms of the following matrix equation:

$$\eta = B\eta + \Gamma\xi + \zeta, \quad (4)$$

where η is the vector ($m \times 1$) that denotes the latent dependent variables; and ξ is the vector ($n \times 1$) representing the latent independent variables. B is a $m \times m$ matrix of structure coefficients connecting the latent dependent variables to one another. Γ denotes an $m \times n$ matrix of structure coefficients whereas the latent independent variables relate to the latent dependent variables. The error term ζ of Equation (4) is a vector that contains the equation prediction errors or its disturbance terms.

From Equations (2), (3), and (4), it is possible to generate the final matrix of covariance terms implied by the overall model (i.e. the matrix Σ). The empirical estimation of the coefficients is achieved by comparing the sample covariance matrix S with the estimated covariance matrix Σ of theoretical model. The process of fitting continues until the two covariance matrices are close enough that the differences might reasonably be attributed to a sampling error. Several fitting functions or estimation procedures are available; in the article, we used the Maximum Likelihood (ML) estimation,[7] which is the most common fitting process.

Adopting the maximum likelihood estimation, the fit model is the product ($N-1$) F_{ML}, where F_{ML} is the value of the fit function minimised in the ML estimation and ($N-1$) is one less than the sample size. The product ($N-1$) F_{ML} follows a central chi-square distribution. This statistic is the model chi-square, χ^2_{ML}, also known as the likelihood ratio chi-square or the generalised likelihood ratio.

This approach provides a measure of the goodness of fit based on a comparison of the model implied covariance matrix Σ with the sample covariance matrix S (Lomax and Schumacker 2004). When the estimation method is the maximum likelihood, the basic fit test is the likelihood ratio chi-square (LRχ^2). The LR chi-square tests the difference between a given restricted model $L(\hat{\theta}^R)$, referred to as an estimated covariance matrix, and whatever unrestricted model $L(\hat{\theta}^U)$ would imply a covariance matrix that perfectly corresponds to the data covariance matrix (Equation (5)),

$$\mathrm{LR}\chi^2 = 2\ln\left(\frac{L(\hat{\theta}^U)}{L(\hat{\theta}^R)}\right) \sim \chi^2_M \quad (5)$$

$L(\hat{\theta})$ describes the marginal likelihood for the vector of estimated parameters ($\hat{\theta}$) of the model. The test is based on the null hypothesis that the LRχ^2 is equal to zero, so when there is a good fit, the ratio between $L(\hat{\theta}^U)$ and $L(\hat{\theta}^R)$ is very close to 1, and consequently the logarithmic becomes similar to zero and the null hypothesis cannot be rejected. Therefore, a large *p*-value (>0.05) indicates that the model is consistent with the covariance data.

5.2 The methodological approach for the empirical analysis

The methodology adopted to conduct the empirical analysis has followed several phases. First, based on Kline (2015), an explorative analysis was conducted to test the validity of the hypotheses expressed through the measurement model, which describes the relationships between the latent factors and their reflective indicators. The explorative analysis was carried out through two steps, where first we computed two correlation matrixes, and then carried out two confirmative factor analyses (CFA). Pearson correlations were calculated in the first step to identify the level of correlation between the reflective indicators. According to the SEM approach, the measurement model is adequately defined if the latent variables can capture most of the correlation between their reflective indicators. Therefore, through the first step, it was possible to exclude the indicators that were not significantly correlated with the other indicators belonging to the same latent factor.

The confirmative factor analyses performed in the second step of the explorative analysis are specifically designed to individually test the measurement model of the overall model. The CFA technique analyses *a priori* the relationships between latent factors and the respective indicators, explicitly specifying the number of latent factors (Kline 2015). As a result of first and second step, we obtained a selection of indicators. We, therefore, identified the final path diagram, where the results of the explorative analysis are applied to the theoretical assumptions of theoretical model. As a third step, the final path diagram has been analysed through SEM, meaning that the relationships involving the measurement and structural models have been tested simultaneously.

Finally, we have tested the robustness of the model by analysing the main indicators of the fitting. In particular, the null hypothesis of the LR chi-square test cannot be rejected in all the three models (*p*-value > 0.05), although small P-values are generally admissible (Jöreskog and Sörbom 1993). The fit measures are particularly important for the purposes of this article. Our objective was to obtain a high fit to the data, since the empirical analysis was not the aim of finding or predicting causal relationships but rather to describe the relationships between the factors of the HESs under analysis. According to this intention, a point that must be taken into consideration is that the HES sample analysed in the empirical analysis exactly corresponds to the entire population studied (i.e. all HESs in Europe). Based on this, we can also justify the use of a sample size that is small when compared to the usual number of observations employed in SEM analyses.[8]

6. Results of the empirical analysis

6.1 Structural equation modelling results

The results of the explorative analysis (see Supplementary Annex, Section A3 for details) have led to a simplified path diagram displayed in Fig. 6. In particular, following the procedure described in the previous section, we excluded the *student–teacher ratio*, *expenditure per student*, *R&D expenditure*, and *mobility rate inbound* from the indicators of determinant factors; and the *number of publications* and *PhD graduation rate* from the indicators of performance factors. Although this reduced complexity has cut the number of factors tested in the model, in the authors' opinion, the path model is still of an adequate level of completeness in terms of describing the relationships between the performance of HESs and its determinants. In fact, the results of the explorative analysis allow to estimate a path diagram that is built exclusively on the indicators that better measure the latent factors of our theoretical framework. In this way, the robustness of the estimates will be improved.

The path diagram in Fig. 6 has been tested using SEM in SPSS AMOS 24.0. The results of the SEM estimates, expressed in

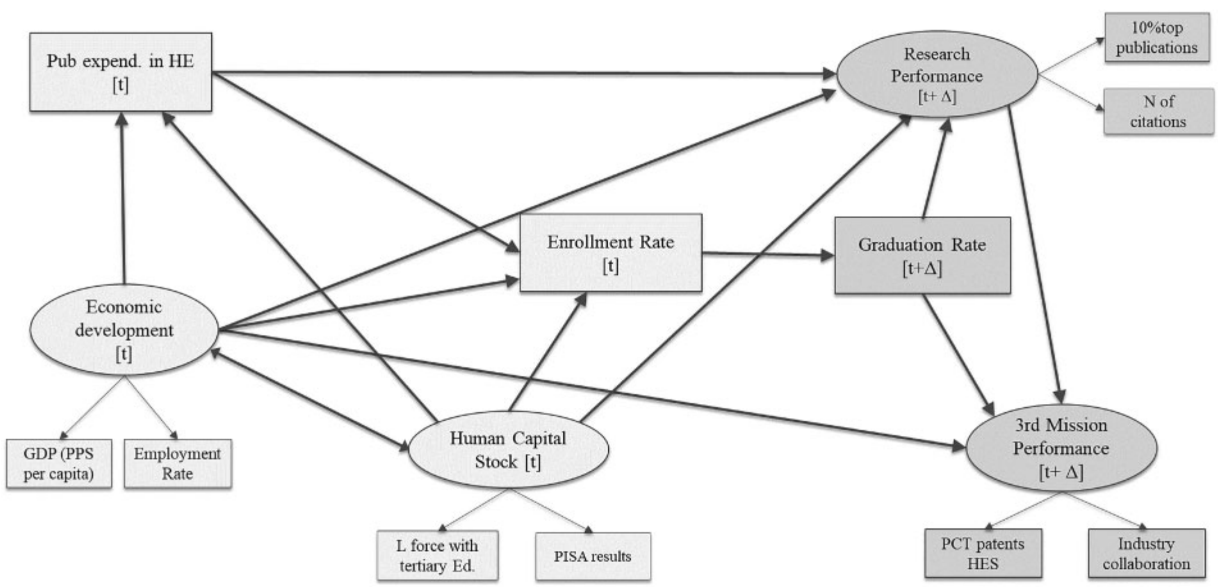

Figure 6. Path diagram for SEM estimation – Final model. *Note*: The figure displays the path diagram estimated through SEM. This path diagram is the result of the explorative investigation on the complete path diagram of Fig. 2. The boxes represent the observable variables and the circles indicate the latent factors. According to path diagram conventions, the paths (the lines) represent the direct effects of a variable on another. The double arrow represents a correlation effect between two variables. The figure also indicates the time lag between the determinants (period *t*) and performance (period $t+\Delta$). *Source*: Authors.

Table 6. SEM estimates for Models 1, 2, and 3.

			M1 (2000–2005)			M2 (2005–2010)			M3 (2010–2014)		
			Standard Estimate	SE	P	Standard Estimate	SE	P	Standard Estimate	SE	P
pub_exp	←	Economic development	0.558	0.013	**	0.639	0.012	***	0.508	0.017	**
Research Perf.	←	Economic development	0.973	0.357	***	0.965	0.168	***	0.975	0.044	***
ER	←	pub_exp	0.724	6.532	***	0.555	8.724	***	0.270	7.298	NS
Employ	←	Economic development	0.610			0.707			0.734		
GDP_pc	←	Economic development	0.924	397.55	***	0.901	340.93	***	0.864	351.25	***
L_forceHE	←	HC stock	0.536	0.072	**	0.672	0.067	***	0.583	0.094	**
PISA	←	HC stock	0.828			0.704			0.669		
GR	←	ER	0.250	0.003	NS						
10% top	←	Research perf.	0.983	0.028	***	0.979	0.043	***	0.966	0.316	***
Citations	←	Research perf.	0.975	0.004	***	0.989			0.960		
Third-mission perf.	←	Research perf.	0.995	0.012	***	0.977	0.22	***	0.961	0.151	***
Third-mission perf.	←	GR				0.213	0.005	**	0.202	0.006	*
industry_col	←	Third-mission perf.	0.915			0.899			0.885		
PCTpat_HE	←	Third-mission perf.	0.751	0.161	***	0.838	0.148	***	0.838	0.159	***
Economic development	<->	HC stock	0.761	35.462	*	0.943	38.179	*	0.951	28.308	*
e(employ)	<->	e (GDP_pc)							0.544	5647.8	*
e(GR)	<->	HC stock	0.543	1.082	**	0.300	52.082	†			
P-level	>0.05 good fit		0.102			0.121			0.080		
Chi-square			51.677			51.740			53.108		
CMIN/DF	<3 good fit (Kline 2015)		1.292			0.958			1.328		
RMSEA	<0.05 good fit (MacCallum et al., 1996)		0.102			0.097			0.108		
CFI	>0.95 good and >0.90 acceptable		0.955			0.958			0.936		

Notes: The above table gives the results of SEM estimations using SPSS AMOS 24. The coefficients for standardised direct effects are also displayed. The standardised coefficients are in deviation standard units and are independent of the units in which all variables are measured; therefore, they are not affected by the choice of identification constraints. The estimates are reported for the three time-moments of analysis (Models 1, 2, and 3).

†P ≤ 0.10; *P ≤ 0.05; **P ≤ 0.01; ***P ≤ 0.001.

NS: Non-significant

Source: Authors.

standardised coefficients, are summarised for the three models (M1, M2, and M3) in Table 6 and reported through the relative path diagrams (Fig. 7).

Taking into consideration the first research question, the empirical analysis provides interesting findings about the relationships between the HES performance and their respective determinants, that is, the latent factors of the framework. In terms of size, the most significant relationships are (i) the effect of economic development on research performance, and (ii) the effect of research performance on the third-mission performance. Consequently, economic development also has a high indirect effect on the third-mission performance; the standardised coefficients are 0.969, 0.943, and 0.937 in M1, M2, and M3, respectively. Moreover, the effect of economic development on research performance is not mediated (see Baron and Kenny 1986) by the public expenditure on HE, as we have hypothesised in the conceptual framework. In fact, although economic development significantly influences public expenditure, public expenditure has no significant effect on research performance.

Confirming the hypothesis of theoretical framework, the results report a significant and high correlation between HC stock and economic development; moreover, they show a significant direct effect of research performance on the third-mission performance. However, on comparing the empirical results with the conceptual framework, one of the most relevant findings is the absence of statistically significant relationships between graduation rates and research performance and, partially, between graduation rates and third mission performance.

In order to address the second research question, we compared the results for different models and substantial dissimilarities have come to light. First, the effect of public expenditure on enrolment rates, highly significant in the first two models, is no more significant in Model 3. The second difference is regarding the effect of graduation rates on the third-mission performance. In Model 1, this effect is not statistically significant whereas it is in Model 2 and 3, even though with low coefficients and poor levels of significance. Lastly, as shown in Table 6, the values of the coefficients related to stable relationships do not vary largely from model to model.

The main measures for goodness-of-fit associated to each model are reported in the bottom part of Table 6. As said previously, the indicators reveal a good fit for each model. In fact, the empirical results show that the P-values of the LR chi-square tests are large (>0.05), and only the root mean square error of approximation (RMSEA) measures indicate a limited fit, since the indicator penalises high levels of complexity, especially in the case of small sample sizes.

6.2 Robustness check analysis—dealing with institutional diversity

As a robustness check, we empirically tested whether the structural (institutional) diversity between HESs affects the results presented in the previous section. The robustness check analysis is based on the division between binary and unitary systems described in Section

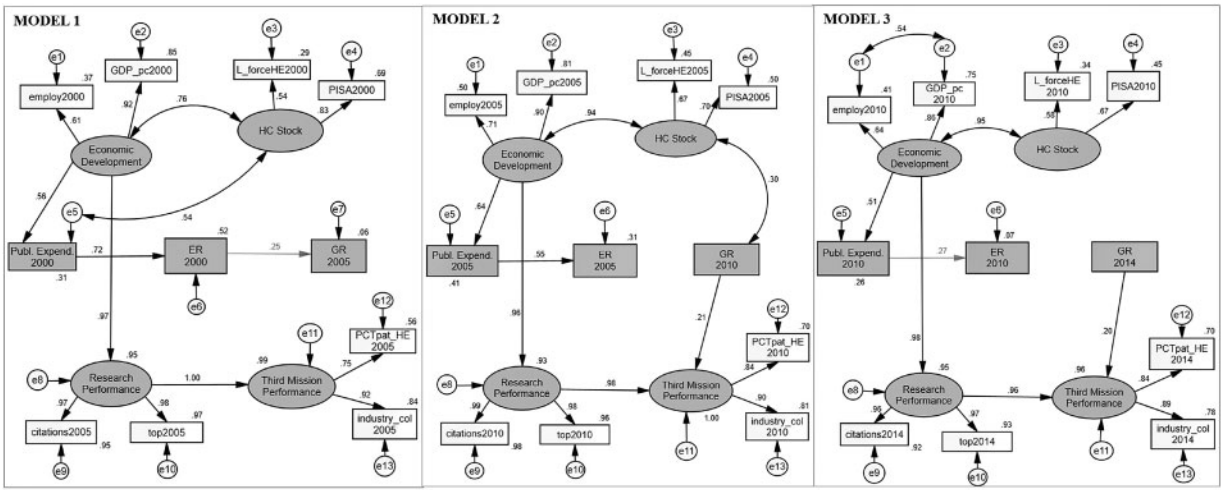

Figure 7. Path diagram for SEM estimation—Final model. *Note*: The figure displays the path diagram estimated through SEM. This path diagram is the result of the explorative investigation on the complete path diagram of Fig. 2. The boxes represent the observable variables and the circles indicate the latent factors. According to path diagram conventions, the paths (the lines) represent the direct effects of a variable on another. The double arrow represents a correlation effect between two variables. The figure also indicates the time lag between the determinants (period *t*) and performance (period $t + \Delta$). *Source*: Authors.

4.4. The last column of Table 5 identifies two groups of HES (binary vs. unitary), which comprise fourteen and fifteen countries, respectively. Based on these two groups, we have conducted a multi-group invariance analysis in SPSS Amos 24 (see Byrne 2004). This approach allows to perform a SEM analysis that estimates separated parameters between binary and unitary systems but performing the tests simultaneously across the two groups. In other words, the multi-group analysis allows to determine whether the two groups are associated to models that are statistically different.

For reasons related to the sample size, we performed a multi-group analysis on a single model that includes data of all cross-sections, that is, 2000 and 2005 for determinants factors, and 2010 and 2014 for performance factors. For this reason, the robustness check analysis wants to be informative only on the differences between binary and unitary systems without allowing to directly compare the results of this model with the ones described in the previous section.

The results are reported in Table 7, which contains the values of chi-square difference test for three nested models. As a first step, we tested differences between the two groups in their measurement models by means of the *measurement weights* model (first line of Table 7), in which the measurement weights of the two groups are constrained to be equal. The result of the test, on the right part of the table, shows no statistical difference between the constrained model (the *measurement weights* model) and the unconstrained one. In fact, the P-value is not statistically significant, meaning that the constrained model has not significant lower fit compared to the unconstrained model ($P > 0.05$). The same test has been performed for *structural weights* and *structural covariances* models that add, compared to the first model, the constraints on structural effects and on structural effects plus structural covariances, respectively. Regarding these two models, the tests show a difference between unitary and binary HESs that is statistically significant. Therefore, in order to provide evidences on which effect causes the structural difference, each single relationship of the structural model has been tested separately. To this end, we constrain one structural relationship at a time, identifying the five models in the bottom part of the table. The results show a statistically significant difference only for the effect of teaching performance on the third-mission performance. In fact, looking at the estimates between the two groups (see Supplementary Annex, Section A4), unitary systems do not show a statistically significant effect of teaching performance on the third-mission performance ($P = 0.812$) whereas for binary systems, the same effect is statistically significant and high ($P = 0.000$, with a standardized weight of 0.546).

These results are also confirmed when we control for the specific cases of structural diversity identified in Section 4.4. In particular, we repeated the multi-group analysis excluding France from the sample of unitary systems, since it is not clearly classifiable in this category (see Section 4.4). Moreover, the same analysis has been carried out excluding both France and Germany, as these are the countries that most are characterised by research activities outside HE institutions. The results of these robustness checks confirm that binary and unitary systems are statistically different because of the effect of teaching performance on the third-mission performance, as we found analysing the sample in which all countries are included (see Supplementary Annex, Section A4).

7. Discussion, policy messages, and concluding remarks

This article analyses the performance of HESs in Europe, exploring some key determinants affecting the performance. The aim of the work arises from the lack of studies evaluating HE performance at system level. With this purpose in mind, we proposed a conceptual framework to describe relationships between performance and its determinants, and tested it by employing SEM for the empirical analysis.

The results lead to four significant observations. The first observation concerns the relationships between the three dimensions of HE performance. While on the one hand, the hypotheses defined in the conceptual framework are confirmed by the significant effect that research performance has on the third-mission performance, on the other hand, the findings show a very limited effect of teaching on the third-mission performance and they do not detect any statistically significant correlation between teaching performance and

Table 7. Results of multi-group analysis.

Model	DF	CMIN	P	NFI Delta-1	IFI Delta-2	RFI rho-1	TLI rho2
Measurement weights	5	7.514	0.185	0.015	0.018	−0.006	−0.008
Structural weights	9	26.942	0.001	0.054	0.064	0.022	0.029
Structural covariances	12	29.909	0.003	0.059	0.071	0.015	0.019
Economic development → Research perf.	1	0.160	0.69	0.000	0.000	−0.005	−0.006
Public expenditure→ ER	1	0.298	0.585	0.001	0.001	−0.004	−0.006
Research perf. → third-mission perf.	1	0.000	0.997	0.000	0.000	−0.005	−0.007
Teaching performance → third-mission perf.	1	15.433	0.000	0.031	0.037	0.036	0.046
Economic development <-> HC stock	1	1.190	0.663	0.000	0.000	−0.005	−0.006

Note: The table displays the results of multi-group analysis. The first line refers to the model in which measurement weights of the two groups are constrained to be equal. The second line refers to the second model nested to the first one, in which also the structural weights of the two groups are constrained to be equal. The third line refers to the third model nested to the second one, in which also the covariance weights of the two groups are constrained to be equal. The last lines are referred to the models on which one structural weigh at the time has been constrained to be equal in both groups.

Source: Authors' elaboration using SPSS Amos 24.

research performance. In other words, HESs with higher (lower) teaching performance are not necessarily those with better (lower) research indicators and they do not necessarily generate better (lower) third-mission outcomes. A potential explanation can be linked to the conflicting effects of teaching activities on research performance, a fact already pointed out in the existent literature exploring teaching vs. research trade-offs at university level: the intensity of teaching can imply that professors have less time for engaging in their research and, on the other hand, students can be involved in joint research, increasing the research productivity (Dundar and Lewis 1998). This relationship could not only apply at institution level but also holds true at system level. Regarding the third-mission performance, research-intensive institutions are normally associated in literature with measures of university–industry links (Abreu et al. 2016), which is indeed the main dimension captured by the factor in the conceptual framework. On the other hand, this could also explain the limited effect of teaching on the third-mission performance. Regarding this specific relationship, the structural diversity of HESs also plays an important role. As showed by the robustness check analysis, the effect of teaching performance on the third-mission performance is statistically significant only for binary systems. This result could suggest that teaching performance in vocational education has a more direct link with the third-mission performance.

The second relevant observation relates to the role that economic development has in terms of affecting the performance of HESs. In our Model 1, economic development affects research performance directly, while it has an indirect influence over third-mission performance. As a point of interest, stronger economic development appears to be an important driver for research-related activities, more so than teaching performance.

Third, the empirical analysis found that HES resources had a moderate effect on HES performance (particularly on research performance) compared to the findings provided in literature (see Hoareau et al. 2013; Williams et al. 2013). In interpreting this result, a point to consider is that the previous studies by Williams et al. (2013) and Hoareau et al. (2013) did not control for the economic factors of the system, for example, GDP or employment rate. Therefore, because of the high effect of economic development on HES resources (see Table 6), these previous works could have overestimated the role of HES resources, ascribing the effects of economic factors instead to HES resources. In the same vein, we must acknowledge that an important limitation of our article is that we have not taken into account the private resources dedicated to HE, recognising that their effect could be significant, especially on research performance.

All these considerations should be interpreted with caution, since teaching performance is expressed through a single indicator, the graduation rate, and the indicators of the third-mission performance present some limitations. In fact, we must be aware that indicators of PCT patent applications and industry collaboration only reflect that a part of third-mission activities is more strictly connected to 'applied' HE research, overlooking the other functions that connect universities to society. Moreover, we are also conscious that the indicators we used for measuring the third-mission performance capture primarily the results of activities in the disciplines of applied sciences. This shortage is also due to the nature of theoretical disciplines and humanities, which are associated with an indirect impact on society, difficult to capture and measure (Molas-Gallart 2002). It follows that the future research should include other indicators for third-mission and teaching performance in order to be able to test more effectively the underlying connection between these two performance factors.

Lastly, our findings highlight substantial changes in the relationships between the performance of HESs and its determinants over the period under study, that is, from 2000 to 2014. The present study does not explore the causes of these changes, but the future research could investigate the effects of policy instruments and HES reforms on the already discovered evolution of HES performance. In fact, taking into account the large time span we are analysing, the political dimension could have an important role in explaining the changes in HESs over time. In other words, it can be the case that reforms intervened during the period under analysis had the effect of changing the relationships between the variables analysed and the performance. For example, these relationships could be affected by the recent reforms that in many European countries were aimed at improving the functional organization of HESs, encouraging the merger of universities (see Kyvik 2004).

This study provides some major contributions for innovating the current state-of-art of literature in the field. We have provided and tested a conceptual framework to represent structural relationships between the performance of HESs (teaching, research, and the third-mission performance) and their respective determinants. This study does not merely contribute to the field's empirical insights but also to theoretical literature, where there is a lack of empirical works that consider teaching, research, and third-mission activities at the same time at system level and not only at institution level. Another contribution of this work is the large cross-country panel data set of European HESs, which covers twenty-nine countries over a 20-year period. Addressing the problem of the deficiency of HE data at country level (Hanushek and Woessmann 2011), the data set would help to generate new empirical studies on HESs. In addition, we introduced a new method to test relationships between variables at system level, using SEM. The SEM approach is a powerful method for understanding and describing complex relationships between latent and exogenous variables and it represents a good alternative to compensate for the shortcomings of other techniques such as input–output analysis (Guerrero et al. 2015). Moreover, we have provided empirical insights into the evolution of the framework and its relationships over the last 15 years, identifying how key PIs have evolved over the period under study, together with their relationship with the corresponding determinants.

The study also presents some limitations. In our empirical work, we have considered that a range of HESs could be very different in terms of their cultural characteristics and organisational structure. As the difference between binary and unitary HESs has been proved to affect the empirical results, other structural diversities could matter also, such as the degree of specialisation between teaching and research activity, different levels of student fees, etc. These conditions could produce some level of bias caused by the unobserved heterogeneity, something that is not explicitly modelled in this article. Consequently, the factors analysed in the article must be interpreted as a means to describe phenomena that are 'system level' in a broad sense, without placing them firmly within a specific national context. Moreover, the problem of data availability limited the number of indicators examined in the models. The work considers only public expenditure without investigating the potential role of private expenditure on HE. Also, teaching and the third-mission performance are represented only partially, although they can be considered as good proxies for the wider dimension of analysis, but can only give a partial representation of the reality. The novel data set that we have proposed is just a first step to cover the lack of indicators for HESs. This effort should be carried forward to improve the quality and availability of data.

Within its specific scope, the article provides significant contributions in terms of policy-making. The results portray the HES as a complex system of interconnections between the three HES activities and the factors that influence them (i.e. the determinants). In describing these interconnections, the article seeks to offer a holistic view that can help policy makers to monitor HE reforms more effectively and take the informed future decisions. In line with this purpose, the research provides a set of systemic measures of HE performance, responding to the need for better HES indicators as expressed forcefully by governments and international organisations (Martin et al. 2011).

Notes

1. PISA is a triennial international survey where education systems are evaluated by testing the skills and knowledge of 15-year-old students.
2. PRIN Project "Comparing Governance Regime Changes in Higher Education: systemic performances, national policy dynamics, and institutional responses. A multidisciplinary and mixed methods analysis", funded by the Italian Ministry of Education.
3. We excluded Luxembourg because being such a small country there were too many outlier values in terms of both GDP per capita and population, the latter being used as the denominator of some of the variables considered. We included Norway and Switzerland because, while they are not part of European Union, they are European countries and their HE systems are relevant to this study.
4. The four cross sections were taken at 5-year intervals; there is, however, a 4-year interval between the last two years selected, as data are largely unavailable for 2015. This time breakdown was believed to be appropriate for capturing variations in HE systems over the last 20 years.
5. Purchasing Power Standards (PPS).
6. PISA is a triennial international survey which evaluates education systems by testing the skills and knowledge of 15-year-old students. The process is designed so that the mean score for each section (mathematics, reading, and science) across OECD countries is set at 500 and the standard deviation is set at 100.
7. The ML estimation method assumes that there is a multivariate normality for the observed variables (the sufficient conditions are that the observations are independent and identically distributed and that kurtosis is zero). The multivariate normality was verified through the results of the tests presented in the Supplementary Annex (see Section A3).
8. According to the rule of thumb, a minimum sample size is between 100 and 200 observations (see Bentler and Chou 1987; Jackson 2003). However, sample size requirements have been queried in a branch of literature where it was demonstrated that the minimum number of observations required varies considerably and that common rules of thumb are totally inadequate (see Wolf et al. 2013). This is also proved by the fact that well-known SEM experts have published research with samples of 100 or less observations (e.g. Browne et al. 2002).
9. From a technical perspective, the reader should be aware of a problem in comparing the indicators over time: Following the transition from ISCED-97 to ISCED 2011, *Education at a Glance* changed the classification of educational levels, affecting the relative indicators from the 2014 edition onwards.

Supplementary data

Supplementary data is available at *SCIPOL* online.

Acknowledgements

This paper reports the results obtained through a research funded under the PRIN Project "Comparing Governance Regime Changes in Higher Education: systemic performances, national policy dynamics, and institutional responses. A multidisciplinary and mixed methods analysis", funded by the Italian Ministry of Education. We are grateful to Prof. Rocco Mosconi for his assistance in setting the econometric modelling. Colleagues of the REMAPS research group at Politecnico di Milano School of Management continuously provided stimuli and suggestions. A preliminary version of this work was presented at the 5th Workshop on 'Efficiency in Education', Budapest (HU) from 19–20 October 2017; Conference on 'Higher Education in Modern Ecosystems: Efficiency, Society and Policies', Augsburg (DE) from 12 to 14 March 2018; 'LEER Conference on Education Economics', Leuven (BE) from 28 to 29 March 2018; 'XXVII Meeting of the Economics of Education Association', Barcelona (ESP) from 28 to 29 June 2018; 'EAIR Forum', Budapest from 26 to 29 August 2018; '6th International Workshop on Efficiency', Huddersfield from 6 to 7 September 2018; '15th International Conference on Higher Education Reform', Baltimore from 24 to 26 September 2018; events at which we received a number of very useful comments. All potential errors are solely our responsibility.

Conflict of interest statement. None declared.

References

Abreu, M., Demirel, P., Grinevich, V. et al. (2016) 'Entrepreneurial Practices in Research-Intensive and Teaching-Led Universities', *Small Business Economics*, 47/3: 695–717.

Agasisti, T. (2011) 'Performances and Spending Efficiency in Higher Education: A European Comparison Through Non-Parametric Approaches', *Education Economics*, 19/2: 199–224.

Baron, R. M., and Kenny, D. A. (1986) 'The Moderator–Mediator Variable Distinction in Social Psychological Research: Conceptual, Strategic, and Statistical Considerations', *Journal of Personality and Social Psychology*, 51/6: 1173.

Bentler, P. M., and Chou, C. P. (1987) 'Practical Issues in Structural Modeling', *Sociological Methods & Research*, 16/1: 78–117.

Bonaccorsi, A., and Daraio, C. (2007) *Universities and Strategic Knowledge Creation: Specialization and Performance in Europe.* Cheltenham: Edward Elgar Publishing.

Browne, M. W., MacCallum, R. C., Kim, C.-T. et al. (2002) 'When Fit Indices and Residuals are Incompatible', *Psychological Methods*, 7/4: 403–21.

Byrne, B. M. (2004) 'Testing for Multigroup Invariance Using AMOS Graphics: A Road Less Traveled', *Structural Equation Modeling*, 11/2: 272–300.

Cave, M., Hanney, S., and Henkel, M. (1995) 'Performance Measurement in Higher Education – Revisited', *Public Money & Management*, 15/4: 17–23.

Cunha, M., and Rocha, V. (2012) 'On the Efficiency of Public Higher Education Institutions in Portugal: An Exploratory Study', *FEP Working Papers*, 1: 34–5.

de Rassenfosse, G., Dernis, H., and Boedt, G. (2014) 'An Introduction to the Patstat Database with Example Queries', *Australian Economic Review*, 47/3: 395–408.

——, and Williams, R. (2015) 'Rules of Engagement: Measuring Connectivity in National Systems of Higher Education', *Higher Education*, 70/6: 941–56.

di Pietro, G. (2004) 'The Determinants of University Dropout in Italy: A Bivariate Probability Model with Sample Selection', *Applied Economics Letters*, 11/3: 187–91.

Dobbins, M., and Knill, C. (2009) 'Higher Education Policies in Central and Eastern Europe: Convergence Toward a Common Model?', *Governance*, 22/3: 397–430.

Docampo, D. (2011) 'On Using the Shanghai Ranking to Assess the Research Performance of University Systems', *Scientometrics*, 86/1: 77–92.

——, and Cram, L. (2017) 'Academic Performance and Institutional Resources: A Cross-Country Analysis of Research Universities', *Scientometrics*, 110/2: 739–64.

Dundar, H., and Lewis, D. R. (1998) 'Determinants of Research Productivity in Higher Education', *Research in Higher Education*, 39/6: 607–31.

Filippakou, O., Salter, B., and Tapper, T. (2012) 'Higher Education as a System: The English Experience', Higher Education Quarterly, 66/1: 106–22.

Guerrero, M., Cunningham, J. A., and Urbano, D. (2015) 'Economic Impact of Entrepreneurial Universities' Activities: An Exploratory Study of the United Kingdom', *Research Policy*, 44/3: 748–64.

Hanushek, E. A., and Woessmann, L. (2011). 'The Economics of International Differences in Educational Achievement', in E. Hanushek, S. Machin, and L. Woessmann (eds.) *Handbook of the Economics of Education*, pp. 89–200. Amsterdam: Elsevier.

Hardeman, S., and Van Roy, V. (2013) *An Analysis of National Research Systems (II): Efficiency in the Production of Research Excellence*. Luxembourg: Publications Office of the European Union.

Hien, P. D. (2010) 'A Comparative Study of Research Capabilities of East Asian Countries and Implications for Vietnam', *Higher Education*, 60/6: 615–25.

Higher Education Funding Council for England. (2011) *Decisions on Assessing Research Impact*. Bristol: Higher Education Funding Council for England.

Hoareau, C., Ritzen, J., and Marconi, G. (2013) 'Higher Education and Economic Innovation, a Comparison of European Countries', *IZA Journal of European Labor Studies*, 2/1: 24.

Jackson, D. L. (2003) 'Revisiting Sample Size and Number of Parameter Estimates: Some Support for the N:Q Hypothesis', *Structural Equation Modeling*, 10/1: 128–41.

Jallade, J. P. (1992) 'Undergraduate Higher Education in Europe: Towards a Comparative Perspective', *European Journal of Education*, 27/1/2: 121–44.

Jöreskog, K. G. and Sörbom, D. (1993) *LISREL 8: Structural Equation Modeling with the SIMPLIS Command Language*. Skokie, IL: Scientific Software International.

Kline, R. B. (2015) *Principles and Practice of Structural Equation Modeling*. New York, NY: Guilford Publications.

Kyvik, S. (2004) 'Structural Changes in Higher Education Systems in Western Europe', *Higher Education in Europe*, 29/3: 393–409.

Lomax, R. G., and Schumacker, R. E. (2004) *A Beginner's Guide to Structural Equation Modeling*. London: Psychology Press.

Lucas, S. R. (2001) 'Effectively Maintained Inequality: Education Transitions, Track Mobility, and Social Background Effects', *American Journal of Sociology*, 106/6: 1642–90.

MacCallum, R. C., Browne, M. W., and Sugawara, H. M. (1996) 'Power Analysis and Determination of Sample Size for Covariance Structure Modeling', Psychological Methods, 1: 130–49.

Martin, B. R. (2012) 'Are Universities and University Research Under Threat? Towards an Evolutionary Model of University Speciation', *Cambridge Journal of Economics*, 36/3: 543–65.

Martin, M., Sauvageot, C., and Tchatchoua, B. (2011) *Constructing an Indicator System or Scorecard for Higher Education: A Practical Guide*. Paris: UNESCO, International Institute for Educational Planning.

Molas-Gallart, J., Salter, A., Patel, P. et al. (2002) Measuring Third Stream Activities. *Final Report to The Russell Group of Universities*. Brighton: SPRU, University of Sussex.

Obadić, A., and Aristovnik, A. (2011) 'Relative Efficiency of Higher Education in Croatia and Slovenia: An International Comparison', *Amfiteatru Economic Journal*, 13/30: 362–76.

Organization for Economic Cooperation and Development (OECD). (2017) *Benchmarking Hhigher Education System Performance: Conceptual Framework and Data, Enhancing Higher Education System Performance*. Paris: OECD.

Phillimore, A. J. (1989) 'University Research Performance Indicators in Practice: The University Grants Committee's Evaluation of British Universities', 1985–86', *Research Policy*, 18/5: 255–71.

Richman, B. and Farmer, R. (1974) *Leadership Goals and Power in Higher Education*. San Francisco, CA: Jossey-Bass.

Saisana, M., d'Hombres, B., and Saltelli, A. (2011) 'Rickety Numbers: Volatility of University Rankings and Policy Implications', *Research Policy*, 40/1: 165–77.

Santoalha, A., Biscaia, R., and Teixeira, P. (2018) 'Higher Education and Its Contribution to a Diverse Regional Supply of Human Capital: Does the Binary/Unitary Divide Matters?', *Higher Education*, 75/2: 209–30.

St. Aubyn, M. S., Garcia, F., and Pais, J. (2009) *Study on the Efficiency and Effectiveness of Public Spending on Tertiary Education.* Bruxelles: Directorate General Economic and Financial Affairs, European Commission.

Triventi, M. (2014) 'Higher Education Regimes: An Empirical Classification of Higher Education Systems and its Relationship with Student Accessibility', *Quality & Quantity*, 48/3: 1685–703.

van Zeebroeck, N., and van Pottelsberghe de la Potterie, B. (2011) 'Filing Strategies and Patent Value', *Economics of Innovation and New Technology*, 20/6: 539–61.

Veugelers, R. (2016) 'The Embodiment of Knowledge: Universities as Engines of Growth', *Oxford Review of Economic Policy*, 32/4: 615–31.

Williams, R., and de Rassenfosse, G. (2016) 'Pitfalls in Aggregating Performance Measures in Higher Education', *Studies in Higher Education*, 41/3: 51–62.

——, ——, Jensen, P. et al. (2013) 'The Determinants of Quality National Higher Education Systems', *Journal of Higher Education Policy and Management*, 35/6: 599–611.

Wolf, E. J., Harrington, K. M., Clark, S. L. et al. (2013) 'Sample Size Requirements for Structural Equation Models: An Evaluation of Power, Bias', *and Solution Propriety', Educational and Psychological Measurement*, 73/6: 913–34.