ARTICLE - SPECIAL ISSUE

Framework and operationalisation challenges for quantitative comparative research in higher education

Marco Seeber [iD]

Department of Political Science and Management, University of Agder, Kristiansand, Norway

Abstract

The increasing availability of data on higher education systems, institutions and their members creates new opportunities for comparative research adopting a quantitative approach. The value of future studies crucially depends on the capability to recognise and address some major methodological challenges existing in quantitative comparative research in higher education. The higher education context presents in fact specific features that can hinder comparisons, and political and social processes occurred in recent decades further enhanced complexity. This article aims to discuss key challenges that are currently met in quantitative comparative research in higher education while developing the conceptual framework and in research operationalisation, to discuss possible solutions and the value of configurational and multilevel analytical approaches in identifying meaningful objects of comparisons, to take into account the complexity of the higher education context and in identifying causal relationships.

Sammendrag

Den økende tilgjengeligheten av data som omhandler høyere utdanningssystemer, institusjoner og deres ansatte, gir nye muligheter for komparativ forskning med en kvantitativ tilnærming. Fremtidige studiers verdi vil, i stor grad, avhenge

av deres evne til å gjenkjenne og adressere de viktige metodologiske utfordringene som en finner i kvantitativ komparativ forsking på høyere utdanning. Konteksten til høyere utdanning kan ha karakteristika som hindrer sammenlikning, noe som er blitt ytterligere forsterket av politiske og sosiale prosesser de siste tiårene. Denne artikkelen diskuterer sentrale utfordringer som en møter i kvantitativ komparativ forsking på høyere utdanning under utarbeidelsen av det metodisk konseptuelle rammeverket, og i operasjonaliseringen av forsking, for å diskutere mulige løsninger og verdien av konfigurasjons- og flernivå analytiske tilnærminger for å identifisere meningsfulle objekter en kan sammenlikne, gitt kompleksiteten i høyere utdanningsfeltet, og identifisere årsakssammenhenger.

1 | INTRODUCTION

In recent decades, technological, political and managerial transformations have changed the landscape for comparative research in higher education, unveiling new opportunities and new challenges.

Technological progress with respect to software, computing power, digitalisation, connectivity and communication has made it easier to retrieve, store and analyse data. Moreover, the global competition for prestige, growing demands for accountability and efficient use of resources, as well as the introduction of new steering instruments and the professionalisation of administrative staff, have increased the *need* and *capability* of higher education institutions (HEIs) to collect information (Chirikov, 2013; Mathies & Välimaa, 2013). In parallel, university associations increasingly gather information on their members, and national datasets have become more accessible thanks to the Internet, and more comparable through the standardisation of indicators. International organisations such as the Organisation for Economic Co-operation and Development (OECD) and the European Union (EU) have also played important roles in creating and widely disseminating multinational datasets. As a result of these processes, scholars now have at their disposal an unprecedented wealth of data on higher education systems, institutions, scientists and students, as well as computing instruments and power to analyse them.

While the greater availability of data[1] and analytical tools create new opportunities for quantitative comparative research and favours a shift from a 'case-oriented' towards a 'variable-oriented' approach (Bleiklie, 2014), each higher education system still possesses peculiar features which can hinder country comparisons, as well as between HEIs and individuals in different systems. Moreover, political and social processes have occurred in recent decades to make the higher education context more complex to analyse. Namely, globalisation, modernisation policies and multilevel governance have spurred new competitive and institutional forces that operate beyond and within the traditional national boundaries, forcing comparative thinkers to adopt more sophisticated frameworks.

The goal of this article is therefore to address some key challenges currently met in quantitative comparative research in higher education. The structure of the article reflects key phases in developing a quantitative study: (a) setting the framework, (b) operationalisation and (c) analysis. Quantitative research and comparative research in the social sciences each display distinctive challenges, e.g., of validity and reliability vis-à-vis conceptual

equivalence. This article focuses on the distinct challenges emerging in higher education research when quantitative and comparative approaches are combined. The first section examines common issues related to the development of a conceptual framework for a quantitative comparative study in higher education, namely in identifying meaningful objects of comparison and taking into account the complexity of the higher education context. The second section elaborates on the key challenges emerging in research operationalisation, whereas the third section discusses the usefulness of two analytical approaches—configurational and multilevel—in addressing the challenges discussed as well in detecting causal relationships.

2 | THE FRAMEWORK OF A QUANTITATIVE COMPARATIVE STUDY IN HIGHER EDUCATION

Developing the conceptual framework of a quantitative comparative study raises specific challenges for the higher education researcher, when it comes to define the objects of comparison[2] and the factors that matter for the phenomenon under examination.

While comparative research explores a given phenomenon in different macrosocial contexts, like countries, societies, nations or regions (Teichler, 1996), and uses attributes of the macrosocial contexts as potential explanatory variables (Ragin, 2014), the object of comparison is not necessarily a macro-level phenomenon, but it may well be—for instance—a phenomenon at the meso level (organisational level) or micro level (individual level) (Bleiklie, 2014), such as HEIs' or departments' strategies, or students' satisfaction in different countries. In addition, variables at the macro level often derive from the aggregation of organisational and individual data, like a country number of HEIs or total scientific production. Therefore, a comparative higher education study not only entails defining the macrosocial units being compared (e.g., what higher education systems or countries), but also what organisations and/or individuals must be considered, and to ascertain the implications of different sampling decisions.

Decisions regarding which HEIs must be considered for the comparative analysis, namely the definition of the *institutional perimeter*, bear great importance and should be carefully pondered (Huisman, Lepori, Seeber, Frølich, & Scordato, 2015). HEIs are commonly defined as institutions awarding degrees at least at the bachelor level. However, systems greatly differ in the types of organisations they comprehend and their relative proportions, such as according to their legal status (public or private), type (universities and universities of applied sciences), and their functions (teaching only vs. teaching and research). For example, Spain is a unitary system including only universities that all have research activity as a core task, whereas Belgium is a binary system including universities of applied sciences—for which research is not a core task—as well as universities, which are strongly research-oriented. Hence, a study comparing the research quality of HEIs in the two countries will provide very different results depending on whether both kinds of institutions or only universities are considered for Belgium. Moreover, a comparative study may wish to consider not only HEIs that formally belong to the national higher education system, but all institutions located in a country, such as including private HEIs that do not award state-recognised degrees. Comparability issues may also emerge due to outliers, namely institutions that display peculiar traits and whose inclusion or exclusion can strongly impact the results (Huisman et al., 2015).

Systems may also differ regarding the definition of the individual members. For example, PhD students employed as teaching assistants or research assistants are included among academic staff in some systems and not in others (Lepori et al.,2018). Thus, among the HEIs included in the European Tertiary Education Register (ETER),[3] British and German HEIs exclude PhD students from the academic staff count, whereas Dutch universities include them (Lepori et al., 2018). Such differences can distort comparisons, for example on gender balance or the internationalisation of the academic staff—since the share of female and foreign staff tends to be much larger among PhD students than among professorial staff. The researcher should therefore ascertain the attributes held by the same units in different systems, and eventually remedy the differences that may distort the comparison (in the

case of comparing the internationalisation of academic staff, for instance, by excluding PhD students from the pool of the academic staff).

Beyond clarifying the objects of comparison, a second pivotal step is to identify the factors that matter for the phenomenon of interest, for example by focusing on relationships between macro-level constructs (e.g., investment in the higher education system and the system's educational output), macro-level and meso-level constructs (e.g., system policies and universities' performance, or vice versa), as well as between macro-level and micro-level constructs (e.g., systems' policies and teachers' behaviours, or vice versa). In this regard, comparative higher education research presents specific and increasing challenges due to the existence of multiple contexts in which individuals, organisations and systems are embedded and intertwined. First, higher education has been long regarded a peculiar social setting, characterised by strong national differences and the prominence of scientific disciplines' system of beliefs, goals and social structures, which differ from each other and cut across national systems (Clark, 1983; Whitley, 2000). Second, while HEIs have long been known to operate as loosely coupled structures with little ability to influence their members' behaviour and shape their environment, global competition and modernisation policies have spurred universities to actively steer their members, and develop clear boundaries and strategies (Brunsson & Sahlin-Andersson, 2000; Krücken & Meier, 2006; Seeber et al., 2015). Third, globalisation processes and the emergence of multilevel governance approaches have created new institutional and competitive forces beyond and within the national level (Enders, 2004; Jones & Oleksiyenko, 2011; Maassen & Olsen, 2007).

In turn, comparative higher education studies that focus primarily on system differences, or system and individual level variables and their relationships, are no longer sufficient (Kosmützky, 2015). Comparative higher education research must now take into account the multiple embeddedness of individuals, within disciplines and HEIs that possess distinctive values and objectives, and the multiple embeddedness and interactions of HEIs with local, national and supranational contexts conveying different competitive and institutional pressures (Frølich, Huisman, Slipersæter, Stensaker, & Bótas, 2013; Hüther & Krücken, 2016; Marginson, 2006; Seeber, Cattaneo, Huisman, & Paleari, 2016). Neglecting such complexity may threaten the validity of the comparative study.

As a matter of example, Figure 1 juxtaposes a simple conceptualisation of the higher education context (left side) and a complex one (right side). The first only considers that micro-level units—like students or scientists, represented by letters—are embedded into different higher education systems. The latter also considers that individuals are simultaneously embedded into different HEIs and disciplines—with distinct norms and pressures, and that HEIs are themselves embedded into different systems and supranational contexts—such as European universities in the European Higher Education Area (EHEA).
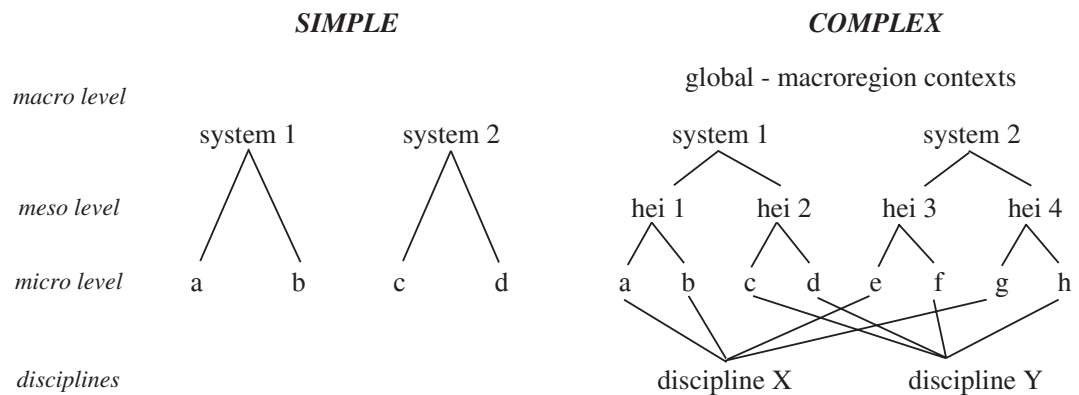

FIGURE 1 Examples of conceptualisations of the higher education context

3 | OPERATIONALISATION: CHALLENGES FOR COMPARATIVE HIGHER EDUCATION

Quantitative social science aims to test hypotheses on relationships—correlational or causal—between variables (Ragin, 2014). Testing these relationships requires a process of operationalisation, whereby abstract phenomena are represented via concepts and operational definitions (i.e., indicators) that allow to measure and quantify them.

Operationalisation's main challenge is to provide suitable representations of the phenomena being investigated. At the same time, comparative research requires that concepts and indicators can be meaningfully employed across different contexts.

3.1 | Concepts

The phenomena of central interest to higher education researchers are often multidimensional and abstract (like student satisfaction, teaching quality, research productivity), which are more complex to operationalise compared to mono dimensional and concrete phenomena (such as the number of students). The definition of concepts, their attributes and meanings, are not straightforward for multidimensional and abstract objects of analysis. This problem is heightened by the fact that the field of higher education research hosts scholars with different research backgrounds, who may have different ideas about how a given construct should be conceptualised. For instance, scholars comparing universities' efficiency in education often conceptualise educational output quantitatively, in terms of the number of students enrolled and/or degrees produced (e.g., Abbott & Doucouliagos, 2003; Johnes, 2006). Some researchers may find this conceptualisation insufficient, because it ignores that the cost of producing a degree differs greatly across disciplines, nor does it consider the quality of the educational output or the students' improvement vis-à-vis their initial capabilities. More detailed conceptualisations may then include the cost and quality of education, with operational definitions consisting of the standard cost for producing each graduate in different disciplines or students' test scores (e.g., Witte & López-Torres, 2017).

In choosing among possible conceptualisations, the researchers should take into consideration not only which one is better in terms of representing the underlying phenomenon, but also whether it is suitable for comparing systems. In this respect, a major challenge is that the key attributes of concepts employed in higher education research often differ across systems. This is well illustrated by the previously mentioned example of academic staff. Namely, while ETER sets the defining attributes of academic staff as being related to the combinations of a role (staff, PhD students) and activities (research, teaching, public service),[4] yet variations exist across systems in PhD students' teaching and research duties that affect whether or not they are included among the academic staff.

A related challenge emerges when concepts developed in one context are used in new contexts, where the original definition may be unsuitable. Higher education scholars will meet this problem more frequently in the future, as data availability extends the possibility of comparative studies to new systems. It is then important to review this problem and possible solutions with an example.

In a study of the internationalisation activities (IA) of European HEIs, Seeber, Meoli, and Cattaneo (2018) conceptualised three types of HEIs' internationalisation profile: basic, academic and entrepreneurial. *Entrepreneurial* HEIs, for example, develop both IA aimed to improve research and teaching, as well as IA aimed to attract resources. Let's suppose a researcher aims to investigate HEIs' IA in African countries by using the conceptualisations extrapolated from the analysis of European HEIs. This is potentially problematic, as all HEIs in the new context may belong to one typology—namely the conceptualisation is not helpful in discerning differences, or HEIs may display portfolios of IA that do not fit the existing conceptualisations, such as an hypothetical portfolio comprising *only* IA aimed at attracting resources. One option is to run the clustering analysis with the new sample and identify new concepts and their defining attributes. However, the risk of a 'sample-dependent' approach (Cooper & Glaesser, 2011; Kent, 2015) is that the emerging concepts may have little in common with the original

ones, meaning comparisons with the European context are unfeasible. A second option is to start from the defining attributes of the original concepts and adapt them, resulting in the typical challenge of comparative studies of managing the trade-off between *specification* and *abstraction*, between system-specific and system-inclusive concepts (Smelser, 1976). In pursuing abstraction, the researcher will relax the constraints of the original definition—for example, by including new attributes or by considering units that only possess a few of the defining attributes as members. The potential downside is conceptual stretching (Sartori, 1970), namely that the boundaries of the concept become so blurred that it loses its original meaning. An alternative solution is to add new 'defining attributes' (such as, recruiting international research talent) and include both the original and the new defining attributes.

In any case, decisions upon a concept's attributes and meaning (its intension) are intertwined with decisions regarding the set of entities to which the concept should apply (its extension; namely, what is to be considered an entrepreneurial HEI in internationalisation?). Given the set of defining attributes, different approaches exist to determine the extension. For example, in the *family resemblance* approach, the members are those units that have at least one or some attributes in common with every other unit. In the *radial* approach, a prototype member is defined—e.g., by having all attributes—and real members must possess a certain number of attributes of the prototype (Collier & Mahon, 1993).

In sum, applying a concept in a new context should achieve a suitable balance between generalisation and ability to represent local differences. No single approach is always preferable to the others in this regard. Rather, the researcher should weigh the pros and cons in terms of the resemblance of the new concept vs. the original concept, its usefulness in describing the new context, and for developing broader comparisons.

3.2 | Indicators

Indicators represent operational definitions of a concept, which allow its measurement. Comparative researchers should develop or rely on indicators—and related measures—that are not *biased* and that are *equivalent* (He & van de Vijver, 2012).

Indicators are biased when affected by systematic measurement anomalies. An important precondition for meaningful comparisons is in fact that indicators and related measures are the same in different contexts. For example: data on personnel may be expressed in terms of head counts, or full-time equivalent; financial data should be converted in the same currency or, given that exchange rates can vary rapidly, in purchasing power standards. In some cases, data need to be rearranged to allow for comparability. For example, some systems encompass three tenured academic ranks (assistant, associate and full professors—e.g., in the Netherlands), while others encompass two (e.g., in Norway, associate and full professors), so that researchers may need to re-elaborate classifications in order to make them comparable.[5]

Indicators are equivalent when the resulting scores' relationship to the whole construct that is measured is the same across systems. This property can be illustrated with an example. Only a small part of icebergs is observable, while 90 per cent lies beneath sea level; yet, we can compare the size (phenomenon-construct) of icebergs (objects compared) through the small part that is visible (proxy-measure), because this remains in constant proportion to the whole for all icebergs.[6]

However, how can researchers design indicators that satisfy such properties and that can be safely employed for comparative research?

In many cases, the finer the operationalisation, the better the representation of the whole phenomenon, and the better the equivalence. For instance, a researcher aiming to compare a systems' research output (proxy) can rely on the number of articles included in the Web of Science (WOS) database. Yet in many countries, monographs are still very common, or incentives reward publications indexed in Scopus database rather than in WOS, so relying on the WOS only will implicitly overestimate the production of systems that focus on such a database

compared to other systems. Instead, using a more comprehensive proxy that includes both types of publication archives will provide more meaningful comparisons.

In other cases, increasing the precision/comprehensiveness of the indicator may not be necessary or desirable. This is true even for seemingly simple concepts that encompass a single and concrete dimension. A notorious example is the *coastline paradox*, formulated as a puzzle by the mathematician Benoît Mandelbrot to determine 'how long is the coast of Britain?' (1967). This problem appears to be rather straightforward, but the length of the coastline depends on the scale of measurement: the shorter the yardstick the longer the measured length, as the ruler would be laid along a more curvilinear route. Hence, if the coast of Britain is measured with a yardstick of 200 km, it measures 2,400 km; if a 50-km yardstick is used, the coast measures 3,400 km, and so forth. However, as the yardstick becomes more and more precise and decreases towards zero, the measured length increases ad infinitum, leading to a measure that is of little empirical value. In higher education research this is a very common problem. For example, in order to measure the social impact of an HEI's activities, a researcher can think about a very large number of indicators covering the many ways in which HEIs' activities impact society. Yet one should wonder if the increased precision also delivers measures that are superior for comparison and whether the gain is worth the cost and time needed for more data collection. In some cases, increased precision may even lead to measures that fare worse for comparative purposes. For instance, a researcher comparing a system's performance in third mission activities can rely on the number of patents produced and the amount of resources attracted from third parties. However, HEIs in some countries may be accurate in collecting information about similar revenues, while others may be not. In such cases, an indicator that is less representative of the whole, but based on more homogeneously collected data, will guarantee a stronger degree of equivalence.

It is often the case that full equivalence cannot be achieved, and the choice of the indicator will affect the comparison in favour of one system or another. In other words, indicators are rarely 'neutral' to the comparison. Once more, the example of the coast length is illustrative: because the coastlines of Russia and Norway are jagged to a different extent, the former may be longer or shorter than the latter depending on the yardstick's length. In any case, researchers should avoid operationalisations that engender blatant comparative biases and acknowledge the consequences of indicators' choices on comparison.

Decisions upon whether and how to normalise the indicators are also very relevant for meaningful comparisons. For example, the number of universities from a country listed in the universities' ranking have been used to compare the quality of higher education systems governance and funding policy. Since US universities dominate such rankings, it is frequently deducted that US governance and funding structures are the best and should be imitated by other systems (e.g., Aghion, Dewatripont, Hoxby, Mas-Colell, & Sapir, 2007). However, the number of HEIs in the ranking is largely affected by the sheer size of a country higher education system, such as in terms of investment in research. Hence, when the number of HEIs in the university ranking is used to infer the quality of the higher education governance, this number should be at least normalised according to proxies of a system's size.[7]

Specific issues of comparability exist for indicators based on the judgement of experts and the perceptions of non-experts. Experts' judgements or individual perceptions may in fact be consciously or unconsciously affected by the individual characteristics or role of the respondent, and by the context in which the respondent is embedded.

Expert-based indicators rely on judgements from people who have better knowledge or hold positions that allow them to give an accurate evaluation and information. Experts' judgements are particularly valuable when it is impossible to observe, measure or experiment on the phenomenon in question, and when indicators based on concrete dimensions provide incomplete or questionable proxies (Benoit & Wiesehomeier, 2009). For example, the number of publications and citations provide a proxy of scientific output, yet the value of scientific production cannot be fully captured by numbers alone and judgements from reputed academics can provide valuable, complementary information. Indicators for measuring the decision-making power of different academic bodies provide another example. Proxies can be constructed based on universities' statutes, which specify academic bodies' formal powers—yet some bodies or persons can have little formal decision-making power, but have strong

decision-making power in practice, thanks, for instance, to a favourable network position (Padgett & Ansell, 1993). In a similar case, a survey of people holding insider positions, such as employees in the central administration, can provide more valid information on actual power than the information derived from official documents. In the case of experts' judgement, comparability is at stake when respondents in different units display systematic differences in traits or roles. For instance, when asked if internationalisation is important for their institution, rectors may give systematically higher scores than deans. If systems differ in the proportion of respondents in the two roles, comparisons will be distorted. Therefore, experts should have similar characteristics in different systems, and/or the inferential analysis should include specific variables to take such effects into account.

For given constructs, there are no real experts with privileged information. For example, when it comes to concepts that pertain to cognitive dimensions or dimensions which are socially constructed, such as an organisation's culture, identity or quality of the working environment. These attributes are diffused through the organisation and possibly heterogeneously, so there is no privileged point of view to observe them. In this case, scholars develop surveys to collect perceptions from a variety of members from the units analysed (i.e., 'non-experts'). For indicators derived from perceptions, a major threat to comparability occurs when the measure is biased towards the point of view of as a specific subgroup. To be comparable, samples should be representative of the whole (system or organisation) population, along the individuals' traits—such as rank, role, gender, age—and the variety of the relevant contexts in which they are embedded—such as department/discipline, and the type of HEIs in which they work. For example, academics working in a technical university/department may hold different views on the purpose of education than respondents from a university/department specialised in the humanities. When the sample is not representative of the whole population, the researcher can weight responses from different groups differently in order to rebalance their over- or under-representation in the sample (so-called 'post-stratification', Davern, 2008).

Unwarranted response variations can also be triggered by country differences in contextual pressures, such as on what is socially desirable and what is not (Kreuter, Presser, & Tourangeau, 2008). The same question or measure may even hold different meanings in different countries, cultures (Välimaa & Nokkala, 2014) or, more subtly, items translated in different languages may convey different meanings (Brady, 1985). An extensive literature has explored comparability of survey items in cross-cultural and cross-national survey research (Holland & Wainer, 1993; Peng, Peterson, & Shyi, 1991; Strauss, 1969), and several types of equivalence have been identified, which can be grouped in two fundamental domains (Johnson, 1998). Interpretive equivalence concerns with similarities in how abstract or latent concepts are interpreted across cultures, including, for instance, conceptual equivalence, when the construct has similar meanings within the contexts compared (Okazaki & Sue, 1995), and functional equivalence, which exists to the extent that the concept serves similar functions within each society being investigated (Singh, 1995). Procedural equivalence regards instead the measures and procedures used to make comparisons, and emphasises the applicability of identical procedures, including among others stimulus equivalence (Anderson, 1967) and text equivalence (Alwin, Braun, Harkness, & Scott, 1994).

Several techniques are also available for evaluating the comparability of survey measures across nations or cultures (Braun & Johnson, 2010; Johnson, 1998). Equivalence across contexts can be pursued, for instance, by improving response anchors, deleting the more problematic answers, correcting for respondents' differential interpretations (King, Murray, Salomon, & Tandon, 2004), developing several indicators and retaining those that are correlated across countries, as such correlation signals that they relate to the same underlying concept (Przeworski & Teune, 1966).

For example, surveys exploring questionable research practices can solely inquire about perceptions—for instance: 'Do you think that there is a problem of questionable research practices in your university?' However, respondents in different countries may hold different ideas of what a questionable research practice is, so conceptual equivalence should be ascertained in advance—for example, through a vignette design. At the same time, procedural equivalence may be at stake because responses to such a question can be affected by contextual factors, such as how much time the mass media in the country dedicate to scandals in science, which might be decoupled

from the 'true' level of questionable research practices. Hence, researchers can opt for operationalisations that are less vulnerable to cultural and contextual biases, such as questions about objective facts, like: 'Do you know of any colleague ever fabricating data?',[8] although also this measure may have limitations in terms of procedural equivalence, because the proportion of questionable practices in research consisting of 'fabricating data' may vary across countries.[9]

In turn, operationalisation via survey items presents trade-offs in terms of comparability, which should be carefully assessed.

4 | PROMISING APPROACHES TO COMPARATIVE ANALYSIS

This section discusses the value of configurational and multilevel research approaches in addressing some of the challenges discussed so far, namely in setting an appropriate institutional perimeter, taking account of the complex social structure of higher education systems and identifying causal relationships, which is often a prime goal of comparative research (Allardt, 1990; Ragin, 2014; Reale, 2014; Verba, 1967).[10]

Configurational research aims to identify groups of organisations that resemble each other along important dimensions, through statistical techniques such as cluster analysis. This approach assumes that organisations can be better understood by identifying groups of organisations with distinct, internally consistent traits, rather than by seeking to uncover causal relationships that hold true for all organisations (Ketchen, Thomas, & Snow, 1993; Short, Payne, & Ketchen, 2008).

The configurational approach is valuable to define meaningful institutional perimeters—i.e., what HEIs are comparable in different systems and should be therefore considered in the analysis. In fact, country variation in HEIs' organisational taxonomies may hinder functional equivalence, and hide differences or similarities that threaten the validity of the comparison: HEIs with similar traits and functions may be named differently (such as *Fachhochschulen* in Germany and *Hogescholen* in the Netherlands), or HEIs with substantial differences may have the same official category (the label 'university' can be used by 'world-class' research-intensive institutions as well as by regional, teaching-oriented ones). Thus, researchers should take advantage of configurational approaches to identify in the first place which HEIs belong to the same population and can be rightfully compared.

The configurational approach can also be valuable to identify causal relationships between constructs, particularly when the relationship between an independent and a dependent variable is not the same for the whole sample—for instance, when the relationship is positive for a subsample of cases and negative for another one. Hence, once internally coherent groups are identified, the researcher can explore what organisational traits and contextual conditions relate to specific configurations. The configurational approach can also be used in combination with the method of difference (Goedegebuure & van Vught, 1996; Mill, 1967) to explore causal relationships. Namely, by identifying cases that are classified in different groups and display similar traits and environmental conditions except one, this one trait or condition can be regarded as the cause for the different classification. For instance, in Seeber et al.'s (2018) study, Spanish and Italian HEIs display very similar contextual conditions and traits, namely they are all universities of unitary systems, with a Napoleonic administrative tradition, endowed with similar resources and for which tuition fees represent between 10 per cent and 20 per cent of the total revenues. Yet Spanish HEIs are much more often *entrepreneurial* in internationalisation than Italian HEIs. A major difference between the two systems is their language, which could explain the difference. In fact, around 500 million people speak Spanish outside Spain, whereas far fewer Italian (native) speakers live outside Italy. Hence, Spanish HEIs have a much greater potential to attract students from abroad than Italian HEIs by developing entrepreneurial IA (Seeber et al., 2018).

A second important analytical approach is represented by multilevel models, which are particularly valuable to represent and address the complexity of the higher education settings, where individuals and organisations are embedded in multiple social environments at the same time (see Figure 1). Multilevel regression models consider

that observations (like individuals or HEIs) are not independent from each other, as they share the same environment and they affect each other (Snijders & Bosker, 2012). Multilevel models split the variance of the dependent variable (such as individual performance) between levels (like systems, HEIs, disciplines), hence diagnosing the extent to which a phenomenon is affected by factors at each level (Jones & Subramanian, 2012; Snijders & Bosker, 2012). For example, the more universities in the same country display a similar behaviour between each other and different from universities in other countries, the more the variance at the country level will be proportionally greater and statistically significant, thus signalling that the causes of the behaviour should be searched at the country level. On the contrary, if less than 5% of the variance is accounted for by between-countries variations, it can be inferred that system policies and country traits have a minor impact. For instance, by adopting a multilevel approach to explore HEIs' rationales to internationalise, Seeber et al. (2016) found that country-level factors matter only for few rationales and systems, and that HEIs' internationalisation rationales are also affected by the extent to which they are embedded in the global competitive arena.

Multilevel models are also valuable to overcome some notable challenges of comparative studies in detecting causal relationships. When comparisons occur between macro-level constructs, the fact that macro-level social units (higher education systems) are often limited in number may lead to what is known as the 'many variables, small N' problem, namely that too many explanatory variables may exist for relatively few cases (Lijphart, 1971). The increased availability of data at meso and micro level combined with multilevel models can often overcome this obstacle, because the number of observations for each country becomes much larger and it is possible to control for alternative explanatory factors at meso and micro level. For example, the number of universities in a country listed in universities' ranking may not be affected only by a system's governance, funding structure, language, size and investment in research (Li, Shankar, & Tang, 2011), but also by its HEIs' characteristics, like their internal governance or their size.

5 | CONCLUSIONS

This article discusses some key challenges that currently characterise quantitative comparative research in higher education as well as possible solutions. Some choices and limitations that underpin this work should be considered. First, the article focuses on challenges that are relevant for a vast array of comparative studies—like the definition of the objects of comparison—and that most frequently undermine the validity of quantitative comparative studies in higher education. The list of challenges examined is far from exhaustive, and future studies should explore other important methodological aspects. Second, the article focuses on methodological challenges that are relevant for the kind of data and analytical tools that are currently available to most higher education researchers, such as surveys, institutional research, databases developed from large organisations. Namely, data that are: (a) structured and (b) with a size that is manageable by common hardware and software tools. The article does not discuss challenges related to the rise of so-called 'big data', which are typically (a) unstructured and (b) with a size beyond the processing capability of common hardware and software tools. These kinds of data require distinct analytical approaches—such as machine learning aiming at pattern recognition—and imply a new array of methodological challenges for comparative higher education scholars.

Given nowadays' rapid pace of social, political and technological transformation, some reflections are warranted upon the future opportunities and challenges of (quantitative) comparative higher education research. On the one hand, despite a long-standing policy pressure aimed at blending the peculiar features of the higher education sector, by transforming universities from loosely coupled to complete organisations (Brunsson & Sahlin-Andersson, 2000) and increasing external guidance and standardisation of research and education activities (Hood, James, Peters, & Scott, 2004; Musselin, 2007; Paradeise & Thoenig, 2013), nevertheless academic tasks, structures and identities have often proven resilient towards attempts of steering and homogenisation (Musselin, 2007; Seeber et al., 2015; Thoenig & Paradeise, 2014). Thus, future research needs to consider the

lasting complexity of the higher education context. On the other hand, more recent political and technological processes may have durable effects on the opportunities and modes of quantitative comparative higher education. First, the scientific, economic and political hegemony of Western liberal democracies, which eased the spread of institutions and models of Western systems and organisations (Mounk & Foa, 2018), is increasingly contested. For example, the proportion of scientific articles from China compared to the US has grown from 9 per cent to 88 per cent between 1996 and 2018 (Source: Scimago). In 2003 no Chinese university could be found among the top 200 positions of the Academic Ranking of World Universities (ARWU) Shanghai ranking, whereas 19 are found in 2019, second only to US universities. This process weakens the isomorphic pressure towards a single template, as other systems can claim a role model or their distinctiveness (Marginson, 2017). This shift can possibly spur diversity of practices and governance arrangements in higher education systems and institutions around the globe, or even reverse the direction of isomorphic forces, creating opportunities and need for broader international comparisons. At the same time, broader comparative studies heighten methodological challenges, because the more the compared systems differ in terms of their cultural and political background, the more they likely differ in terms of understandings and assumptions about key constructs. Moreover, shifts in global economic and political influence are also resulting in geopolitical frictions, which may threaten the scientific collaboration needed for large-scale comparative studies.

A second major process looming over quantitative social sciences is the rise of the 'big data' and the explosion in computing power. It is estimated, for example, that 90 per cent of all digital data in the world was created in the last two years (McAfee & Brynjolfsson, 2017). Access to a much larger and deeper set of information on systems, organisations and individuals creates huge opportunities for research, but also new methodological and ethical challenges for comparative higher education studies (Foster, Ghani, Jarmin, Kreuter, & Lane, 2016). Future research should address both foreseeable challenges, such as of operationalisation in spite of different regulations on privacy and ethical standards, as well as challenges that will emerge with the diffusion of new research approaches.

ORCID

Marco Seeber 🔗 https://orcid.org/0000-0002-0162-6289

ENDNOTES

[1] The article focuses on kinds of data currently available to most higher education (HE) researchers. We do not refer here—but discuss in the conclusion—the so-called 'big data'.

[2] Regarding the challenges of selecting units for comparison in quantitative studies, see also Ebbinghaus (2005).

[3] ETER is a European Commission initiative consisting in a database collecting information on the basic characteristics, geographical position, educational activities, staff, finances and research activities of HEIs in Europe. ETER provides data at the level of individual institutions, which complement educational statistics at the country and regional level provided by EUROSTAT (the EU's statistical office); see also: www.eter-project.com.

[4] ETER defines as academic staff: (a) staff whose primary assignment is instruction, research or public service, (b) staff who hold an academic rank, like professor, assistant professor, lecturer or an equivalent title, (c) staff with other titles (like dean, head of department, etc.) if their principal activity is instruction or research, and (d) PhD students employed for teaching assistance or research.

[5] The development of the Changing Academic Profession survey implied a similar challenge, and academic ranks were grouped into junior and senior positions (Rostan et al., 2014).

[6] I am grateful for this metaphor to Daniele Archibugi, who used it during a project meeting.

[7] A country number of HEIs in the ranking is affected by other factors unrelated to the quality of the governance and funding structure, such as whether it is an English-speaking country (Li et al., 2011), and most popular rankings also consider indicators related to past scientific performance (such as the number of Nobel prize winners from the university, ARWU, 2018), that implicitly favours systems with an established scientific tradition.

[8] For a meta-analysis on research malpractices see e.g., Fanelli (2009).

[9] A similar issue is illustrated by surveys exploring countries' level of corruption, which employ both questions based on perceptions of corruption, i.e., 'Do you think that corruption is a problem in this country?' and about objective facts, i.e., 'Have you paid a bribe in the past 12 months?' Countries may rank very differently according to the two indicators. For instance, Italy ranks among the fairly corrupted countries according the first question (75th position), and among the least corrupted according to a fact-related question (15th position) (Transparency International, 2013).

[10] By comparing different systems, a researcher can in fact deduct which policies, for instance, are more effective. Increased capability to retrieve data at the organizational and individual levels opens up the possibility for comparative studies to not only explore relationships between macro-level constructs (e.g., system policy → system efficiency), but also relationships between macro- and micro-/meso-level constructs (e.g., system policy → individual/organizational behaviour) and micro-/meso–macro-level constructs (e.g., individual/organizational behaviour → system efficiency).

REFERENCES

Abbott, M., & Doucouliagos, C. (2003). The efficiency of Australian universities: A data envelopment analysis. *Economics of Education Review*, 22(1), 89–97. https://doi.org/10.1016/S0272-7757(01)00068-1

Aghion, P., Dewatripont, M., Hoxby, C., Mas-Colell, A., & Sapir, A. (2007). Why reform Europe's universities? *Policy Brief 2007/04*, Bruegel, Brussels.

Allardt, E. (1990). Challenges for comparative social research. *Acta Sociologica*, 33, 183–193. https://doi.org/10.1177/000169939003300302

Alwin, D. F., Braun, M., Harkness, J., & Scott, J. (1994). Measurement in multi-national surveys. In I. Borg & P. Mohler (Eds.), *Trends and perspectives in empirical social research* (pp. 26–39). Berlin, Germany: de Gruyter.

Anderson, R. B. W. (1967). On the comparability of meaningful stimuli in cross-cultural research. *Sociometry*, 30(2), 124–136. https://doi.org/10.2307/2786222

ARWU. (2018). *Academic ranking of world universities—Methodology*. Retrieved from http://www.shanghairanking.com/ARWU-Methodology-2018.html

Benoit, K., & Wiesehomeier, N. (2009). Expert judgments. In S. Pickel, G. Pickel, H. J. Lauth, & D. Jahn (Eds.), *Methoden der vergleichenden Politik-und Sozialwissenschaft* (pp. 497–516). Wiesbaden, Germany: Neue Entwicklungen und Anwendungen.

Bleiklie, I. (2014). Comparing university organizations across boundaries. *Higher Education*, 67(4), 381–391. https://doi.org/10.1007/s10734-013-9683-z

Brady, H. E. (1985). The perils of survey research: Inter-personally incomparable responses. *Political Methodology*, 11(3/4), 269–291.

Braun, M., & Johnson, T. P. (2010). An illustrative review for detecting inequivalences. In J. A. Harkness, M. Braun, B. Edwards, T. P. Johnson, L. E. Lyberg, P. P. Mohler, … T. W. Smith (Eds.), *Survey methods in multinational, multiregional, and multicultural contexts* (pp. 375–394). Hoboken, NJ: Wiley.

Brunsson, N., & Sahlin-Andersson, K. (2000). Constructing organizations: The example of public sector reform. *Organization Studies*, 21(4), 721–746. https://doi.org/10.1177/0170840600214003

Chirikov, I. (2013). Research universities as knowledge networks: The role of institutional research. *Studies in Higher Education*, 38(3), 456–469. https://doi.org/10.1080/03075079.2013.773778

Clark, B. R. (1983). *The higher education system*. Berkeley, CA: University of California Press.

Collier, D., & Mahon, J. E. (1993). Conceptual 'stretching' revisited: Adapting categories in comparative analysis. *American Political Science Review*, 87(4), 845–855. https://doi.org/10.2307/2938818

Cooper, B., & Glaesser, J. (2011). Using case-based approaches to analyse large datasets: A comparison of Ragin's fsQCA and fuzzy cluster analysis. *International Journal of Social Research Methodology*, 14(1), 31–48. https://doi.org/10.1080/13645579.2010.483079

Davern, M. E. (2008). Representative sample. In P. J. Lavrakas (Ed.), *Encyclopedia of survey research methods* (pp. 720–721). London, UK: Sage Publications.

Ebbinghaus, B. (2005). When less is more: Selection problems in large-N and small-N cross-national comparisons. *International Sociology*, 20(2), 133–152. https://doi.org/10.1177/0268580905052366

Enders, J. (2004). Higher education, internationalisation, and the nation-state: Recent developments and challenges to governance theory. *Higher Education*, 47(3), 361–382. https://doi.org/10.1023/B:HIGH.0000016461.98676.30

Fanelli, D. (2009). How many scientists fabricate and falsify research? A systematic review and meta-analysis of survey data. *PLoS One*, *4*(5), e5738. https://doi.org/10.1371/journal.pone.0005738

Foster, I., Ghani, R., Jarmin, R. S., Kreuter, F., & Lane, J. (2016). *Big data and social science: A practical guide to methods and tools*. Chapman and Hall/CRC.

Frølich, N., Huisman, J., Slipersæter, S., Stensaker, B., & Bótas, P. C. P. (2013). A reinterpretation of institutional transformations in European higher education: Strategising pluralistic organisations in multiplex environments. *Higher Education*, *65*(1), 79–93. https://doi.org/10.1007/s10734-012-9582-8

Goedegebuure, L., & Van Vught, F. (1996). Comparative higher education studies: The perspective from the policy sciences. *Higher Education*, *32*(4), 371–394. https://doi.org/10.1007/BF00133253

He, J., & van de Vijver, F. (2012). Bias and equivalence in cross-cultural research. *Online Readings in Psychology and Culture*, *2*(2), 8. https://doi.org/10.9707/2307-0919.1111

Holland, P. W., & Wainer, H. (1993). *Differential item functioning*. Hillsdale, NJ: Lawrence Erlbaum.

Hood, C., James, O., Peters, B. G., & Scott, C. (2004). *Controlling modern government*. London, UK: Edward Elgar.

Huisman, J., Lepori, B., Seeber, M., Frølich, N., & Scordato, L. (2015). Measuring institutional diversity across higher education systems. *Research Evaluation*, *24*(4), 369–379. https://doi.org/10.1093/reseval/rvv021

Hüther, O., & Krücken, G. (2016). Nested organizational fields: Isomorphism and differentiation among European universities. In E. P. Berman & C. Paradeise (Eds.), *The university under pressure* (pp. 53–83). Bingley, UK: Emerald Group Publishing.

Johnes, J. (2006). Data envelopment analysis and its application to the measurement of efficiency in higher education. *Economics of Education Review*, *25*(3), 273–288. https://doi.org/10.1016/j.econedurev.2005.02.005

Johnson, T. P. (1998). Approaches to equivalence in cross-cultural and cross-national survey research. In J. Harkness (Ed.), *Cross-cultural survey equivalence* (pp. 1–40). Mannheim, Germany: Zentrum für Umfragen, Methoden und Analysen – ZUMA.

Jones, G. A., & Oleksiyenko, A. (2011). The internationalization of Canadian university research: A global higher education matrix analysis of multi-level governance. *Higher Education*, *61*(1), 41–57. https://doi.org/10.1007/s10734-010-9324-8

Jones, K., & Subramanian, S. V. (2012). *Developing multilevel models for analysing contextuality, heterogeneity and change*. Bristol, UK: Centre for Multilevel Modelling.

Kent, R. A. (2015). *Analysing quantitative data: Variable-based and case-based approaches to non-experimental datasets*. Thousand Oaks, CA: Sage.

Ketchen, D. J., Jr., Thomas, J. B., & Snow, C. C. (1993). Organizational configurations and performance: A comparison of theoretical approaches. *Academy of Management Journal*, *36*(6), 1278–1313.

King, G., Murray, C. J., Salomon, J. A., & Tandon, A. (2004). Enhancing the validity and cross-cultural comparability of measurement in survey research. *American Political Science Review*, *98*(1), 191–207. https://doi.org/10.1017/S000305540400108X

Kosmützky, A. (2015). In defence of international comparative studies. On the analytical and explanatory power of the nation state in international comparative higher education research. *European Journal of Higher Education*, *5*(4), 354–370.

Kreuter, F., Presser, S., & Tourangeau, R. (2008). Social desirability bias in CATI, IVR, and web surveys. The effects of mode and question sensitivity. *Public Opinion Quarterly*, *72*(5), 847–865.

Krücken, G., & Meier, F. (2006). Turning the university into an organizational actor. In G. S. Drori, J. W. Meyer, & H. Hwang (Eds.), *Globalization and organization: World society and organizational change* (pp. 241–257). Oxford, UK: Oxford University Press.

Lepori, B., Bonaccorsi, A., Daraio, A., Daraio C., Gunnes H., Hovdhaugen E., … Wagner-Schuster D. (2018). *Implementing and disseminating the European tertiary education register. Handbook for data collection*. Brussels, Belgium: European Union.

Li, M., Shankar, S., & Tang, K. K. (2011). Why does the USA dominate university league tables? *Studies in Higher Education*, *36*(8), 923–937. https://doi.org/10.1080/03075079.2010.482981

Lijphart, A. (1971). Comparative politics and the comparative method. *American Political Science Review*, *65*(3), 682–693. https://doi.org/10.2307/1955513

Maassen, P., & Olsen, J. P. (2007). *University dynamics and European integration*. Dordrecht, The Netherlands: Springer.

Mandelbrot, B. (1967). How long is the coast of Britain? Statistical self-similarity and fractional dimension. *Science*, *156*(3775), 636–638.

Marginson, S. (2006). Dynamics of national and global competition in higher education. *Higher Education*, *52*(1), 1–39. https://doi.org/10.1007/s10734-004-7649-x

Marginson, S. (2017). The world-class multiversity: Global commonalities and national characteristics. *Frontiers of Education in China*, *12*(2), 233–260. https://doi.org/10.1007/s11516-017-0018-1

Mathies, C., & Välimaa, J. (2013). Is there a need for a European institutional research? *Tertiary Education and Management*, *19*(1), 85–96. https://doi.org/10.1080/13583883.2012.747558

McAfee, A., & Brynjolfsson, E. (2017). *Machine, platform, crowd: Harnessing our digital future*. New York, NY: W. W. Norton & Company.

Mill, J. S. (1967) (orig. 1843). *A system of logic: Ratiocinative and inductive*. Toronto, Canada: University of Toronto Press.

Mounk, Y., & Foa, R. S. (2018). The end of the democratic century: Autocracy's global ascendance. *Foreign Affairs*, *97*, 29–36.

Musselin, C. (2007). Are universities specific organisations? In G. Krücken, A. Kosmützky, & M. Torka (Eds.), *Towards a multiversity? Universities between global trends and national traditions* (pp. 63–84). Bielefeld, Germany: Transcript.

Okazaki, S., & Sue, S. (1995). Methodological issues in assessment research with ethnic minorities. *Psychological Assessment*, *7*(3), 367–375. https://doi.org/10.1037/1040-3590.7.3.367

Padgett, J. F., & Ansell, C. K. (1993). Robust action and the rise of the medici, 1400–1434. *American Journal of Sociology*, *98*(6), 1259–1319. https://doi.org/10.1086/230190

Paradeise, C., & Thoenig, J. C. (2013). Academic institutions in search of quality: Local orders and global standards. *Organization Studies*, *34*(2), 189–218. https://doi.org/10.1177/0170840612473550

Peng, T. K., Peterson, M. F., & Shyi, Y.-P. (1991). Quantitative methods in cross-national management research: Trends and equivalence issues. *Journal of Organizational Behavior*, *12*(2), 87–107. https://doi.org/10.1002/job.4030120203

Przeworski, A., & Teune, H. (1966). Equivalence in cross-national research. *Public Opinion Quarterly*, *30*(4), 551–568. https://doi.org/10.1086/267455

Ragin, C. C. (2014). *The comparative method, moving beyond qualitative and quantitative strategies*. Berkeley, CA: University of California Press.

Reale, E. (2014). Challenges in higher education research: The use of quantitative tools in comparative analyses. *Higher Education*, *67*(4), 409–422. https://doi.org/10.1007/s10734-013-9680-2

Rostan, M., Finkelstein, M., & Huang, F. (2014). Concepts and methods. In F. Huang, M. Finkelstein, & M. Rostan (Eds.), *The internationalization of the academy* (pp. 23–36). Dordrecht, The Netherlands: Springer.

Sartori, G. (1970). Concept misformation in comparative politics. *American Political Science Review*, *64*(4), 1033–1053. https://doi.org/10.2307/1958356

Seeber, M., Cattaneo, M., Huisman, J., & Paleari, S. (2016). Why do higher education institutions internationalize? An investigation of the multilevel determinants of internationalization rationales. *Higher Education*, *72*(5), 685–702. https://doi.org/10.1007/s10734-015-9971-x

Seeber, M., Lepori, B., Montauti, M., Enders, J., de Boer, H., Weyer, E., … Reale, E. (2015). European universities as complete organizations? Understanding identity, hierarchy and rationality in public organizations. *Public Management Review*, *17*(10), 1444–1474. https://doi.org/10.1080/14719037.2014.943268

Seeber, M., Meoli, M., & Cattaneo, M. (2018). How do European higher education institutions internationalize? *Studies in Higher Education*, 1–18. https://doi.org/10.1080/03075079.2018.1541449

Short, J. C., Payne, G. T., & Ketchen, D. J., Jr. (2008). Research on organizational configurations: Past accomplishments and future challenges. *Journal of Management*, *34*(6), 1053–1079. https://doi.org/10.1177/0149206308324324

Singh, J. (1995). Measurement issues in cross-national research. *Journal of International Business Studies*, *26*(3), 597–619. https://doi.org/10.1057/palgrave.jibs.8490188

Smelser, N. J. (1976). *Comparative methods in the social sciences*. Englewood Cliffs, NJ: Prentice Hall.

Snijders, T., & Bosker, R. (2012). *Multilevel analysis*. London, UK: Sage Publications.

Straus, M. A. (1969). Phenomenal identity and conceptual equivalence of measurement in cross-national comparative research. *Journal of Marriage and the Family*, *31*(2), 233–239. https://doi.org/10.2307/349938

Teichler, U. (1996). Comparative higher education: Potentials and limits. *Higher Education*, *32*(4), 431–465. https://doi.org/10.1007/BF00133257

Thoenig, J. C., & Paradeise, C. (2014). Organizational governance and the production of academic quality: Lessons from two top US research universities. *Minerva*, *52*(4), 381–417. https://doi.org/10.1007/s11024-014-9261-2

Transparency International. (2013). *Global corruption barometer 2013*. Retrieved from https://www.transparency.org/gcb2013/in_detail

Välimaa, J., & Nokkala, T. (2014). The dimensions of social dynamics in comparative studies on higher education. *Higher Education*, *67*(4), 423–437. https://doi.org/10.1007/s10734-013-9684-y

Verba, S. (1967). Some dilemmas in comparative research. *World Politics*, *20*(1), 111–127. https://doi.org/10.2307/2009730

Whitley, R. (2000). *The intellectual and social organization of the sciences*. Oxford, UK: Oxford University Press.

Witte, K. D., & López-Torres, L. (2017). Efficiency in education: A review of literature and a way forward. *Journal of the Operational Research Society*, *68*(4), 339–363. https://doi.org/10.1057/jors.2015.92