

Research-driven classification and ranking in higher education: an empirical appraisal of a Romanian policy experience

Gabriel-Alexandru Vîiu¹  · Mihai Păunescu¹ · Adrian Miroiu¹

Received: 16 July 2015 / Published online: 2 February 2016
© Akadémiai Kiadó, Budapest, Hungary 2016

Abstract In this paper we investigate the problem of university classification and its relation to ranking practices in the policy context of an official evaluation of Romanian higher education institutions and their study programs. We first discuss the importance of research in the government-endorsed assessment process and analyze the evaluation methodology and the results it produced. Based on official documents and data we show that the Romanian classification of universities was implicitly hierarchical in its conception and therefore also produced hierarchical results due to its close association with the ranking of study programs and its heavy reliance on research outputs. Then, using a distinct dataset on the research performance of 1385 faculty members working in the fields of political science, sociology and marketing we further explore the differences between university categories. We find that our alternative assessment of research productivity—measured with the aid of Hirsch’s (Proc Natl Acad Sci 102(46):16569–16572, 2005) *h*-index and with Egghe’s (Scientometrics 69(1):131–152, 2006) *g*-index—only provides empirical support for a dichotomous classification of Romanian institutions.

Keywords Research evaluation · Higher education · Institutional classification · Institutional ranking · *h*-index · *g*-index

Introduction

Since the beginning of the 1980s nationally relevant university research coupled with the pressure for accountability have increasingly shaped the policies and priorities of individual universities (Geuna 2001). Since then, the growing importance of research has been

The present paper is a revised and expanded version of the work (Vîiu et al. 2015) previously published in the *Proceedings of ISSI 2015 Istanbul*.

✉ Gabriel-Alexandru Vîiu
gabriel.viiu@yahoo.com

¹ Department of Political Science, National School of Political and Administrative Studies, Povernei Street 6, 010643 Bucharest, Romania

continually underscored by transnational policy documents such as the EU 2020 Strategy, by implementation of performance-based research funding mechanisms which create new competitive pressures within national university systems (Hicks 2012) and, perhaps most visibly and controversially, by national and international university rankings which fuel debates surrounding ‘world-class universities’ (Sadlak and Liu 2007; Salmi 2009; Shin and Kehm 2013). It is now well established that “international rankings of universities have become both popular with the public and increasingly important for academic institutions” (Buela-Casal et al. 2007, p. 351). At the same time rankings have also become “successful as an agenda-setting device for both politicians and for the higher education sector” (Stensaker and Gornitzka 2009, p. 132).

In this paper we present an empirical exploration of the research-driven ranking and classification processes directed toward the Romanian higher education institutions (henceforth “HEIs”) in the policy context of a new Law on National Education. In accordance with the new law a comprehensive process of evaluation was conducted in Romania in 2011 with the dual aim of (1) classifying HEIs (at the global, institutional level) and (2) ranking their constituent study programs. The ranking and classification were conducted using a *common methodology* that heavily emphasized the *research productivity* of university staff. Our primary objective is to contribute to a better understanding of the relation between the classification and ranking processes by discussing the methodological outline of the official evaluation and by analyzing its results. To achieve this goal we rely on official documents and on data collected with regard to the actual results of the classification and ranking processes. A secondary objective of our paper is to investigate the consistency of the institutional classification categories used in the official evaluation. To do this we employ an alternative dataset on research performance, measured using Hirsch’s (2005) well-known *h*-index but also with the *g*-index which—for the set of papers of an individual researcher—represents “the largest rank (where papers are arranged in decreasing order of the number of citations they received) such that the first *g* papers have (together) at least g^2 citations” (Egghe 2006, p. 144). Our goal is to investigate whether an alternative assessment of research based on such indices confirms the official classification of institutions which was largely determined by their research performance.

Background

Theoretical considerations

Higher education in recent years has witnessed the emergence of numerous university rankings which have been the focus of comprehensive studies that aimed to investigate their methodological underpinnings, theoretical outlook and practical consequences (e.g.: Dill and Soo 2005; Salmi and Saroyan 2007; Usher and Medow 2009; Rauhvargers 2011). In a more recent study Hazelkorn (2013) noted no less than 10 global rankings and at least 60 countries that have introduced national rankings. All these studies highlight (among other aspects) the fundamental importance that ranking systems generally attach to research performance, the deleterious consequences that rankings may have for institutional diversity and quality and, perhaps most importantly, the methodological caution which should be exercised when undertaking and interpreting rankings.

As more and more rankings have been developed over the years and as concerns have mounted regarding their implications and methodological problems (e.g.: van Raan 2005a;

Billaut et al. 2010; Longden 2011) the adjacent subject of university classification has also received increased attention (see for example Shin 2009). This has been the case especially at the broader European level where the international ranking impetus has been critically received by scholars and policymakers and carried forward in a new direction with the introduction of the U-Map and U-Multirank initiatives which, unlike pre-existing commercial rankings, focus on a user-driven approach and emphasize multidimensionality in evaluation.

Classification of universities has tended to be a much less debated subject than rankings but these two distinct processes are nonetheless naturally interwoven with each other. On the one hand, due to strictures of comparability “classification is a prerequisite for sensible rankings” (van der Wende 2008, p. 49) although some instances may also be found in which classifications are undertaken subsequent to (and within the boundaries of) ranking exercises (e.g.: Cheng and Liu 2006). On the other hand, classifications are often interpreted as rankings even though this is clearly against the intentions of the classifying agency. Shulman (2005) and McCormick (2008) provide several examples of how the Carnegie Classification of US HEIs is actually understood as a form of ranking by several types of stakeholders.

A useful analytical distinction made between classifications and rankings involves conceptualizing them in the context of the broader notion of institutional diversity which itself may be divided into vertical diversity and horizontal diversity. According to van Vught (2009) the former refers to differences between higher education institutions owing to prestige and reputation while the latter stems from differences in institutional missions and profiles. In light of this distinction, classifications are “eminently suited to address horizontal diversity” (van Vught and Ziegele 2011, p. 25) while rankings “are instruments to display vertical diversity in terms of performance by using quantitative indicators” (Kaiser et al. 2012, p. 888). A fundamental difference therefore seems to separate the two notions: while classifications ideally do not imply value judgements (i.e. separation on a continuum from ‘better’ to ‘worse’), rankings are the very epitome of such judgements which is why they are heavily contested.

Institutional diversity plays an important role in higher education policies around the world due to its status as “an inherent good” and is often the object of governmental policies which seek to maintain or even increase it (Huisman et al. 2007). However, at the European level one of the consequences of governmental intervention is that “institutional diversity is more often based on regulation than on the actual characteristics or performances of the institutions involved” (Huisman and van Vught 2009, p. 19). In other words, diversity is often specified *ex ante* by legal mandate, rather than *ex post* through analysis of HEIs and their activities. Shin (2009) provides a similar account of the prevalence of legal classification in the case of Korean universities and undertakes a distinct classification using hierarchical cluster analysis which he offers as an empirically-grounded alternative to predetermined benchmarks. As will be argued below, the Romanian experience of classification and ranking is a good example of the legalistic approach to diversity.

The Romanian policy of classification and ranking

In 2011, following the provisions of a new law on education a comprehensive national evaluation was conducted for the first time by the Romanian Ministry of Education with the aim of classifying all accredited HEIs and, additionally, of ranking all accredited study programs offered by the universities. This process was by far the most elaborate evaluation of the Romanian system of higher education and the first one to explicitly undertake an

official classification of HEIs and an official ranking of their study programs on the basis of quantitative indicators.

With regard to the classification process the law stipulated that all universities must be classified as belonging to one of the following three classes: (1) *universities focused on education, henceforth to be referred to as Type I universities*; (2) *universities focused on education and research, henceforth to be referred to as Type II universities* and (3) *universities focused on advanced research and education, henceforth to be referred to as Type III universities*. This would point toward a functional differentiation with regard to research capacity but the law also stipulated that the allocation of public funding was to be a function of the results of the classification process: Type I universities could only receive public funding for study programs at the bachelor level, those of Type II could receive funding for programs at both bachelor and master level, while only Type III universities were eligible to receive public funding for all programs (including Ph.D.).¹

With regard to the ranking of study programs, the law on education did not contain any detailed provisions. However, a subsequent government decision (789/03.08.2011) established five distinct hierarchical classes: A (high quality), B, C, D and E (poor quality). These *program ranking classes* should not be confused with the broader *university classes*, which is why use is made in the current paper of the labels *Type I, II* and *III* to designate university classes.

A detailed methodology for the classification and ranking processes was made public through Ministry of Education Order 5212/26.08.2012. This methodology outlined a complex system of criteria, performance indicators, variables and weights. At the most general level, *four common evaluation criteria were used for both classification and ranking purposes*: (1) research; (2) teaching; (3) relation to the external environment; (4) institutional capacity. The most important aspect in the evaluation process was the research performance—broadly assimilated with specific research outputs—of the staff working in the universities and/or the study programs under assessment. This is especially significant for our later use of the *h* and *g*-indices.

Operationally, the ranking and classification processes relied on a list of 60 ranking domains which clustered the various study programs of the universities. Although the main evaluation criteria had different weights across these 60 domains (for example research had a greater weight for physics, chemistry or geology than it did for law or sociology), a typical distribution of these weights was the following: research—0.50, teaching—0.25, relation to the external environment—0.20, institutional capacity—0.05.² Each evaluation criterion was further divided into indicators and variables, all with their own predefined weights used to calculate intermediate scores. For example, the research criterion, which due to its weight was the chief determinant of the overall results of the ranking and classification processes, was divided into four indicators covering (1) research output, (2) research funding, (3) international recognition and (4) Ph.D. programs. The research output indicator had a weight of 0.75 within the broader research criterion and it comprised the following variables: the number of publications in journals indexed in the ISI Web of Knowledge, the relative influence score of publications (derived from the article influence score published in Thompson Reuters'

¹ This provision was later annulled through a Government Emergency Ordinance adopted in December 2013.

² For all ranking domains belonging to natural sciences, mathematics, biomedical sciences and engineering—in total 33 out of the 60 domains—the research criterion had a weight of 0.60 whereas for most domains belonging to social sciences (for example political science or administrative science) it had a weight of 0.50 and for those within humanities (for example philosophy or history) it had a weight of 0.40.

Journal Citation Reports), the number of articles published in journals indexed in international databases, the number of books and book chapters published with national and international printing houses. Data for all variables had to be reported by each university for the period of 5 years preceding the time of the evaluation.

The ranking process of study programs entailed the calculation of an overall *aggregated index of ranking* (AIR) based on the four general evaluation criteria and their attached weights. As a final step in the ranking of a study program, its AIR was compared to the highest one obtained among all the similar study programs and, based on certain predefined intervals, it was finally assigned to one of the five ranking classes A–E. Similarly, for the more general purpose of university classification a separate *aggregated index of classification* (AIC) was calculated at the global level of each university. The AIC was a product of three factors: (1) a cumulative factor that combined the scores obtained for two specific research indicators (the relative influence score of publications in Thompson Reuters-indexed journals and the number of books published with international printing houses); (2) a second factor calculated as the sum of the scores obtained by each of the study programs organized by the HEI under assessment for the four general evaluation criteria mentioned in the previous paragraphs and (3) an indicator based on the confidence level given to the HEI by the Romanian Agency of Quality Assurance in Higher Education following its periodic evaluations.

Without going into further details, it must be stressed that the methodological outline of the official evaluation conducted for purposes of university classification actually had the general underpinning of a ranking. This has been alluded to by Andreescu et al. (2015), Miroiu et al. (2015) and Andreescu et al. (2012) and is primarily a consequence of the fact that the classification was based on the composite scores of university performance which were sorted in descending order and clustered in accordance with predefined thresholds. Moreover, the classification relied on the research scores obtained by the constituent study programs of the universities and, therefore, on the partial results of the ranking process of these programs. In effect research—especially the bibliometric indicators derived from Thompson Reuters databases—was the object of double counting, once at the individual level of the study programs and once more at the aggregated level of the HEIs. Based only on the analysis of the methodology used in 2011, it may be argued that the entire classification process was actually hierarchical in nature and that vertical, not horizontal differentiation was a foreseeable consequence not only at the level of study programs (where ranking was explicit) but also with regard to the more general level of universities (where ranking was disavowed in favour of the more neutral label of “classification”). However, the preceding argument is theoretical in nature and no empirical analysis has so far been undertaken with regard to the relation between the actual results of the classification and the results of the program rankings. In addition, no independent empirical test of the three university classification categories has been conducted, either relying on the performance indicators initially used by the Ministry, or on alternative measures of research performance. In the following paragraphs we will address both issues in an attempt to answer several questions related to the classification and ranking processes.

Research questions

Given the unique nature of the classification and ranking processes undertaken by the Romanian Ministry of Education several important aspects invite questioning and empirical study. We will confine our analyses to the following:

1. Did the overlap in methodology with the program rankings have empirically discernible consequences for the more general process of university classification? Is there a significant degree of association between particular types of universities and particular classes of study programs? If so, which programs are more common in which types of university?
2. Since the classification process relied heavily on research outputs, can an alternative assessment of the research productivity of universities confirm the threefold classification? Are there significant differences with regard to the research productivity of faculty members *between* the three university types? Furthermore, are there significant differences with regard to the research productivity of faculty members *within* the three university types?

The first set of questions addresses the official university classification and the study program ranking processes in tandem and implies an investigation of data on the official results. The second set of questions only addresses the university classification process and will be explored using a distinct approach which will be described in the subsequent section.

Methodology

In order to investigate our first set of research questions we created a comprehensive dataset of the results of the ranking process for all the study programs evaluated in 2011. We then added the results of the classification of universities in order to obtain a final dataset comprising all the study programs, the ranking class in which they were placed following the evaluation process and the class in which the university managing them was placed following the separate evaluation for classification. This primary dataset contains 1056 observations of distinct study programs. To test for the level of association between ranking and classification results we created contingency tables for the occurrence of particular study programs (i.e. ranked in class A, B, C, D, E) in the three types of universities (i.e. Type I, Type II and Type III). Additionally, a Chi-squared test was also used to investigate the association between the classification and ranking categories.

To explore the second set of research questions we used a distinct dataset composed of information on 1385 Romanian faculty members active in the fields of political science, sociology and marketing. Specifically, we used their *h* and *g*-index to conduct an alternative assessment of university research output. These 1385 staff members represent the full populations of staff employed in political science, marketing and sociology study programs and they are spread out across 64 departments (study programs) and 34 distinct universities. Information on the identity of the staff members was obtained from the Romanian Agency of Quality Assurance in Higher Education (ARACIS) and, for each of the staff members in this second dataset, the *h* and *g*-index were extracted using Anne Harzing's Publish or Perish software (Harzing 2007) following a procedure previously employed in Vîiu et al. (2012) in an examination of political science departments. Specifically, once the academic population of each individual department was determined based on the lists provided by ARACIS, the name of each individual was queried in the Publish or Perish software and the raw values of the *h* and *g*-index were extracted following careful investigation of the results and subsequent to the removal of duplicate entries and redundant items. This data collection process was carried out between October 2012 and April 2013, first for the academic staff belonging to political science programs,

then for those belonging to sociology and, finally, for those engaged in marketing programs.

In keeping with standard practice used to report the collection of bibliometric data we need to render explicit some methodological options which may be conceived as technical limitations and which should be borne in mind when interpreting the results. First, self-citations were not excluded in the process of calculating the h and g indices, partly because this could only have been achieved through an enormously time-consuming manual process, partly because we adopted the assumption that self-citations are randomly occurring events that would not significantly distort the final results given the scale of the dataset. A similar limitation is that fractional counting was not employed to distinguish single-authored papers from multiple-authored ones, thereby allowing the latter contributions to be factored into the calculation of each individual index fully. A final option we must mention is the reference window selected for evaluation: since avoiding short time windows is a sensible recommendation often reiterated in the literature on bibliometric indicators (e.g.: van Raan 2005b; Glänzel and Henk 2013) and since our aim was that of evaluating in a global manner the entire research output of the human resources engaged in the Romanian universities, we allowed the entire output of any given scholar to contribute to his or her indices. We thus opted for an inclusive approach which takes into account the full achievements of each staff member, not only those of a short and potentially arbitrary window.

A further important methodological choice we wish to clarify regards the very indicators selected to conduct the alternative research assessment. Since there are numerous bibliometric indicators which we could have used—some with more desirable properties than the ones we selected—it is important to explain that we opted for the h and g -indices owing to considerations related to a specific policy context relevant for Romanian higher education: subsequent to the ranking and classification processes of 2011 the National Council for Higher Education Funding—a subordinate body of the Ministry of Education responsible with the funding of public universities—incorporated into its funding methodology the results of these processes. However, it later devised a new methodology for the allocation of funding to universities, one which incorporated the h -index as an important indicator to assess research performance.³ The new funding methodology is being piloted in 2015 and is to be effectively implemented in 2016. Given the interest in the h -index as a research evaluation mechanism and its intended use in the allocation of funding to universities we also decided to use this particular indicator in our assessment. Additionally, we opted to also use the g -index because it has a higher capacity to discriminate between average scientists (Schreiber 2008), a property which is desirable when comparing the overall research performance of multiple units of analysis.

With regard to this secondary dataset, the results of the official classification of Romanian HEIs would imply that there are significant differences between the staff employed in the three university types with respect to their h -index and, even more so, with respect to their g -index. To test this we employ analysis of variance and subsequent Tukey HSD tests to reveal the instances where differences between g -indices are significant. However, given the fact that the distributions of h and g -index values violate the

³ According to the methodology the h -index is to be calculated based on the following databases for academics belonging to natural sciences, mathematics, biomedical sciences and engineering: Thompson Reuters' Web of Science, Elsevier's Scopus, Google Scholar. Different weights are attached to the h -index from each database: 0.5 for Web of Science, 0.3 for Scopus and 0.2 for Google Scholar. For academics in social sciences and humanities the h -index is to be calculated only from Google Scholar.

assumption of normality implicit in analysis of variance (which may nonetheless be robust even in such cases), we also conduct nonparametric tests (the Wilcoxon rank sum test and Kolmogorov–Smirnov test) to further investigate the differences between staff types. Though both of these nonparametric procedures are weak tests (in the sense that they only provide a limited amount of information), they can nonetheless confirm the non-identical distribution of the h and g -scores across the academic staff investigated in our study. More importantly, however, they are useful for the complementary operation, that of establishing that certain subgroups are in fact very similar in nature.

To investigate the consistency of the official university classification process we follow a two-step approach. First we compare the university classes globally, checking whether parametric and nonparametric procedures confirm the threefold classification at the level of the entire dataset. This entails running analyses of variance (and corresponding nonparametric tests) for the h and g -indices of all the 1385 staff members with regard to university type. Following this global comparison the analysis is refined at the level of academic titles held by the staff in order to determine whether or not there is a structural difference between the university classes. This process entails running separate analyses of variance within each of the four staff categories—assistants, lecturers, associate professors and full professors—to determine the degree of differentiation or similarity between university classes in a pairwise manner (for example, by comparing the g -index of lecturers from Type I universities to that of lecturers from Type II universities, by further comparing the g -index of the same lecturers from Type I universities to that of the lecturers from Type III universities and, finally, by comparing the index of lecturers from Type II universities to that of lecturers from Type III universities).

A final and essential methodological remark we wish to stress is the following: although the analyses presented in the following section are based on the raw h and g -index values their results have also been cross-validated by equivalent analyses based on three distinct normalization techniques. First, to account for differences between the three fields in our dataset we applied the statistical procedures detailed above to field-normalized values of the h and g indices. This first normalization approach is proposed by Kaur et al. (2013) and entails normalization of the h and g indices by dividing the raw values with the average ones within each of the three fields. This process of normalization is not without its shortcomings, mostly connected to the issue of establishing adequate reference sets for normalization (Bornmann and Leydesdorff 2014) but it remains “a useful way to accommodate for disciplinary differences” (Harzing et al. 2014). A second normalization procedure we employed was that of standardizing the h and g indices following academic titles. In other words, the raw scores were normalized at the level of each academic rank by dividing them with the corresponding averages (for example, the raw values of associate professors were normalized by dividing them with the mean index of this particular staff group). Finally, we also employed a third, more general and common technique, namely that of taking the square-root of the raw values of the indices in order to normalize their distribution (see also Costas and Bordons 2007).

We wish to stress that the results we obtained are *consistent regardless of the index values taken into account, whether raw, or normalized* according to either of the three procedures just mentioned.⁴ Because of this convergent validity we limit our presentation

⁴ There is of course one logical exception: in the case of index values normalized with reference to academic titles the intra-university tests of difference which compare index values between academic titles within each of the three university types are rendered obsolete. In the case of this specific analysis conducted under this particular normalization technique the differences in index values are naturally very small and statistically insignificant.

from the subsequent section only to the raw values of the *h* and *g*-index as these are more intuitive than any of the normalized variants and capture with greater clarity the differences and/or similarities between (as well as within) the three classes of universities.

Results and discussion

Relation between official ranking and official classification results

With regard to our first set of research questions a review of Table 1 and Fig. 1 indicates that Type I universities (i.e. those classified as being focused on education) have a limited number of top-performing study programs (90 ranked in classes A and B, i.e. 17 % of all study programs in this type of university) but cluster the most programs with middle and low performance (those ranked in classes C, D and E add up to 83 % of programs managed within Type I universities). On the other hand, Type III universities (focused on advanced education and research) hold a total of 185 study programs and 121 of these (over 65 %) are ranked in class A. Another 39 are ranked in class B (thus, over 86 % of the programs in this type of universities are ranked in classes A and B) and only less than 5 % belong to the lower performing classes D and E. Type II universities (focused on both education and research) have mixed results: out of a total of 344 study programs managed by these universities 189 (55 %) are ranked in classes A and B, 28 % are in class C and the remaining 17 % are ranked in C, D and E.

A more detailed study of the relationship between observed and expected count values of the different classes of study programs within each of the three university types is also instructive. This study indicates a negative association between programs ranked in classes A and B and Type I universities. A further negative association can also be observed with

Table 1 Contingency table of ranking classes of study programs and university types

University type	I—Education	II—Education and research	III—Advanced research	Row total
Class of study program in official ranking				
A	22 4.17 %	60 17.44 %	121 65.41 %	203
B	68 12.90 %	129 37.5 %	39 21.08 %	236
C	147 27.90 %	97 28.20 %	17 9.19 %	261
D	112 21.25 %	16 4.65 %	3 1.62 %	131
E	178 33.78 %	42 12.21 %	5 2.70 %	225
Column total	527 100 %	344 100 %	185 100 %	1056

Chi-square test of ranking classes of study programs and university types

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-square	495.43	8	0.000
N of valid cases	1056		

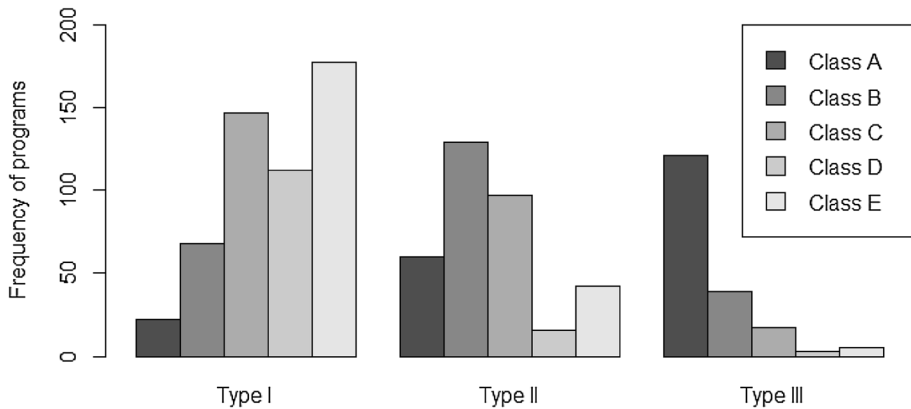


Fig. 1 Distribution of study programs (A–E) across the three university types

regard to programs ranked in classes A, D, and E and Type II universities. Finally, Type III universities are negatively associated with study programs ranked in classes B, C, D, and E. On the other hand, a positive association exists between Type I universities and study programs ranked in classes C, D and E. A further positive association exists between Type II universities and programs ranked in classes B and C. Type III universities are positively associated only with programs ranked in class A.

The results of this analysis paint a rather clear and polarized picture in which universities focused on education generally cluster study programs with poor performance while universities focused on advanced research cluster the programs with high performance. In addition, universities focused on advanced research are fewer and more selective (accounting for a total of only 185 study programs) as compared to universities focused on education (which manage a total of 527 programs). A certain hierarchy is implicit: universities focused on advanced research seem to be “better” than those focused on both education and research which, in turn, are “better” than those focused solely on education. However, as we mentioned earlier, these results were to be expected since both the classification and the ranking evaluation relied on a common methodology which was mostly concerned with research performance. This leads us to our second set of research questions.

Differences in research productivity across and within university types

We now move to explore whether our secondary dataset enables us to distinguish between three university types. In particular, what we want to see is whether the average h and g -indices of all academic staff in Type I universities are significantly lower than the average h and g -indices of staff in Type III universities and also in Type II universities. Table 2 presents the summary statistics of the indices, taking as reference the university type and, separately, the three fields covered by our dataset. One may note that the differences between the three university types seem much more pronounced than the differences between the three academic fields which are quite modest both for the h and for the g -index. This indicates—at least for academics working in Romanian universities—that the field differences between political science, sociology and marketing are negligible and that the research output of Romanian scholars working in these three fields is very similar.

Table 2 Summary statistics of raw *h* and *g*-indices within university classes and academic fields

	Minimum	1st quartile	Median	3rd quartile	Maximum	Range	Mean	SD
Breakdown by university type								
<i>h</i> index								
Type I	0	0	0	1	5	5	0.70	0.98
Type II	0	0	1	1	11	11	0.95	1.39
Type III	0	0	1	3	15	15	1.97	2.28
<i>g</i> index								
Type I	0	0	0	1	8	8	0.88	1.36
Type II	0	0	1	2	16	16	1.29	2.15
Type III	0	0	2	4	24	24	3.00	3.99
Breakdown by academic field								
<i>h</i> index								
Political science	0	0	0	2	13	13	1.04	1.62
Sociology	0	0	1	2	15	15	1.27	2.05
Marketing	0	0	1	1	8	8	0.98	1.20
<i>g</i> index								
Political science	0	0	0	2	22	22	1.39	2.55
Sociology	0	0	1	2	24	24	1.85	3.44
Marketing	0	0	1	2	14	14	1.39	2.01

The ANOVA procedure applied to the raw *h* and *g* values in conjunction with university type yield the results presented in Table 3 (very similar results are obtained, as previously mentioned, under all of the three normalization techniques we employed to cross-validate the robustness of our findings). The subsequent Tukey HSD tests indicate significant differences in mean values between all three university types (although the confidence level for the Type I–Type II distinction is lower, but still above 95 %) and therefore seem to provide empirical ground for the threefold classification which was legally mandated in 2011.

If the nonparametric procedures detailed in “Appendix 1” are applied, however, a slightly different picture begins to emerge: although the Wilcoxon rank sum test still confirms a threefold university classification, the Kolmogorov–Smirnov test indicates that staff working in Type I and Type II universities are not significantly different in terms of their *h* or *g* indices, and, therefore, that these two university classes are not readily distinguishable from one another with regard to their research output. However, the results presented in Table 3 and in “Appendix 1” only provide information on the *global differences* between university types with regard to the *h* and *g*-indices of their entire staff, without further consideration of academic titles. Therefore, in order to test the consistency of the threefold model of classification imposed by the 2011 law, we must explore in greater depth the differences between universities, taking into account more granular differences between their academic staff. We thus set out to test not only the global aggregate differences, but also the *structural patterns* of the three types of universities, taking into account the academic titles of the teaching staff.

Bearing in mind the results of the official evaluation from 2011, we wish to know whether, for example, associate professors from Type I universities are significantly different from associate professors in Type II universities and from those belonging to Type

Table 3 ANOVA of *h* and *g*-indices (raw values) with regard to university class ($N = 1385$)

Model summary for ANOVA of <i>h</i> -index with regard to university type				
	Sum of squares	<i>df</i>	Mean square	<i>F</i>
Between groups	340.8	2	170.4	77.2
Within groups	3048.6	1382	2.2	Sig.
Total	3389.4	1384		0.000

Tukey HSD values for ANOVA of <i>h</i> -index with regard to university type				
Comparison	Difference	Lower bound	Upper bound	<i>p</i> value
Type II–Type I	0.247	0.031	0.462	0.020
Type III–Type I	1.268	1.026	1.511	0.000
Type III–Type II	1.022	0.762	1.281	0.000

Model summary for ANOVA of <i>g</i> -index with regard to university type				
	Sum of squares	<i>df</i>	Mean square	<i>F</i>
Between groups	952.6	2	476.3	81.6
Within groups	8064.9	1382	5.8	Sig.
Total	9017.6	1384		0.000

Tukey HSD values for ANOVA of <i>g</i> -index with regard to university type				
Comparison	Difference	Lower bound	Upper bound	<i>p</i> value
Type II–Type I	0.405	0.054	0.756	0.019
Type III–Type I	2.119	1.725	2.513	0.000
Type III–Type II	1.714	1.293	2.136	0.000

III and, still further, if the associate professors from Type II institutions are different from those from Type III. Similarly, we also wish to know whether assistants, lecturers and full professors from one type of university are different from those belonging to the other two types of universities. Based on such analyses we may draw more general conclusions regarding the degree of structural differentiation that exists between the three types of universities.

Due to the known high correlation between the *h* and *g*-index values (see for instance Bornmann et al. 2011) the following analyses rely only on the *g*-index of the academics in our dataset. We mention in passing, however, the following findings with regard to the relation between the *h* and *g* indices of the academic staff in our dataset: first, the correlation coefficient (Pearson's *r*) for the relation between the *h* and *g* indices recorded in our dataset has a value of 0.954 with a *p* value of 0.000 (95 % CI 0.950–0.959). Second, our analysis of the index values indicates that out of the 1385 academic staff investigated only 21.3 % have a *g*-index whose value is higher (by at least one unit) than that of their *h*-index. This is consistent with Egghe's logic for developing the *g*-index as a specific tool which rewards selective researchers who have a higher impact (more citations) but a lower productivity (fewer papers).

Figure 2 illustrates the distribution of the *g*-indices of the academic staff in our secondary dataset with respect to academic titles and also with regard to the type of university

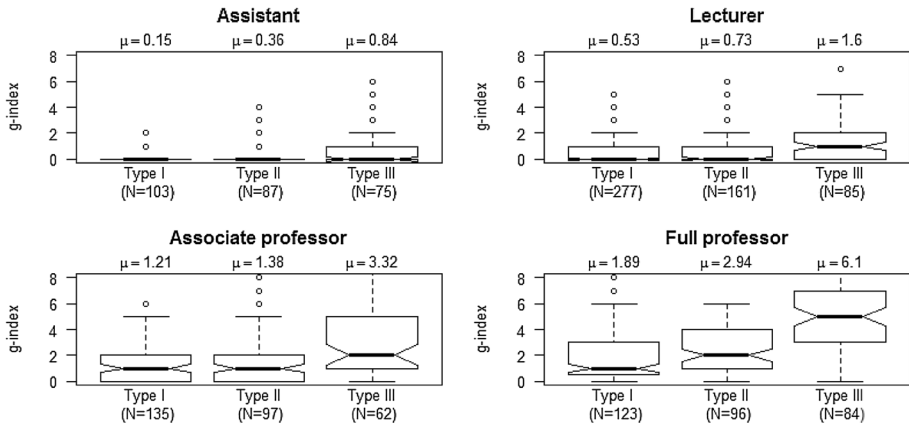


Fig. 2 Distribution of *g*-indices (raw values) by academic title and university type

they belong to. Mean values are presented in the upper sections as μ . An initial visual inspection of the data would seem to indicate that in the case of assistants, lecturers and even associate professors there are no substantial differences between Type I universities and those of Type II. On the other hand, all three staff types working in Type III universities seem to have substantially different *g*-indices compared to the ones from both Type I and Type II universities. A somewhat more nuanced picture emerges when looking at full professors. In this case the *g*-indices are more easily distinguishable between university classes and there indeed seem to be differences not only between Type III and the other two university types, but also between these two.

Based on the information contained in Fig. 2 and on the ANOVA procedures presented in Table 4 we may now answer our secondary research questions. In the case of all staff members (be they assistants, lecturers, associate or even full professors) the parametric statistical procedures show that universities classified within the official evaluation of 2011 as focused on advanced research (Type III) are indeed significantly different from the other two types. In other words, assistants, lecturers, associate and full professors working in these universities focused on advanced research have significantly higher *g*-indices than their counterparts from education-centred universities, as well as from those in universities focused on both research and education. Beyond the clear distinction of staff members working in Type III universities, statistical procedures also confirm something that Fig. 2 reveals in a more intuitive manner: virtually no statistically significant distinction can be made between Type I universities and Type II universities: assistant staff from Type I universities are in no way significantly different from assistant staff working in Type II universities, lecturers from one are in no way different from lecturers in the other and neither are associate professors. Even the apparent differences described by Fig. 2 between full professors from Type I universities and those from Type II universities do not seem to be statistically meaningful either, as can be observed in Table 4. Nonetheless, in the case of full professors belonging to Type I and Type II universities, a clear verdict might prove more elusive because only one of the nonparametric tests detailed in “Appendix 2” agrees with the findings of ANOVA. However, when looking at the overall picture there is compelling evidence that, for all human resources with the potential exception of full professors, Type I and Type II universities are not sufficiently different to have warranted distinct categorization in the 2011 classification process. In conclusion, when looking not only at the aggregate results across university classes, but also at the structure of universities, we find

Table 4 Tests of difference for *g*-index across academic titles and university types1. Model summary for ANOVA of *g*-index of assistant staff with regard to university type

	Sum of squares	<i>df</i>	Mean square	<i>F</i>
Between groups	20.681	2	10.341	13.29
Within groups	203.824	262	.778	Sig.
Total	224.506	264		0.000

Tukey HSD values for ANOVA of *g*-index of assistant staff with regard to university type

Comparison	Difference	Lower bound	Upper bound	<i>p</i> value
Type II–Type I	0.212	−0.090	0.515	0.225
Type III–Type I	0.685	0.369	1.000	0.000
Type III–Type II	0.472	0.145	0.800	0.002

2. Model summary for ANOVA of *g*-index of lecturers with regard to university type

	Sum of squares	<i>df</i>	Mean square	<i>F</i>
Between groups	73.699	2	36.850	25.39
Within groups	754.768	520	1.451	Sig.
Total	828.467	522		0.000

Tukey HSD values for ANOVA of *g*-index of lecturers with regard to university type

Comparison	Difference	Lower bound	Upper bound	<i>p</i> value
Type II–Type I	0.195	−0.086	0.476	0.233
Type III–Type I	1.062	0.711	1.413	0.000
Type III–Type II	0.867	0.487	1.247	0.000

3. Model summary for ANOVA of *g*-index of associate professors with regard to university type

	Sum of squares	<i>df</i>	Mean square	<i>F</i>
Between groups	204.795	2	102.397	24.44
Within groups	1219.205	291	4.190	Sig.
Total	1424.000	293		0.000

Tukey HSD values for ANOVA of *g*-index of associate professors with regard to university type

Comparison	Difference	Lower bound	Upper bound	<i>p</i> value
Type II–Type I	0.167	−0.475	0.808	0.814
Type III–Type I	2.108	1.368	2.848	0.000
Type III–Type II	1.941	1.157	2.725	0.000

4. Model summary for ANOVA of *g*-index of full professors with regard to university type

	Sum of squares	<i>df</i>	Mean square	<i>F</i>
Between groups	914.074	2	457.037	34.83
Within groups	3936.401	300	13.121	Sig.
Total	4850.475	302		0.000

Table 4 continued

Tukey HSD values for ANOVA of *g*-index of full professors with regard to university type

Comparison	Difference	Lower bound	Upper bound	<i>p</i> value
Type II–Type I	1.054	−0.108	2.216	0.084
Type III–Type I	4.213	3.005	5.421	0.000
Type III–Type II	3.159	1.885	4.434	0.000

little empirical ground for the threefold classification. This suggests a dichotomous classification would fit the data better than the threefold model imposed by law.

Furthermore, we can also observe that the statistically significant difference between Type I and Type II higher education institutions that we found in the global analysis of variance is due to the difference between the research performance of the most productive human resource of these institutions, namely the tenured professors.

So far we have argued that the data we have available clearly indicate significant *inter*-university differences (at least insofar as Type III universities are made up of staff with higher indices than both Type I and II universities). In order to further investigate the different contribution of each type of academic staff in our alternative assessment based on the *h* and *g* indices we now turn to an analysis of *intra*-university differences. We have a reasonable expectation that within research universities there is a greater gap between the four staff types with regard to their scientific productivity. In other words, within Type III universities we expect that the *g*-indices of assistants, lecturers, associate and full professors show greater dispersion than the corresponding indices of the equivalent staff that are employed in Type I and Type II universities. If we review the mean *g*-index values in Fig. 2 we can observe that they appear to confirm our expectation. Whereas in the case of Type I universities the gap between an average assistant and an average full professor is 1.74 and in the case of Type II universities this gap is 2.58, in Type III universities the difference is no less than 5.26.

These findings indicate that full professors in research-centred universities have a substantially larger scientific contribution in their fields of study, not only when compared to staff employed in Type I and Type II universities, but also in comparison to their colleagues from the same university class. This suggests more competitive selection mechanisms of highly qualified academic staff in the research-centred universities compared to the other two university classes. These more competitive selection mechanisms may actually explain the institutional differences. Additionally, these findings also confirm the increased contribution of the most productive human resources of the institutions, a phenomenon related to the known skewness of science (Seglen 1992; Albarrán et al. 2011; Ruiz-Castillo and Costas 2014). As noted earlier, the global ranking differences between Types I and Type II universities are mainly caused by the differences between the full professors, whereas for the rest of the academics in these institutions there do not seem to be statistically significant differences.

The overall *intra*-university differences among staff members are also shown to be significant by an analysis of variance where *g*-indices are tested across the four different staff categories within each distinct university class. More detailed results on this are available in “Appendix 3”. A notable fact about the information provided in the final appendix is that while it largely confirms *intra*-university differences among staff members there does seem to be one exception: lecturers and assistants are only significantly different

within Type I universities, whereas in Type II and Type III universities they do not seem to be statistically distinguishable from one another.

Concluding remarks

The boundaries between classification and ranking of higher education institutions are often hard to establish and it is even harder to properly communicate the differences to intended stakeholders. When classification and ranking processes are carried out simultaneously and using common criteria the task of disambiguation becomes virtually impossible and the risk that a classification is perceived as a ranking increases exponentially. In the case of the evaluation conducted in Romania in 2011 the boundaries between classification and ranking were weak from the very inception of these evaluation processes in the law on education.

The official methodology for classification and ranking further obscured the differences between the two due to its reliance on common criteria and indicators, most notably the research performance of academic staff employed by the HEIs. Furthermore, although best practice in the field of classification suggests that this process should be based on empirical data (McCormick and Zhao 2005; van Vught 2009) rather than arise from pre-determined legal categories, the Romanian policy was thoroughly legalistic in its conception and in its implementation. The empirically-driven approach, however, is followed by both the historically well-established Carnegie Classification of US educational organizations as well as by the more recent U-Map effort of classifying European universities. Recent work by Shin (2009), Ortega et al. (2011) and García et al. (2012) have also mainly focused on the empirical construction of classifications through the use of cluster analysis.

By analysing the official methodology we have shown that the classification of Romanian HEIs carried out in 2011 had the underpinning of a ranking. By further analysing the results of both the classification and ranking processes we have shown that there is a clear association between the outcomes of the global process of classification and those of the more specific process of program ranking: a polarized landscape thus emerges in which HEIs classified as focused on education cluster the overwhelming part of poor performing programs, while universities classified as focused on advanced research cluster the better part of the top performing programs.

The intermediate class of universities focused on both education and research presents mixed results. However, by conducting an alternative assessment of the research performance of the individual staff employed by Romanian universities in three fields of study we have shown that the threefold classification may not have a sufficiently robust empirical grounding, at least insofar as social sciences are concerned. By using the *h* and *g*-index as concise measures of research performance we have illustrated the fact that the intermediate universities focused on both education and research may not be sufficiently distinct from the universities focused on education and therefore this intermediate class might have a certain degree of redundancy.

The conclusion regarding the degree to which our alternative assessment confirms the official classification is however ultimately contingent on the level of aggregation taken as reference: when looking in our dataset of 1385 staff members only at the aggregate results across university classes we do find empirical grounding for the three classes defined in 2011. However, when analysing in greater detail the structure based on the academic titles

and positions, we find less empirical grounds for the threefold classification as most of the staff employed in Type I and Type II universities are virtually indistinguishable from one another (i.e. assistants, lecturers and associate professors). It is only full professors that seem to make a more substantial difference between Type I and Type II universities, thus narrowly substantiating a threefold classification which might otherwise well be a simpler dichotomous one. It is thus only at the top academic level of these institutions (tenured professors) that there seem to be significant differences between three types of universities, while for the rest of the academic staff our data cannot fully discriminate between three levels of institutional performance, but only between two (i.e. Type III HEIs on one hand, and Type II/Type I HEIs on the other).

Previous extensive studies on the quality of Romanian higher education (Păunescu et al. 2012; Vlăsceanu et al. 2011; Miroiu and Andreescu 2010) revealed the structural isomorphism of the Romanian higher education organizations. The undifferentiated set of standards that all institutions must comply with for purposes of accreditation and public funding led the institutions to adopt similar strategies for achieving these objectives. This is reflected in the poor differentiation and homogeneity of HEIs as shown by their similar scores in the external evaluation of the accreditation agency, similar missions, similar achievements on various performance indicators, etc. While the present paper finds empirical support for the vertical differentiation between advanced research universities (usually traditional, older universities) and the rest (more recent ones, including all private initiatives), the actual structures of the bulk of HEIs, including Type I and Type II universities, reveal more similarities than differences. These findings should of course be considered under the due caveat that our results are based only on data collected for social sciences.

Acknowledgments Financial support from the National Research Council (Grant Number PN-II-ID-PCE-2011-3-0746) is gratefully acknowledged by Gabriel Vîiu and Adrian Miroiu. The authors would also like to express their gratitude for the helpful comments and suggestions provided to them during the anonymous peer review process, both for the original paper submitted for the 2015 ISSI conference as well as for the subsequent manuscript submitted to *Scientometrics*.

Appendix 1

See Table 5.

Table 5 Results of nonparametric tests for comparison of *h* and *g*-indices between the three university types

Comparison	Wilcoxon rank sum test with continuity correction (test statistic <i>W</i> and <i>p</i> value)	Two-sample Kolmogorov–Smirnov test (test statistic <i>D</i> and <i>p</i> value)
<i>h</i> -index		
Type II–Type I	$W = 129,340, p \text{ value} = 0.013$	$D = 0.062, p \text{ value} = 0.267$
Type III–Type I	$W = 135,360, p \text{ value} = 0.000$	$D = 0.316, p \text{ value} = 0.000$
Type III–Type II	$W = 88,464, p \text{ value} = 0.000$	$D = 0.263, p \text{ value} = 0.000$
<i>g</i> -index		
Type II–Type I	$W = 128,410, p \text{ value} = 0.008$	$D = 0.068, p \text{ value} = 0.186$
Type III–Type I	$W = 136,030, p \text{ value} = 0.000$	$D = 0.312, p \text{ value} = 0.000$
Type III–Type II	$W = 88,538, p \text{ value} = 0.000$	$D = 0.244, p \text{ value} = 0.000$

Appendix 2

See Table 6.

Table 6 Results of nonparametric tests for comparison of *g*-indices between academic staff within the three classes of universities

Comparison	Wilcoxon rank sum test with continuity correction (test statistic <i>W</i> and <i>p</i> value)	Two-sample Kolmogorov–Smirnov test (test statistic <i>D</i> and <i>p</i> value)
Assistant staff		
Type II–Type I	$W = 4008, p \text{ value} = 0.067$	$D = 0.096, p \text{ value} = 0.780$
Type III–Type I	$W = 2654.5, p \text{ value} = 0.000$	$D = 0.294, p \text{ value} = 0.001$
Type III–Type II	$W = 2592, p \text{ value} = 0.007$	$D = 0.199, p \text{ value} = 0.083$
Lecturers		
Type II–Type I	$W = 20,764, p \text{ value} = 0.160$	$D = 0.078, p \text{ value} = 0.572$
Type III–Type I	$W = 7936, p \text{ value} = 0.000$	$D = 0.278, p \text{ value} = 0.000$
Type III–Type II	$W = 5070.5, p \text{ value} = 0.000$	$D = 0.201, p \text{ value} = 0.023$
Associate professors		
Type II–Type I	$W = 6286.5, p \text{ value} = 0.589$	$D = 0.060, p \text{ value} = 0.988$
Type III–Type I	$W = 2217, p \text{ value} = 0.000$	$D = 0.362, p \text{ value} = 0.000$
Type III–Type II	$W = 1705, p \text{ value} = 0.000$	$D = 0.390, p \text{ value} = 0.000$
Full professors		
Type II–Type I	$W = 4958, p \text{ value} = 0.038$	$D = 0.182, p \text{ value} = 0.055$
Type III–Type I	$W = 1920.5, p \text{ value} = 0.000$	$D = 0.537, p \text{ value} = 0.000$
Type III–Type II	$W = 2117, p \text{ value} = 0.000$	$D = 0.376, p \text{ value} = 0.000$

Appendix 3

See Table 7.

Table 7 Intra-university comparison of *g*-indices across staff types

1. Tests of difference for *g*-index across academic titles within Type I universities

Model summary for ANOVA

	Sum of squares	<i>df</i>	Mean square	<i>F</i>
Between groups	228.184	3	76.061	50.72
Within groups	950.763	634	1.500	Sig.
Total	1178.947	637		0.000

Tukey HSD values

Comparison	Difference	Lower bound	Upper bound	<i>p</i> value
Associate professor–Assistant	1.059	0.647	1.472	0.000
Lecturer–Assistant	0.383	0.019	0.747	0.035

Table 7 continued

Tukey HSD values

Comparison	Difference	Lower bound	Upper bound	<i>p</i> value
Professor–Assistant	1.739	1.318	2.160	0.000
Lecturer–Associate professor	−0.677	−1.008	−0.346	0.000
Professor–Associate professor	0.679	0.286	1.073	0.000
Professor–Lecturer	1.356	1.015	1.698	0.000

2. Tests of difference for *g*-index across academic titles within Type II universities

	Sum of squares	<i>df</i>	Mean square	<i>F</i>
Model summary for ANOVA				
Between groups	388.628	3	129.543	34.53
Within groups	1639.372	437	3.751	Sig.
Total	2028.000	440		0.000

Tukey HSD values

Comparison	Difference	Lower bound	Upper bound	<i>p</i> value
Associate professor–Assistant	1.014	0.276	1.751	0.002
Lecturer–Assistant	0.365	−0.300	1.030	0.490
Professor–Assistant	2.580	1.841	3.319	0.000
Lecturer–Associate professor	−0.649	−1.291	−0.007	0.047
Professor–Associate professor	1.566	0.847	2.286	0.000
Professor–Lecturer	2.215	1.571	2.859	0.000

3. Tests of difference for *g*-index across academic titles within Type III universities

	Sum of squares	<i>df</i>	Mean square	<i>F</i>
Model summary for ANOVA				
Between groups	1333.936	3	444.645	38.10
Within groups	3524.064	302	11.669	Sig.
Total	4858.000	305		0.000

Tukey HSD values

Comparison	Difference	Lower bound	Upper bound	<i>p</i> value
Associate professor–Assistant	2.483	0.968	3.997	0.000
Lecturer–Assistant	0.760	−0.638	2.158	0.498
Professor–Assistant	5.267	3.865	6.669	0.000
Lecturer–Associate professor	−1.723	−3.196	−0.249	0.015
Professor–Associate professor	2.785	1.307	4.262	0.000
Professor–Lecturer	4.507	3.149	5.865	0.000

References

- Albarrán, P., Crespo, J. A., Ortuño, I., & Ruiz-Castillo, J. (2011). The skewness of science in 219 sub-fields and a number of aggregates. *Scientometrics*, 88(2), 385–397.
- Andresescu, L., Gheorghiu, R., Irimia, A., & Curaj, A. (2015). Mergers and classifications in Romania: Opportunities and obstacles. In A. Curaj, L. Georghiou, J. C. Harper, & E. Egron-Polak (Eds.), *Mergers and alliances in higher education. International practice and emerging opportunities* (pp. 33–55). Dordrecht: Springer.
- Andresescu, L., Gheorghiu, R., Proteasa, V., & Curaj, A. (2012). Institutional diversification and homogeneity in Romanian higher education: The larger picture. In A. Curaj, P. Scott, L. Vlăsceanu, & L. Wilson (Eds.), *European higher education at the crossroads: Between the Bologna process and national reforms* (pp. 863–885). Dordrecht: Springer.
- Billaut, J.-C., Bouyssou, D., & Vincke, P. (2010). Should you believe in the Shanghai ranking? An MCDM view. *Scientometrics*, 84(1), 237–263.
- Borrmann, L., & Leydesdorff, L. (2014). On the meaningful and non-meaningful use of reference sets in bibliometrics. *Journal of Informetrics*, 8(1), 273–275.
- Borrmann, L., Mutz, R., Hug, S. E., & Daniel, H.-D. (2011). A multilevel meta-analysis of studies reporting correlations between the *h* index and 37 different *h* index variants. *Journal of Informetrics*, 5(3), 346–359.
- Buela-Casal, G., Gutiérrez-Martínez, O., Bermúdez-Sánchez, M. P., & Vadillo-Muñoz, O. (2007). Comparative study of international academic rankings of universities. *Scientometrics*, 71(3), 349–365.
- Cheng, Y., & Liu, N. C. (2006). A first approach to the classification of the top 500 world universities by their disciplinary characteristics using scientometrics. *Scientometrics*, 68(1), 135–150.
- Costas, R., & Bordons, M. (2007). The *h*-index: Advantages, limitations and its relation with other bibliometric indicators at the micro level. *Journal of Informetrics*, 1(3), 193–203.
- Dill, D., & Soo, M. (2005). Academic quality, league tables, and public policy: A crossnational analysis of university ranking systems. *Higher Education*, 49(4), 495–533.
- Egghe, L. (2006). Theory and practise of the *g*-index. *Scientometrics*, 69(1), 131–152.
- García, J. A., Rodríguez-Sánchez, R., Fdez-Valdivia, J., Robinson-García, N., & Torres-Salinas, D. (2012). Mapping academic institutions according to their journal publication profile: Spanish universities as a case study. *Journal of the American Society for Information Science and Technology*, 63(11), 2328–2340.
- Geuna, A. (2001). The changing rationale for European university research funding: Are there negative unintended consequences? *Journal of Economic Issues*, 35(3), 607–632.
- Glänzel, W., & Henk, F. M. (2013). Opinion paper: Thoughts and facts on bibliometric indicators. *Scientometrics*, 96(1), 381–394.
- Harzing, A. W. (2007). Publish or Perish. <http://www.harzing.com/pop.htm>.
- Harzing, A.-W., Alakangas, S., & Adams, D. (2014). *h*_{1a}: An individual annual *h*-index to accommodate disciplinary and career length differences. *Scientometrics*, 99(3), 811–821.
- Hazelkorn, E. (2013). How rankings are reshaping higher education. In V. Climent, F. Michavila, & M. Ripolles (Eds.), *Los rankings universitarios: Mitos y realidades*. Ed. Tecnos.
- Hicks, D. (2012). Performance-based university research funding systems. *Research Policy*, 41(2), 251–261.
- Hirsch, J. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences*, 102(46), 16569–16572.
- Huisman, J., Meek, L., & Wood, F. (2007). Institutional diversity in higher education: A cross-national and longitudinal analysis. *Higher Education Quarterly*, 61(4), 563–577.
- Huisman, J., & van Vught, F. (2009). Diversity in European higher education: Historical trends and current policies. In F. van Vught (Ed.), *Mapping the higher education landscape. Towards a European classification of higher education* (pp. 17–38). Dordrecht: Springer.
- Kaiser, F., Faber, M., & Jongbloed, B. (2012). U-Map, university activity profiles in practice. In A. Curaj, P. Scott, L. Vlăsceanu, & L. Wilson (Eds.), *European higher education at the crossroads: Between the Bologna process and national reforms* (pp. 887–903). Dordrecht: Springer.
- Kaur, J., Radicchi, F., & Menczer, F. (2013). Universality of scholarly impact metrics. *Journal of Informetrics*, 7(4), 924–932.
- Longden, B. (2011). Ranking indicators and weights. In J. C. Shin, R. K. Toutkoushian, & U. Teichler (Eds.), *University rankings. Theoretical basis, methodology and impacts on global higher education* (pp. 73–104). New York: Springer.
- McCormick, A. (2008). The Complex interplay between classification and ranking of colleges and universities: Should the Berlin principles apply equally to classification? *Higher Education in Europe*, 33(2–3), 209–218.

- McCormick, A., & Zhao, C. (2005). Rethinking and reframing the Carnegie Classification. *Change*, 37(5), 51–57.
- Miroiu, A., & Andreescu, L. (2010). Goals and instruments of diversification in higher education. *Quality Assurance Review for Higher Education*, 2(2), 89–101.
- Miroiu, A., Păunescu, M., & Vîiu, G.-A. (2015). Ranking Romanian academic departments in three fields of study using the g-index. *Quality in Higher Education*, 21(2), 189–212.
- Ortega, J. S., López-Romero, E., & Fernández, I. (2011). Multivariate approach to classify research institutes according to their outputs: The case of the CSIC's institutes. *Journal of Informetrics*, 5(3), 323–332.
- Păunescu, M., Florian, B., & Hâncean, M.-G. (2012). Internalizing quality assurance in higher education: Challenges of transition in enhancing the institutional responsibility for quality. In A. Curaj, P. Scott, L. Vlăsceanu, & L. Wilson (Eds.), *European higher education at the crossroads: Between the Bologna process and national reforms* (pp. 317–338). Dordrecht: Springer.
- Rauhvargers, A. (2011). *Global university rankings and their impact*. Brussels: European University Association.
- Ruiz-Castillo, J., & Costas, R. (2014). The skewness of scientific productivity. *Journal of Informetrics*, 8(4), 917–934.
- Sadlak, J., & Liu, N. C. (2007). *The world-class university and ranking: Aiming beyond status*. Bucharest: UNESCO-CEPES.
- Salmi, J. (2009). *The challenge of establishing world-class universities*. Washington, DC: World Bank.
- Salmi, J., & Saroyan, A. (2007). League tables as policy instruments: Uses and misuses. *Higher Education Management and Policy*, 19(2), 31–68.
- Schreiber, M. (2008). An empirical investigation of the g-index for 26 physicists in comparison with the h-index, the a-index, and the r-index. *Journal of the American Society for Information Science and Technology*, 59(9), 1513–1522.
- Seglen, P. O. (1992). The skewness of science. *Journal of the American Society for Information Science*, 43(9), 628–638.
- Shin, J. C. (2009). Classifying higher education institutions in Korea: A performance-based approach. *Higher Education*, 57(2), 247–266.
- Shin, J. C., & Kehm, B. (Eds.). (2013). *Institutionalization of world-class university in global competition*. Dordrecht: Springer.
- Shulman, L. S. (2005). Classification's complexities. *The Chronicle of Higher Education B*, 52, 20.
- Stensaker, B., & Gornitzka, A. (2009). The ingredients of trust in European higher education?. In B. M. Kehm, J. Huisman, & B. Stensaker (Eds.), *The European higher education area: Perspectives on a moving target* (pp. 125–139). Rotterdam: Sense Publishers.
- Usher, A., & Medow, J. (2009). A global survey of university rankings and league tables. In B. M. Kehm & B. Stensaker (Eds.), *University rankings, diversity, and the new landscape of higher education* (pp. 3–18). Rotterdam: Sense Publishers.
- van der Wende, M. (2008). Rankings and classifications in higher education: A European perspective. In J. C. Smart (Ed.), *Higher education: Handbook of theory and research* (Vol. 23, pp. 49–72). Berlin: Springer.
- van Raan, A. F. J. (2005a). Fatal attraction: Conceptual and methodological problems in the ranking of universities by bibliometric methods. *Scientometrics*, 62(1), 133–143.
- van Raan, A. F. J. (2005b). Measurement of central aspects of scientific research: Performance, interdisciplinarity, structure. *Measurement*, 3(1), 1–19.
- van Vught, F. (2009). Diversity and differentiation in higher education. In F. van Vught (Ed.), *Mapping the higher education landscape. Towards a European classification of higher education* (pp. 1–16). Dordrecht: Springer.
- van Vught, F., & Ziegele, F. (Eds.). (2011). *Design and testing the feasibility of a multidimensional global university ranking. Final report*. Consortium for Higher Education and Research Performance Assessment, CHERPA-Network.
- Vîiu, G.-A., Vlăsceanu, M., & Miroiu, A. (2012). Ranking political science departments: The case of Romania. *Quality Assurance Review for Higher Education*, 4(2), 79–97.
- Vîiu, G.-A., Păunescu, M., & Miroiu, A. (2015). Research-driven classification and ranking in higher education. An empirical appraisal of a Romanian policy experience. In A. A. Salah, Y. Tonta, A. A. A. Salah, A. Sugimoto, & U. Al (Eds.), *Proceedings of ISSI 2015 Istanbul* (pp. 622–633). Istanbul: Boğaziçi University Printhouse.
- Vlăsceanu, L., Miroiu, A., Păunescu, M., & Hâncean, M.-G. (Eds.). (2011). *Barometrul calității 2010. Starea calității în învățământul superior din România*. Brașov: Editura Universității Transilvania din Brașov.