

Publisher version available on:

<https://www.tandfonline.com/doi/full/10.1080/1360080X.2020.1846243>

To cite this article:

Emanuel Kulczycki, Ewa A. Rozkosz, Krystian Szadkowski, Kinga Cierieszko, Marek Hołowiecki & Franciszek Krawczyk (2020) Local use of metrics for the research assessment of academics: the case of Poland, *Journal of Higher Education Policy and Management*, DOI: 10.1080/1360080X.2020.1846243

Local use of metrics for the research assessment of academics: The case of Poland

Emanuel Kulczycki,^{1*} Ewa A. Rozkosz,¹ Krystian Szadkowski,¹ Kinga Cierieszko,¹ Marek Hołowiecki,¹ Franciszek Krawczyk¹

¹ [Scholarly Communication Research Group](#), Adam Mickiewicz University in Poznań

* emek@amu.edu.pl

This article discusses the use of bibliometric indicators for the assessment of individual academics. We focused on national indicators for the assessment of productivity in Polish higher education institutions. We analysed whether institutions (N = 768) adopted national templates for their own sets of criteria for intra-institutional evaluations. This study combined an analysis of internal policy documents with semi-structured interviews with deans from institutions in different fields of science. Our findings showed that, despite their high levels of institutional autonomy, the majority of institutions adopted the national criteria for the evaluation of individual academics. This article concludes with recommendations for reducing the negative consequences of local use of national indicators for the assessment of researchers.

Keywords: local use; research evaluation, bibliometric indicators, metric fixation, assessment procedures, Poland

Introduction

The evaluation of higher education institutions (HEIs), academics, and scholarly publication channels has been one of the most widely discussed topics in research on higher education management in recent decades. On the one hand, the widespread increase in assessment exercises has been attributed to the emergence of an audit culture, the evaluative state, university rankings, and a growing emphasis on the financial accountability of educational institutions (Neave, 1998). On the other hand, various studies have highlighted that the higher

education and science landscapes can be transformed by the use of indicators—the key elements of almost all types of evaluation—and their abuses (Aagaard & Schneider, 2017).

Indicators are introduced, not only to monitor, control, and assess institutions and academics, but also to influence and change their everyday work practices and publication behaviours; for instance, Chinese universities introduced various monetary reward systems for papers published in journals indexed in the Web of Science, and such publications play a major role in China's tenure system (Shu et al., 2020). Incentive systems based on either monetary rewards or evaluation regimes can be found, among other countries, in Mexico, Indonesia, Poland, Ukraine, or Turkey (Nazarovets, 2020; Neff, 2018). Publication practices might change as intended by system design, but they can also be 'gamed'. One problem is that any indicator used in an evaluation system can put pressure on the evaluated actors, driving their decisions and consequently corrupting the indicator itself (Campbell, 1976). The gamification of metrics, to meet the expectations of evaluation systems at the lowest possible cost, is widespread in academia (Biagioli & Lippman, 2020).

Research assessments can be based on either expert opinion (e.g. peer reviews) or various metrics; however, the most common method of assessment is a combination of these two approaches. Metrics are useful, but they can be misused and abused like any other numerical indicator; for instance, if two researchers from different disciplines (e.g. history and biology) are evaluated against the same national criteria, solely according to the impact factor of their journal publications, the biologist is likely to be rated more highly, simply because of the database content used for calculations. The use of metrics for research assessment has therefore been widely criticised. In 2013, the San Francisco Declaration on Research Assessment (DORA; <https://sfedora.org/read/>) was published to improve the assessment of research outputs by funding agencies, academic institutions, and other parties. A general recommendation of DORA is to stop using journal-based metrics, such as the impact factor, to assess the quality of individual scientists' contributions in any type of evaluation decision. In 2015, this recommendation was followed by the Leiden Manifesto (Hicks et al., 2015), which highlighted that performance measures should be designed separately for the various research missions of different institutions, groups, and researchers. These two highly discussed policy documents on research assessment both emphasised one key principle: for the assessment of individual researchers, HEIs should not use metrics designed to evaluate publication channels (e.g. the impact factor) or institutions (e.g. metrics and point systems within national research evaluations, such as those used in Denmark, Finland, Norway, and Italy). This means that each unit of assessment (i.e. paper, journal, researcher, and institution), if metrics-based at all, should have its own distinctive set of indicators.

For the last three decades, performance-based research funding systems (PRFSs) have been used as science policy tools for research systems and HEIs aiming to increase their research effectiveness and allocate funding to most productive universities and academics. The main tools for assessing the performance of such systems are research evaluation tools, but not all research evaluation tools tie funding allocations to outputs and performance. The first of these systems was launched in the UK in 1986, and the current Research Excellence Framework was launched in 2014 as its successor (Sivertsen, 2017). Since then, many countries have introduced PRFSs, including Poland in 1991 and Australia in the mid-1990s (Kulczycki et al., 2017). One of the key elements of such tools is a national research evaluation framework by which institutions (i.e. faculties, departments, and disciplines within institutions) are assessed. The results of such assessments are used as inputs to the funding formulae for block grants provided by governments. The most important outputs assessed by such frameworks are scholarly publications, but there are exceptions (e.g. Australia since 2017). National lists of journals and

publishers have been used to indicate which publications can be counted in the evaluation of institutions in various countries (Pölönen et al., 2020). Only articles published in specific lists of journals or books published by listed publishers can be counted as outputs of a given institution.

National PRFSs affect local management practices within the higher education and research sectors, since their outcomes play at least a marginal role in institutions' budgets and are perceived as hallmarks of institutions' prestige. Local management practices, including the evaluation of academics, are often perceived as a link between national models (i.e. incentives) and the behaviours and research practices of individual researchers (Mouritzen & Opstrup, 2020), which may increase the temptation to adopt national criteria at the local level, using solutions designed for evaluating institutions to evaluate individual researchers. Such practices in research evaluation studies are referred to as *local use* (Aagaard, 2015; Hammarfelt et al., 2016; Mouritzen & Opstrup, 2020; Rijcke et al., 2016).

The evaluation of academics by the management in HEIs can be embedded in recruitment processes, tenure procedures, or periodic evaluations (Dobija et al., 2019). Most institutions design their own criteria and expectations (e.g. thresholds) to assess the academics they employ. On the one hand, the particular criteria and yardsticks used for such assessments are influenced by the general competitiveness experienced in a given sector (Musselin, 2018); for instance, competition in higher education is driven by the race for excellence and research universities focus their attention, not only on competition for students, but also on professors with top-tier publications in their portfolios. On the other hand, these criteria are not created in a vacuum, and management may consider the researchers' points of view when determining them. Some scholars have argued that the construction of such criteria is influenced by internal factors (e.g. the culture, structure, and leadership of a given institution) and external factors (e.g. isomorphic pressure to follow the paths of other institutions; Dobija et al., 2019).

Higher education and research institutions are formal organisations capable of completing resource-intensive tasks and achieving complex goals; hence, they need to adapt to changing regulations and contexts (e.g. new national research evaluation models), not only through local use but—in the light of institutional isomorphism and new institutionalism (DiMaggio & Powell, 1983; Dobija et al., 2019)—by becoming more like other institutions in their sectors. According to this perspective, institutions become increasingly similar over time and use similar tools to accumulate advantage and become highly ranked, in order to maximise their advantages (Dey, Milem, & Berger, 1997). This may also lead to the use of similar criteria for the evaluation of their academics.

Organisational institutionalism and isomorphism are reasons for institutions—to a great extent—to use the same bibliometric indicators and adopt national criteria. Meyer and Rowan (1977) showed that organisations are driven to incorporate institutionalised practices and solutions in order to increase their legitimacy and strengthen their prospects for survival. DiMaggio and Powell (1983) proposed three mechanisms of isomorphism (*coercive* stemming from political influences, *mimetic* resulting from standard responses to uncertainty, and *normative* associated with professionalisation) that drive the organisations to adopt 'rationalised myths'—institutionalised practices and solutions. This theoretical framework has been used to investigate transformations of various HE systems; for instance, Hammarfelt and Åström (2015) argued that Swedish HEIs' use of allocation models based on bibliometric measurements was a result of imitation and a consequence of operating under similar constraints. Cai (2010) investigated governance reform in the Chinese HE sector and showed how three types of isomorphism explained why and how global governance reform ideologies

have spread in the Chinese higher education system. Croucher and Woelert (2016) used institutional isomorphism to analyse changes in formal academic organisational structures in Australia.

In some countries, however, similarities of regulations might be relics of the past; for instance, periodic evaluation processes in Polish higher education and research institutions emerged from a long socialist tradition. Cocka (1980) analysed Soviet science policy, which significantly shaped the Polish academic sector, and noted that evaluation tended to be formal and highly structured. The real socialist government in Poland is a thing of the past; nevertheless, when investigating research evaluation practices, one should take into account the possibility that some isomorphism—such as local use—may result, not from adopting similar solutions, but from a longstanding heritage of deeply embedded institutional practices and habits in the science and higher education sectors. In other words, institutions in post-socialist countries may tend to use national-level indicators in their intra-institutional, local assessments because, historically, they were encouraged to do so.

This study focused on the periodic evaluation of academics in Polish HEIs. Specifically, we investigated whether indicators from the Polish PRFS ('national indicators') were adopted locally for the periodic evaluation of academics from 2009–2016. We also investigated how deans—as local managers responsible for designing periodic evaluation procedures—perceived the practice of adopting national criteria at a local level. Until recently, only a limited number of studies had investigated local use (Krog Lind, 2019; Schneider et al., 2016; Woelert & McKenzie, 2018).

This study aimed to answering the following research questions regarding the use of bibliometric indicators:

- (1) What are the characteristics of institutions using national and global indicators for the periodic evaluation of academics?
- (2) What are the patterns of local use?
- (3) What are the deans' attitudes towards research evaluation and local use?

We expected institutions that scored highly in research evaluations to focus on communicating their research results to an international audience and, therefore, use mostly global indicators (i.e. not copied from the national PRFSs), because the highest number of points in the national system are assigned for publications in the top-tier international journals. As Kulczycki et al. (2017) showed, high evaluation scores in Polish evaluations related to publications in top-tier international journals. We also assumed that institutions with lower scores would publish their results mainly through local and domestic scholarly communication channels. We thus expected such institutions to prefer national indicators, because they covered domestic journals and publishers. The results achieved by institutions in national research evaluations, as well as the size of the institutions, have a direct impact on the resources allocated to them. For many large institutions, these funds constitute the lion's share of all funds available for research. We therefore expected that large institutions would be likely to use national indicators to meet the criteria of the national PRFS and thus increase their chances of accessing resources (see Sauder and Espeland, 2009). Aagaard (2015) argued that, in the humanities and social sciences, there is a tendency to decouple organisational practices and the use of metrics. The origins of science metrics related to the so-called 'hard sciences' (e.g. the Science Citation Index in the 1960s), and many metrics are incompatible with the specifics of the so-called 'soft sciences'. The main reason for this disjuncture is that metrics do not include, for instance, scholarly book publications, which are essential communication channels in these disciplines. We therefore

assumed that the introduction of metrics as management tools would be less frequent in the humanities and social sciences than in the so-called ‘hard sciences’.

This paper contributes to the advancement of knowledge in the field of HEI policy and management, since it yielded results from large-scale research concerning assessment practices and local use of research evaluation criteria in Poland. This was accomplished by analysing regulatory documents for the periodic evaluation of academics, which allowed us to determine whether the patterns of local use were influenced by institutional size (expressed as full-time equivalent [FTE] researchers) and quality (expressed as the scientific category based on the evaluation).

Background and context

Poland has a large science and higher education system that has experienced a rapid process of massification and now supports universal access to higher education. In 1990, there were 403,000 students, whereas in 2005, the number had increased to almost two million. In 1992, there were 14 private HEIs, compared to 324 private HEIs in 2007 (Kwiek & Szadkowski, 2018). Presently, the system comprises more than 400 HEIs in both the public and private sectors, the largest of which are public ones such as the Warsaw University and the Jagiellonian University, while the private sector consists of micro or small institutions. Like other post-socialist Eastern European countries, Poland has a separate research organisation—the Polish Academy of Science—and numerous institutions in addition to university research institutions. This is a part of the country’s Soviet legacy. Another distinctive feature of the Polish system is the collegial character of its national governance structure, with the academic community having a substantial degree of autonomy at both the national and institutional levels. A further peculiarity of the Polish HE system is the fact that faculties represent the basic organisational unit in universities, which could be considered networks of loosely coupled faculties, rather than self-standing institutions. Until 2018, universities in Poland were treated by law as a federation of faculties. The true organisational and decision-making power and funds were assigned to faculties, rather than universities, which were more ‘brands’ than institutions. University faculties were autonomous in various fields and whole universities were often unregulated (e.g. regarding the evaluation of researchers). Polish national evaluations, until 2018, considered the faculty, not the university, as the main point of reference; therefore, in this paper, we refer to faculties as institutions.

The first version of a PRFS was introduced in Poland in 1990 (Kulczycki, 2017). Since then, the Polish PRFS has evolved through several cycles of four-year evaluation periods. In 1999, new regulations introduced metrics-based methods of evaluation and the assignment of specific numbers of points to various outputs, such as scholarly monographs, journal articles, and grants. At the same time, the State Committee for Scientific Research published the first edition of the Polish Journal Ranking, which is a national list of selected peer-reviewed journals, to support the research evaluation system. In 2010, the Ministry of Science and Higher Education introduced a new parametric model of evaluation, which was used, with some changes, for the 2013 and 2017 evaluations. These two evaluations were crucial from the perspective of this study, since they covered the period 2009–2016.

During the 2013 and 2017 evaluations, nearly 1,000 units (i.e. faculties, basic and applied research institutes, and institutes of the Polish Academy of Sciences) were evaluated. Each institution was evaluated according to four criteria: (1) scientific and creative achievements (e.g. monographs, journal articles, and patents), (2) scientific potential, (3) the material effects

of scientific activity (e.g. external funding), and (4) other effects of scientific activity. For each evaluated item, a given institution (i.e. faculty or institute) obtained a specified number of points. In this way, research output and productivity were translated into a key metric called *Polish points*. These points were used to rank institutions and assign scientific categories. The quality of a given institution was expressed by one of four categories: A+, A, B, or C. The best scientific units—those in the A+ category—received much greater financing (150 per cent of that allocated to the A category) for a period of four years. The financial consequences of the evaluation affected the block funds that are annually distributed to the scientific units. Since the evaluations are conducted every four years, assigned categories remain in place until the next evaluation. This means that the goal of attaining the highest scientific category may be a driving factor for many institutions to introduce evaluation systems based on national criteria.

In the analysed period (2009–2016), all Polish HEIs were obliged by the state to evaluate all their academic staff. Academics with doctoral or habilitation degrees were evaluated at least once every two years, and professors were evaluated at least once every four years (Act of 27 July 2005 Law on Higher Education). It is noteworthy that HEIs have been allowed to autonomously design their procedures for the evaluation of individual researchers. According to the state obligation, a periodic evaluation serves to assess whether academic staff members contribute to three areas of activity: research, teaching, and administrative and organisational work. A negative evaluation result could be grounds for the termination of employment (Act of 27 July 2005 Law on Higher Education).

Materials and methods

For the purpose of this research, we conducted a mixed-methods study comprising two phases. We investigated all institutions in both public and private universities that were evaluated during the 2017 national evaluation. In the first phase, we sent postal requests to the deans of the respective institutions ($N = 768$), asking them to share all the documents relating to the performance evaluations of the researchers employed by their institutions in the years 2009–2017. We also requested documents relating to academic promotions, but only a few institutions published and used such documents, and thus we did not include them in the analysis.

We received 244 responses (response rate = 31.7 per cent), of which 212 contained the documents that we requested. In total, we received 1,386 documents from 212 institutions (mean number of documents per institution = 6.5). The other 32 of the 244 responses contained only a cover letter ($N = 23$), stating that the dean refused to send the documents ($N = 7$; 5 of which were motivated by data protection concerns), or an empty envelope ($N = 2$). An empty envelope signified that such documents were not used by the respective institution, in accordance with the procedure described in our letter to the deans (see Appendix 1). The number of responses from HEIs in the private sector was low. Distinguishing them in a separate category for the analysis would likely yield invalid results; hence, we analysed the aggregate data for the public and private sectors, combined in the results section.

Table 1 shows the total number of institutions and the proportion of institutions included in this study according to Organisation for Economic Cooperation and Development (OECD) fields and the arts field, which is the seventh major field included in Polish higher education and evaluations. In total, we analysed the documents of 212 (28 per cent) out of 768 institutions. The highest proportions of analysed institutions were in the natural sciences (39 per cent) and

the medical and health sciences (38 per cent); the lowest proportions were in the arts (11 per cent).

Table 1. *The number of institutions (public and private sectors, combined) in Poland (N = 768) and the proportion of institutions included in the analysis.*

| OECD field | Total number of institutions in Poland | Number of analysed institutions (including private) | % |
|-----------------------------|--|---|-----|
| Social sciences | 208 | 65 (12) | 31% |
| Engineering and technology | 159 | 30 (0) | 19% |
| Arts | 99 | 11 (0) | 11% |
| Natural sciences | 96 | 37 (0) | 39% |
| Medical and health sciences | 79 | 30 (0) | 38% |
| Humanities | 79 | 27 (2) | 34% |
| Agricultural sciences | 48 | 12 (0) | 25% |
| Total | 768 | 212 (14) | 28% |

In the first phase of the study, the documents were coded according to two main categories: (1) use of Polish points (national indicators) and (2) use of bibliometric indicators (global indicators) to evaluate individual researchers. In the second phase, we conducted in-depth semi-structured interviews with selected deans ($N = 7$). All the interviewees worked in institutions that had used both national and global indicators to evaluate individual researchers. Interviews were conducted with representatives of each research field (i.e. humanities, science, medical sciences, natural sciences, agricultural sciences, economics and finance, and the arts) from seven different universities. We used multiple qualitative methods (triangulation of interviews with document analysis). The document analysis was conducted to help in selecting interviewees and the interviews served to deepen our understanding of the use of metrics in a given HEI, as well as to investigate the rationales underpinning local use. The aim of these interviews was to understand the deans' perceptions of formulating individual researcher evaluation criteria, since they were obligated by the state to introduce individual researcher evaluation systems in their HEIs. The interviews were conducted in Polish, and the average interview length was approximately 34 minutes. All the interviews were audiotaped, independently transcribed, and then processed in MaxQDA 12. The transcripts were coded, and categories were developed, refined, and validated. The coding scheme addressed the research questions, as well as the results of the document analysis in the first phase of the study. We focused on whether the criteria established by HEIs were, in the deans' opinions, adequate tools for individual researcher evaluation. We also focused on the role of the national research evaluation system in HEIs' formulation of individual research evaluation criteria. Both phases of this study were conducted between September 2018 and January 2020.

Results

Part I: Using national and global indicators

The documents used for the periodic evaluation of individual researchers consisted of sets of tables providing detailed information regarding publications, grants, courses, and organisational activities. All the analysed institutions followed the statutory obligations and

asked for evidence of outputs from three fields of activities: research, teaching, and organisational work.

In the quantitative document analysis, we focused on the two key themes that emerged from the qualitative analysis: the use of Polish points (national indicators) and the use of other bibliometric indicators (global indicators) the evaluation criteria.

Table 2 shows the proportions of institutions using national indicators and global indicators for the periodic evaluation of individual researchers, according to institution size in terms of FTE researchers, and the scientific categories assigned to institutions in the 2017 evaluations. We included 210 of 212 institutions in this analysis. Two institutions were excluded; one because no scientific category was assigned to it in 2017, and the other because it was the only institution in the C category, with 100–200 FTE researchers.

Table 2. *Institutions using national and global indicators (N = 210) by size and scientific category.*

| Institution size | Characteristic | Scientific category | | | |
|--|---|---------------------|-------|-------|-------|
| | | A+ | A | B | C |
| Over 200 FTE researchers | Number of institutions | 3 | 15 | 3 | N/A |
| | % of institutions using national indicators | 66.7% | 86.7% | 66.7% | N/A |
| | % of institutions using global indicators | 100.0% | 86.7% | 33.3% | N/A |
| Over 100 and up to 200 FTE researchers | Number of institutions | 11 | 35 | 31 | N/A |
| | % of all institutions using national indicators | 63.6% | 68.6% | 83.9% | N/A |
| | % of institutions using global indicators | 90.9% | 80.0% | 67.7% | N/A |
| Up to 100 FTE researchers | Number of institutions | 5 | 37 | 58 | 12 |
| | % of institutions using national indicators | 100.0% | 64.9% | 75.9% | 41.7% |
| | % of institutions using global indicators | 60.0% | 59.5% | 72.4% | 16.7% |
| Total | Number of institutions | 87 | 19 | 92 | 12 |
| | % of institutions using national indicators | 73.7% | 70.1% | 78.3% | 41.7% |
| | % of institutions using global indicators | 84.2% | 72.4% | 69.6% | 16.7% |

In total, 72.4 per cent of the institutions used national indicators in their evaluation procedures. All the smallest institutions in the highest category (A+) used national indicators. The lowest proportion of institutions (41.7 per cent) fell into the lowest category (C). With the exception of the institutions in the lowest category, at least 64 per cent of institutions in a given group (size/scientific category) used national indicators. Larger institutions within the A+ or A categories more often used global indicators than smaller institutions (up to 100 FTE researchers), whereas this pattern was reversed in the case of the institutions in the B category.

A different picture emerged when we analysed the documents according to the use of: (1) other indicators, such as the number of citations, the impact factor, and the h-index, and (2) bibliographic databases or lists, such as Web of Science, Scopus, Google Scholar, and Journal Citation Reports (all these indicators and databases are labelled in Figure 1 as global indicators). In total, 69.0 per cent of the institutions used global indicators.

All the largest institutions in the highest category (A+) used global indicators, whereas only 16.7 per cent of the institutions in the lowest category used these indicators and databases. In the two institutions in the highest categories (A+, A), institution size played a role in the patterns

of indicator use: a greater proportion of larger units (in terms of FTE researchers) used global indicators. By contrast, small scientific institutions (below 100 FTE researchers) in the B category often used global indicators.

Table 3 displays the proportions of institutions using national indicators and global indicators for the periodic evaluation of individual researchers according to their OECD fields (OECD, 2011) and the scientific categories assigned to institutions in the 2017 national evaluations. We included 208 institutions in this analysis. Four out of 212 institutions were excluded; one because no scientific category was assigned to it in 2017, and the other three because we determined that, for the group (field/scientific category), we would have to include at least three institutions to compare the proportion of institutions across the groups.

The results showed no clear patterns of national and global indicator use according to the OECD fields. One might expect that institutions from the so-called ‘soft’ fields (humanities and social sciences) would use these indicators less often; however, this was not the case for this study, since both national and global indicators were commonly used in all fields of science.

Table 3. *Use of national and global indicators by institution (N = 208) across OECD fields and scientific categories.*

| Institution size | Characteristic | Scientific category | | | |
|---|---|---------------------|--------|--------|-------|
| | | A+ | A | B | C |
| <i>Agricultural and veterinary sciences</i> | Number of institutions | N/A | 5 | 5 | N/A |
| | % of institutions using national indicators | N/A | 80.0% | 100.0% | N/A |
| | % of institutions using global indicators | N/A | 100.0% | 100.0% | N/A |
| <i>Arts</i> | Number of institutions | N/A | 5 | 6 | N/A |
| | % of all institutions using national indicators | N/A | 20.0% | 83.3% | N/A |
| | % of institutions using global indicators | N/A | 20.0% | 66.7% | N/A |
| <i>Engineering and technology</i> | Number of institutions | 2 | 11 | 12 | 4 |
| | % of institutions using national indicators | 50.0% | 36.4% | 83.3% | 25.0% |
| | % of institutions using global indicators | 100.0% | 90.9% | 75.0% | 0.0% |
| <i>Humanities</i> | Number of institutions | 2 | 15 | 9 | N/A |
| | % of institutions using national indicators | 100.0% | 86.7% | 66.7% | N/A |
| | % of institutions using global indicators | 100.0% | 80.0% | 55.6% | N/A |
| <i>Medical and health sciences</i> | Number of institutions | 2 | 14 | 14 | N/A |
| | % of all institutions using national indicators | 100.0% | 92.9% | 71.4% | N/A |
| | % of institutions using global indicators | 100.0% | 92.9% | 57.1% | N/A |
| <i>Natural sciences</i> | Number of institutions | 8 | 13 | 16 | N/A |
| | % of institutions using national indicators | 62.5% | 76.9% | 93.8% | N/A |
| | % of institutions using global indicators | 87.5% | 61.5% | 81.3% | N/A |
| <i>Social Science</i> | Number of institutions | 4 | 24 | 30 | 7 |
| | % of institutions using national indicators | 75.0% | 66.7% | 70.0% | 57.1% |
| | % of institutions using global indicators | 50.0% | 58.3% | 66.7% | 42.9% |
| <i>Total</i> | Number of institutions | 18 | 87 | 92 | 11 |
| | % of institutions using national indicators | 72.2% | 70.1% | 78.3% | 45.5% |
| | % of institutions using global indicators | 83.3% | 72.4% | 69.6% | 27.3% |

Figure 1 displays the number of institutions using specific indicators (national and global combined) for their evaluations according to their OECD fields. In each field, the most frequently used indicators were Polish points.

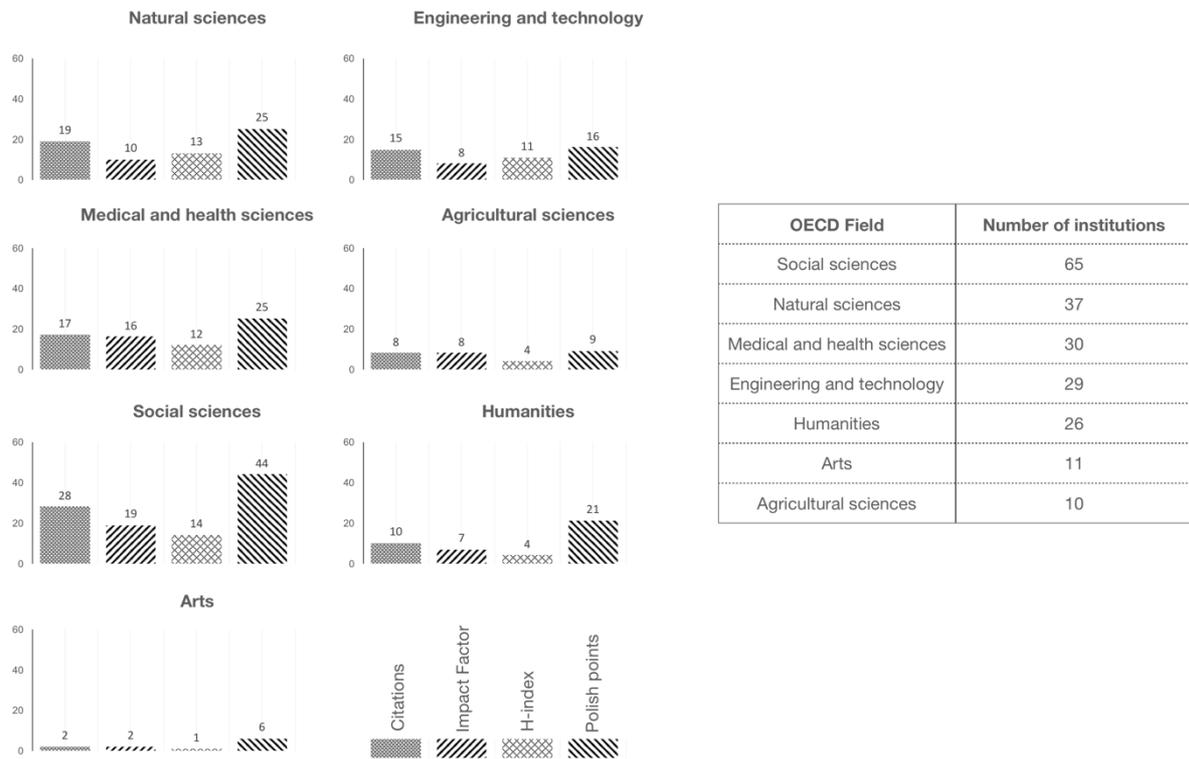


Figure 1. The number of institutions ($N = 208$) using indicators for their evaluations across OECD fields.

The results presented in Part I indicated that the majority of Polish institutions across all fields used both national and global indicators for the evaluation of researchers. It was clear that institutions assigned a higher scientific category, reflecting the quality of the institution in terms of evaluation results, tended to use indicators that were not based directly on the Polish research evaluation system, such as citations or the impact factor. Moreover, large institutions (in terms of FTE researchers) tended to use global indicators more often than small institutions, but this pattern was not so clear for the use of national indicators. In general, we concluded that institutions in higher evaluation categories tended to use global indicators more often than national indicators.

Part II: Patterns of local use

In the content analysis of the documents, we investigated the regulations used by all 152 institutions that applied national indicators. Firstly, we explored the materials to determine whether the institutions explicitly referred to the national research evaluation system, not only using Polish points or the Polish Journal Ranking, but also incorporating the results of the national evaluation exercise into their goals and values. We found that 15 institutions (9.8 per cent of those using national indicators) explicitly referred to the evaluation, as illustrated by the following quote: ‘The evaluation of this area is intended to reward those forms of scientific activity that contribute the most to the institution’s high score in the parametric evaluation and thus make it competitive among the best scientific institutions in the country’ (ID_207_Social_Sciences). All these institutions highlighted that employees’ publications were counted by the institution if they were counted in the national evaluation exercise. Surprisingly, 13 of these 15 institutions were from the social sciences ($N = 9$) and humanities ($N = 4$). However, one could argue that the national evaluations, which count publications in all languages in both national and local journals, may be a better reference for setting the goals of institutions in these fields than hard bibliometric indicators, which promote international visibility and mostly Anglophone journals.

Secondly, we investigated whether institutions used national indicators to set the thresholds for positive evaluation results. We found that 19 institutions across all fields (except the arts) determined a minimum number of Polish points needed to pass the threshold, while 13 institutions expressed the threshold as a minimum number of papers published in journals indexed in the Polish Journal Ranking.

Thirdly, we searched for field differences in local use and found that the examined institutions applied the same tools, expressed requirements in nearly identical ways, and used official terms from ministerial acts and regulations.

In summation, we found that the local use of national criteria for the periodic evaluation of academics, when it appeared, was similar in form and logic across all institutions. Although the examined institutions had the autonomy to shape their own procedures, all of them followed national solutions and imitated tools used by other institutions, with no consideration of specific disciplines.

Part III: Deans’ attitudes toward evaluation and local use

After coding the interviews with the deans, we initiated a second cycle of coding and revealed the four following themes: (1) actors involved in designing regulations of HEI evaluation procedures, (2) understanding of the evaluation criteria, (3) attitudes of the deans towards the evaluation of individual researchers, and (4) influence of national research evaluations on individual researcher evaluations.

Actors involved in designing regulations

All the interviewed deans told us that the evaluation procedures at their universities and faculties were designed in a top-down manner. This meant that the rector or dean appointed a commission, which designed the first draft. In the next phase, these drafts were discussed and revised, mostly by university authorities or full professors, as described by Interviewee 2: ‘The

first proposals came from the dean's circle and were sent to the employees by email. We collected suggestions, and these suggestions were either immediately implemented or recorded as alternative means of evaluation for discussion' (Interviewee 2).

Only one dean mentioned that representatives of the researchers (junior and senior) and trade unions were involved in the design of the first version of the evaluations. The interviewees mentioned that the final form of the evaluations was determined by consensus (i.e. 'the best it could be,' as described by Interviewee 2) reached between the authorities and the employees, who ultimately had to accept the rules according to which they would be evaluated.

Understanding of the evaluation criteria

All the deans confirmed that the areas for evaluation fully adhered to the statutory obligations and that their researchers were assessed according to their activities and outputs in terms of research, teaching, and administrative work. All the deans mentioned that student survey results were taken into account as an important factor; however, most of the interviewed deans stressed that, ultimately, the assessment of research activity was the most important and decisive element of the entire evaluation.

The deans also explained that no minimum thresholds were provided for their evaluation procedures. In other words, the researchers were not informed of what, precisely, they had to do to meet the evaluation criteria (e.g. no minimum number of publications was suggested). It was up to each researcher's supervisor to check whether the requirements were met, but these requirements were not explicitly communicated.

All the interviewed deans perceived the evaluation frameworks of their faculties as expert-based evaluations in which experts referred to various bibliometric and scientometric indicators. Two deans highlighted that these indicators and criteria made the evaluation an objective process but that, ultimately, the subjective opinion of experts (i.e., the researchers' supervisors) was decisive.

Bibliometric indicators were perceived by most of the deans as useful ways of assessing researchers, especially since this exercise is a time-consuming burden for a faculty and its employees (both academic and administrative). Bibliometric indicators, such as citations and the h-index, were treated as additional information, rather than as thresholds to be exceeded.

Attitudes of the deans towards the evaluation of individual researchers

In almost all the interviews, the deans highlighted that evaluation serves to determine which employees should be dismissed. Most of the deans argued that this characteristic of the evaluation process results from statutory guidelines that require such assessments to lead to one of two outcomes: a positive or negative assessment: 'From a dean's point of view, I believe that it is the primary tool that allows us to check the activities of the faculty's academic staff and to say goodbye to those who clearly have not met the basic criteria' (Interviewee 1).

Some of the deans mentioned that the legal guidelines lack a *conditional positive assessment* option, which could serve as a tool to motivate researchers. Other deans, however, admitted that this evaluation serves as an adequate motivator to encourage researchers to do the minimum work required.

All the deans perceived the evaluation as, not only a way of controlling and monitoring their employees and their entire institution, but as a useful tool for providing constructive feedback.

The deans also highlighted an evaluation burden in terms of hours spent by themselves, the researchers, and the administrative staff on reporting, analysing, and evaluating hundreds of researchers at the same time.

Influence of national research evaluations on individual researcher evaluations

All the deans pointed to the national research evaluation system as a reference for designing procedures for the evaluation of individual researchers. The deans also mentioned that researchers should be evaluated according to the criteria by which universities are evaluated, because the fate of their institutions depended on the researchers' work and activity. The evaluation of individual researchers was perceived by most of the deans as a way of achieving more points in the evaluation of their universities. Polish points were thus used in all the institutions in which the interviewees worked.

The deans perceived local use of the national system at their institutions as an aspect of 'playing the game', but it was not perceived as gaming. They seemed to believe they were forced to play this game because it was the best way to achieve better results in the evaluation of their universities.

Some deans mentioned that they initially wanted to modify the rules and avoid incorporating the national rules directly into their local individual evaluation systems; for instance, one dean mentioned that, in their faculty, articles not published in top-tier journals were not counted in the evaluations of individual researchers, even when they were counted in the evaluation of the faculty. This led the employed researchers to rebel against the imposed rules. Specifically, these researchers wanted to copy the entire set of national rules into the rules for their individual evaluations to include—as one of the deans put it— 'poor publications' in the assessments of their research activity. Researchers perceived the disallowing of the local use of the national rules as an unfair decision by the dean, because meeting the national expectations was easier than meeting the institutional ones.

Discussion and conclusion

This study revealed that, despite the high levels of autonomy granted to HEIs in Poland, the majority adhere to the national regulations and use the same indicators to evaluate individual researchers. In our opinion, the local use of national regulations is both (1) an effect of institutions' adaptations to the changing landscape of the higher education and science system and (2) inherited from past socialist methods of governing the research sector through central plans and regulations, with measures and aims for institutions designed at the national level.

Our findings highlighted that national criteria are often adopted and used for the periodic evaluations of individual researchers. In general, this practice can be observed in all fields (including the arts) and in all types of institutions, independent of size (FTE researchers) and performance (expressed in terms of scientific categories). Interviews revealed that the deans of HEIs in Poland generally perceived local use as a natural consequence of the simple existence of indicators at the national level. This situation pushes institutions, deans, and researchers into a game that cannot be avoided, because the structural conditions underpinning it are designed at the national level.

Adopting national evaluation criteria at the individual level is a common practice among Polish HEIs. This practice is not limited to any particular field, despite the fact that deans can make

autonomous decisions regarding individual assessment procedures and adjust criteria to individual fields. Moreover, deans tend to use similar arguments to rationalise their decisions. In particular, they believe that using national indicators, such as Polish points, in the assessment of employed researchers will lead to improvement of outcomes in the national evaluations. Furthermore, deans highlighted the need to use evaluation criteria that are acceptable to employees. From their points of view, decisions regarding the choice of criteria should be legitimised by their employees. We observed that the evaluation system exerts pressure that motivates HEIs to mimic each other; therefore, local use can be perceived as isomorphic. When an institution uses national indicators, it acts like other institutions in the higher education and science sector and, in this way (regardless of whether it is a good or bad solution), legitimises its own decisions.

Mechanisms relating to coercive isomorphism, stemming from the political and policy impact of national regulations on the universities and their faculties in Poland, could be a possible explanation for why institutions adopt the national indicators. This local use might be understood as a direct response to the government mandating the performance evaluations of individual researchers, as well as the national evaluation exercise. As DiMaggio and Powell (1983) showed, mimetic processes occur more often in organisations with large numbers of personnel and exert great pressure on the programs and services offered by such organisations. This could explain why the larger institutions in the A+ or A categories used global indicators more often than the smaller ones in our sample. Given that the official institutional goal is to gain legitimacy (external recognition), and not necessarily to increase the effectiveness (it is more important to follow the national rules and obligations than to design effective evaluation regimes at the institutional level), adaptation of institutional standards—in this case the use of national indicators—may function while, in general, being decoupled from actual organisation practices (Meyer & Rowan, 1977). Furthermore, Western European and British universities (ranked in the top 100 in various global university rankings) are used as examples of good practices in debates about the future of the Polish higher education sector. They serve as templates and models for the effective organisation of research work. This is where mimetic and normative pressures come into the picture and could explain the use of global indicators. Global bibliometric indicators are used to establish the positions of these Western universities in the rankings, as well as to enable Polish ministerial authorities and heads of faculties or universities to measure the distance between them and their Polish counterparts; therefore, mimetic imitation of their practices (seen through the prism of indicators only) is institutionally considered to be reasonable.

One of the main goals of an institution over the course of the evaluation procedure is to be placed in the highest scientific category. Given that national criteria have been and will be adopted, it is necessary to design them in such a way that they will not only be suitable, but also beneficial, on both the institutional and individual levels. The improvement of this situation and the reduction of local use would be possible if the policymakers and decision-makers responsible for the national research evaluations invited deans and researchers to fully participate in defining the criteria and phases of evaluation.

We suggest that three key solutions, if taken into account in the periodic evaluation of academics, can reduce the negative consequences of this form of assessment. Firstly, the evaluations should assess, not only how many papers a researcher publishes (or how many Polish points have been collected) during the evaluation period, but also what progress has been made since the last evaluation. Secondly, the results of the research evaluation should be

expressed, not only as a result on a grade scale, but also in the form of written feedback, with the institution encouraged to set the goals for the next evaluation period. Thirdly, if thresholds for positive grades are determined according to national indicators, they should be fitted to the given field and should incentivise a lower number of publications in high-quality scholarly publication channels, rather than simply a high number of publications.

Finally, a number of general, practical implications arise from our study and go beyond the scope of Polish, national evaluations and their structural dynamics. Every evaluation exercise with results based on the aggregation of individually met indicators (e.g. a specific number of particular publication types or publications in specific outlets) is either open to gamification or makes its subjects individual targets of institutional pressure. For this reason, we suggest—in line with the Leiden Manifesto (Hicks et al., 2015)—that, at the institutional level, specific, indicator-oriented targets should be formulated as team and/or unit targets, rather than individual ones. This should, however, be accompanied by promoting, and facilitating, a more collaborative approach to team management, the resolution of problems, and the creation of space for quality-driven peer assessment within the groups. We also observed that the universalisation of the bibliometric indicators used for assessment within different disciplines is an ongoing worldwide trend, which contributes to the homogenisation of norms, forms, languages, and values for particular disciplines. National policymakers, especially outside the Anglophone countries, should be particularly careful to keep their national systems flexible, rather than designing them to meet indicators over which they have no sovereign control. This is of particular importance for the humanities and social sciences, which are crucial for the preservation and stimulation of local cultures, languages, and traditions. We suggest developing discipline-sensitive national assessment methods for the humanities and social sciences that would align with the recent recommendations of the European Network for Research Evaluation in the Social Sciences (Ochsner et al. 2020), which state that:

An adequate research evaluation in the SSH (and beyond) corresponds to the research practices, makes its policy goals explicit, involves all stakeholders, ensures diversity of evaluation practices by respecting research's local and international embeddedness, uses a broad range of explicit quality criteria adequate for the discipline(s) under evaluation, evaluates each criterion separately, is based on informed peer review combining qualitative judgment and quantitative information, respects interdisciplinarity and does not prioritize some fields over others.

The evaluation of academics should serve, not only to achieve better results in the national evaluation of institutions, but to support the development of researchers and the quality of their work.

Funding

This work was financially supported by the National Science Centre in Poland (Grant Number UMO-2017/26/E/HS2/00019).

Declaration of interest statement

The authors declare no conflicts of interest.

References

- Aagaard, K. (2015). How incentives trickle down: Local use of a national bibliometric indicator system. *Science and Public Policy*, 42(5), 725–737. <https://doi.org/10.1093/scipol/scu087>
- Aagaard, K., & Schneider, J. W. (2017). Some considerations about causes and effects in studies of performance-based research funding systems. *Journal of Informetrics*, 11(3), 923–926. <https://doi.org/10.1016/j.joi.2017.05.018>
- Biagioli, M., & Lippman, A. (Eds.). (2020). *Gaming the metrics: Misconduct and manipulation in academic research*. MIT Press.
- Cai, Y. (2010). Global Isomorphism and Governance Reform in Chinese Higher Education. *Tertiary Education and Management*, 16(3), 229–241. <https://doi.org/10.1080/13583883.2010.497391>
- Campbell, D. T. (1976). Assessing the impact of planned social change. *Occasional Paper Series*, 8. [https://doi.org/10.1016/0149-7189\(79\)90048-X](https://doi.org/10.1016/0149-7189(79)90048-X)
- Croucher, G., & Woelert, P. (2016). Institutional isomorphism and the creation of the unified national system of higher education in Australia: An empirical analysis. *Higher Education*, 71(4), 439–453. <https://doi.org/10.1007/s10734-015-9914-6>
- Dey, E. L., Milem, J. F., & Berger, J. B. (1997). Changing Patterns of Publication Productivity: Accumulative Advantage or Institutional Isomorphism? *Sociology of Education*, 70(4), 308. <https://doi.org/10.2307/2673269>
- DiMaggio, P. J., & Powell, W. W. (1983). The Iron Cage Revisited: Institutional Isomorphism and Collective Rationality in Organizational Fields. *American Sociological Review*, 48(2), 147–160.
- Dobija, D., Górska, A. M., & Pikos, A. (2019). The impact of accreditation agencies and other powerful stakeholders on the performance measurement in Polish universities. *Baltic Journal of Management*, 14(1), 84–102. <https://doi.org/10.1108/BJM-01-2018-0018>
- Hammarfelt, B., & Åström, F. (2015). *The multi-layered and multilevel use of bibliometric measures in Swedish universities: Isomorphism, translation and strategic choice*. 10.
- Hammarfelt, B., Nelhans, G., Eklund, P., & Åström, F. (2016). The heterogeneous landscape of bibliometric indicators: Evaluating models for allocating resources at Swedish universities. *Research Evaluation*, 25(3), 292–305. <https://doi.org/10.1093/reseval/rvv040>
- Hicks, D., Wouters, P., Waltman, L., de Rijcke, S., & Rafols, I. (2015). Bibliometrics: The Leiden Manifesto for research metrics. *Nature*, 520(7548), 429–431. <https://doi.org/10.1038/520429a>

- Krog Lind, J. (2019). The missing link: How university managers mediate the impact of a performance-based research funding system. *Research Evaluation*, 28(1), 84–93. <https://doi.org/10.1093/reseval/rvy038>
- Kulczycki, E. (2017). Assessing publications through a bibliometric indicator: The case of comprehensive evaluation of scientific units in Poland. *Research Evaluation*, 26(1), 41–52. <https://doi.org/10.1093/reseval/rvw023>
- Kulczycki, E., Korzeń, M., & Korytkowski, P. (2017). Toward an excellence-based research funding system: Evidence from Poland. *Journal of Informetrics*, 11(1), 282–298. <https://doi.org/10.1016/j.joi.2017.01.001>
- Kwiek, M., & Szadkowski, K. (2018). Higher Education Systems and Institutions, Poland. *Encyclopedia of International Higher Education Systems and Institutions*, 1–10. https://doi.org/10.1007/978-94-017-9553-1_392-1
- Meyer, J. W., & Rowan, B. (1977). Institutionalized Organizations: Formal Structure as Myth and Ceremony. *American Journal of Sociology*, 83(2), 340–363. <https://doi.org/10.1086/226550>
- Mouritzen, P. E., & Opstrup, N. (2020). *Performance Management at Universities: The Danish Bibliometric Research Indicator at Work*. Springer International Publishing. <https://doi.org/10.1007/978-3-030-21325-1>
- Musselin, C. (2018). New forms of competition in higher education. *Socio-Economic Review*, 16(3), 657–683. <https://doi.org/10.1093/ser/mwy033>
- Nazarovets, S. (2020). Controversial practice of rewarding for publications in national journals. *Scientometrics*. <https://doi.org/10.1007/s11192-020-03485-7>
- Neave, G. (1998). The Evaluative State Reconsidered. *European Journal of Education*, 33(3), 265–284.
- Neff, M. W. (2018). Publication incentives undermine the utility of science: Ecological research in Mexico. *Science and Public Policy*, 45(2), 191–201. <https://doi.org/10.1093/scipol/scx054>
- Pölönen, J., Guns, R., Kulczycki, E., Sivertsen, G., & Engels, T. C. E. (2020). National Lists of Scholarly Publication Channels: An Overview and Recommendations for Their Construction and Maintenance. *Journal of Data and Information Science*, 0(0). <https://doi.org/10.2478/jdis-2021-0004>
- Rijcke, S. de, Wouters, P. F., Rushforth, A. D., Franssen, T. P., & Hammarfelt, B. (2016). Evaluation practices and effects of indicator use: A literature review. *Research Evaluation*, 25(2), 161–169. <https://doi.org/10.1093/reseval/rvv038>
- Schneider, J. W., Aagaard, K., & Bloch, C. W. (2016). What happens when national research funding is linked to differentiated publication counts?: A comparison of the Australian and

Norwegian publication-based funding models. *Research Evaluation*, 25(3), 244–256.
<https://doi.org/10.1093/reseval/rvv036>

Shu, F., Quan, W., Chen, B., Qiu, J., Sugimoto, C. R., & Larivière, V. (2020). The role of Web of Science publications in China's tenure system. *Scientometrics*, 122(3), 1683–1695.
<https://doi.org/10.1007/s11192-019-03339-x>

Sivertsen, G. (2017). Unique, but still best practice? The Research Excellence Framework (REF) from an international perspective. *Palgrave Communications*, 3, 17078–17078.
<https://doi.org/10.1057/palcomms.2017.78>

Woelert, P., & McKenzie, L. (2018). Follow the money? How Australian universities replicate national performance-based funding mechanisms. *Research Evaluation*, June, 1–12.
<https://doi.org/10.1093/reseval/rvy018>

Appendix 1

The template of the letter (translated from Polish) sent to all deans and heads of higher education and research institutions in Poland.

Dear Sir/Madame,

I run the Scholarly Communication Research Group at the Adam Mickiewicz University in Poznań. Within the framework of a grant financed by the National Science Centre, we are conducting research on the effects of the Polish research evaluation system. In order to better understand the publication patterns of Polish researchers, we are collecting documents relating to the periodic evaluation of academic staff members and the evaluation of candidates for scientific promotions.

In connection with this research, I take the liberty of asking you to send to our group all the internal documents that you used in your institutions in the years 2009–2017 for:

1. The periodical evaluation of academic staff members and
2. The evaluation of candidates for scientific promotions.

I kindly ask you to send electronic versions of these documents or paper versions in an enclosed, prepaid return envelope. I would appreciate it if you could send us these documents by 15 November 2018.

If your institutions have not used such documents, I kindly ask you to send back the empty (enclosed) envelope, which will also provide important information for our research.

We guarantee full confidentiality and anonymity of the information provided to us. I would also like to emphasize that the analysis of the submitted documents will be used to write scholarly publications, and the research results will be anonymized (i.e., the data of individuals and persons who made their documents available to us will not be disclosed).

I would like to thank you for all your help, which will allow us to conduct research and better understand the functioning of science in Poland.

In the case of any additional questions, please contact me.

Kind regards,

Appendix 2

Topics and sample questions for interviews

Category I. Opening question.

Topic 1. Definition of a good researcher.

1.1. What does it mean to be a good researcher at your faculty?

Category II. Participation in the development of evaluation criteria.

In accordance with the act on higher education, each faculty member carries out periodic evaluations of researchers at least once every four years. This means that the faculty is obliged to evaluate all employed researchers.

Topic 2. Knowledge of evaluation criteria.

2.1. What criteria were used to assess the researchers?

2.2. How do these criteria relate to the model of a good scientist at your faculty, which we discussed earlier?

2.3. How do you assess the criteria used by the faculty you work at?

Category III. Application of the criteria.

Topic 3. Active participation in the development of criteria.

3.1. How were the evaluation criteria and rules established?

3.2. Who was involved in developing the criteria for the periodic evaluation of researchers?

3.3. What do you think about the provisions on the periodic evaluation of academic staff in the Law on Higher Education? How did these provisions affect the work on the assessment form?

Category IV. Methods of assessment.

Topic 4. Bibliometric indicators.

Some faculties decide to use bibliometric indicators to periodically evaluate academic researchers.

4.1. What do you see as the benefits of using bibliometric indicators, such as numbers of publications, citations, and points, in the evaluation of individual researchers?

Topic 5. Reference values.

5.1. What requirements must a researcher meet to earn a positive assessment?

If the interviewee can explain:

5.2. How difficult do you think it is to meet these requirements?

If the interviewee can explain:

5.3. Does your faculty apply thresholds/minima points?

If yes:

5.4. Why did the faculty decide to introduce thresholds/minima points?

Category V. Function and effects of evaluation.

Topic 6. Evaluation function.

6.1. What is the purpose of the evaluation of researchers?

6.2. How did the faculty inform researchers of the results of the evaluation?

6.3. What should a researcher do in the case of obtaining a negative evaluation? What should they do in the case of obtaining a positive evaluation?

6.4. How often is a negative grade awarded at your faculty?

Conclusion.

7.1. How do you assess how the last researcher evaluation was conducted?

7.2. How do researchers assess this procedure?

7.3. What could be changed or improved in the rules for the periodic evaluation of researchers applied by your faculty?

7.4. Please provide a metaphor for the researcher evaluation process at your faculty.

7.5. Would you like to add something to our conversation that was not addressed by me?

*Those are all the questions. I am informing you that I am already turning off the recorder.
Thank you very much for your time and participation in our research.*