Introduction and context

Reflection serves as a circulation system for philosophy and a backbone for epistemology. Locke (1689) explores human understanding through the different manifestations of reflection – perceiving, thinking, doubting, believing, reasoning, knowing and willing. Reflection of scientific activity has evolved over the years from complete agnostics or relative doubts in the methods and integrity of individual researchers into comprehensive systems of science evaluation against sets of priorities and budgets. One may argue that this specific reflection was transformed from individual will and reasoning into an instrument of power.

The institutionalisation of science evaluation is intertwined with the development of the audit culture (Shore and Wright, 1999). The audit culture is viewed very differently from liberal and leftist perspectives. While the liberals would see audit culture as a progress in democratic governance, the leftists would see it as a strengthening of the status-quo and an instrument to control independent thinking.

Objectively, audit practices are a set of actions and control processes that are carried out by authorised control bodies within the framework of collected and analysed financial and non-financial information for the purpose of assessing the management of financial resources and the accountability of stakeholders with a view to achieving a potential improvement of the process.

It is a common belief that scientific audits and open science make science more transparent and thus more efficient in context of informing better policy decisions. Reichmann and Weiser's (2022) reflection on the science-policy relationships sheds doubt if deeper scrutiny is needed on the science part, but instead advocate widening of policy-making process. Policymakers seek information that is timely, relevant, credible, and available. Audit practices might and could contribute to these ends, however not without participatory engagement in policymaking.

Auditing has proliferated in virtually all spheres of social and economic life, not just the accounting and financial fields. It goes hand in hand with the emergence of new standards (i.e., environmental) and rising compliance costs.

Auditing practices have deep historical roots – from managing the relationship between landlords and peasants to modern corporations with dispersed ownership. Auditing is rooted in the need to have a sound system of checks and balances and perform various control activities. Independent auditing emerged and became increasingly sought after in the late nineteenth century. The United Kingdom, France, Germany, Italy, the Netherlands and Austria were among the first to impose legal regulations on this activity. Auditing was professionalised out of the need to guarantee that managers of corporations do not harm the interests of the state and those of the stakeholders. Classical financial audits were enhanced with performance audits as it became easier to manipulate the accounts over time.

Simultaneously, universities evolved significantly from a typical guild organisation (Medieval Bologna and Paris) the sole purpose of which is to produce educated people. They grew in number and size, changed the way recruitment of professors was performed (in 11th century it was the students' guild that appointed professors based on reputation and not on formal qualifications) and how universities were governed (a top-down corporate approach or bottom-up cooperative self-governance). The Bologna process, which was initiated in 1999, specifically focused on quality assurance as an integration instrument among different national higher education systems and individual European universities. For some observers quality assurance might come at the expense of academic freedom and independence, a former priority of the Bologna Declaration from 1988 (the Magna Charta Universitatum).

The "Standards and Guidelines for Quality Assurance in the European Higher Education Area"[1] was adopted in 2005, and was revised in 2015 by the European ministers of education. Although the standards were not meant to be applied to research per se, they do reflect the relationship between research and education Moreover, the underlying principles of quality assurance are the same as in every audit process and research evaluation, or to put it in a more abstract way – in every professionalised organised reflection: independence, objectivity, confidentiality, integrity and responsibility for the opinion expressed.

So, research evaluation developed together with the spread of the audit culture and accelerated due to integrative demands within the world of

1 http://www.ehea.info/media.ehea.info/file/2015_Yerevan/72/7/European_Standards_and_Guidelines_for_Quality_Assurance_in_the_EHEA_2015_MC_613727.pdf

higher education in the European Union. First in Britain in the early 1990s and then followed by other European countries, the elite higher education started to engage wider audiences expecting to reach half of secondary education graduates. Increasing the share of people with higher education is set as one of the main aims in the program documents of the EU. The Bologna process expanded the base by transforming the system of Fachhochshules (Germany, Austria, Cyprus and others) into universities of applied sciences which award bachelor and master degrees.

New universities flourished also in Eastern European countries after the political changes in 1989. The students enrolled in tertiary education in the EU have risen 1.5 times between 2000 and 2020, according to the World Bank data/UNESCO Institute of Statistics. Higher education institutions in Europe also mark increase in numbers, especially in new member states and associated countries during the last two decades (European Education and Culture Executive Agency, Eurydice, 2020).

These trends were accompanied by vast diversification of new educational programs and growing concerns about the overall quality of education. The new universities (often wrongly called red brick universities, as the term originated only for the civic universities in XIX century England) embraced the evaluation processes to increase public confidence. Yet, in certain countries the evaluation processes turned out to be highly bureaucratised and resulted in a self-replicating system.

So, when assessing the research assessments one should employ a cost-benefit approach. Do research assessments add value? To whom? Who pays the costs associated with them? Are they just a public cost, a fraction of all public investments in education and research or they are paid unevenly by some sub-group of researchers and universities?

The difference in academic and political "cultures" and “languages”, including the typical time-frames (longer horizons for researchers), knowledge and facts, even reputation mechanics, create a niche opportunity, where the evaluation practices, in various scope and format, could provide what both parties are looking for (Reichmann & Wieser, 2022).

At the same time research evaluation, as well as quality assurance in a larger higher education context, emerged as additional markets and source of income for key stakeholders of the system which is being evaluated. The current book studies this specific quasi-market of research/scientific assessments from diverse institutional perspectives. A contextual issue which drives our explorations is the complex relationship between the diffusion of audit culture in universities and the quality and interoperability of universi-

ties across different countries. The main institutional driver for that market is the Bologna process with the synchronization of bachelor, master and doctoral degrees, as well the standardisation of accreditation agencies to a certain degree.

Evaluation practices are relevant and additive, to a large extent, to the principles of good governance: openness, participation, accountability, efficiency, ethics and reasonable financial management, etc. The concept of governance is understood in many ways by different people. Its definition often depends on the objectives pursued, the actors involved and the socio-political environment in which these objectives are to be achieved, but the principles remain essentially unchanged.

Taking the position that the main mission of evaluation is to improve the internal research process, it can be further extrapolated that evaluation can be and is a necessary condition for the subsequent growth of international mechanisms, in which it is postulated that competitiveness is the first requirement to have access to financial instruments.

Historically, the development of research was initially supported only by wealthy individuals, churches or national resources and the issue of evaluation and control did not dominate the development policy. At a later stage, however, other financial flows in support of research entered – second and third – and were implemented either on a project-based competitive basis or at the request of a donor or creator.

After the second half of the last century, research evaluations developed and diversified. Not only did they have different goals, but they were aimed at different levels – local, national, transnational, that is. regional), European, trans-European. If one looks only at one type of evaluations or assessment, it would be difficult to understand or meta-assess its applicability and usability. Therefore, a comparative analysis of different evaluations would help us to better understand the very nature of the process, the motivations of those involved and the impact on the system.

At the beginning of the 1990s, all of the new EU member states were still implementing a science policy and evaluation mechanisms that were a continuation or a replica of the ones in the USSR. The Soviet audit culture in universities was exactly what Shore and Wright (1999) were referring to – a political structure for staff control, which assured a patronised career development nurtured by the party favouring loyal professors. The political institutionalisation was in the form of higher attestation committees (or VAKs), which had the power – the upper chamber – to stop or further any career development despite the assessment of the lower chamber.

The academic landscape was sharply divided in two parts – higher education institutions and universities, which were mainly educationally-oriented, and the centralised academies of sciences which only focused on doing pure research without including an educational component. Of course, there were various diffusions between the respective groups, which led to some institutional integration in the mid to late 1980s (at least in the case of Bulgaria).

The liquidation of the VAKs, followed by the decentralization of the career development system for scholars took place at the beginning of the new century. The effect was sometimes controversial, because the desire for a rapid career growth in the field of research, combined with personal assessment systems, which were not always sufficiently demanding, reflected on the quality of work and, in some cases, on the devaluation of certain research positions. Regardless of some imperfections, however, it is very important that the new system, which copies European practices in its main part, guarantees relative academic freedom and that independent evaluators, in the form of juries, are neither political bodies, nor politically engaged.

The assessment criteria that were applied were largely typified, following international trends, but they were not sustainable over time. They were often influenced by sporadic "modern trends" that were introduced quickly, without analysis and evaluation of the impact, which sometimes led to quite unpleasant consequences. Then they would disappear, but the inherited problems would remain much longer.

The introduction of a criteria-based objective system regarding the assessment of research organisations was also influenced by the manifest accreditation system of Great Britain (at that time it was a European practice, and Great Britain was a member of the EU). Accountability to society, imposed by Margaret Thatcher as a result of the outcomes of the white paper on education, was very well received in almost all member states and membership candidates. This seemed reasonable because public funds were being spent. National Accreditation Agencies, which have a similar mission, almost identical criteria that were largely a replica of the British system, were formed over a short period of time in the countries.

The EU agencies themselves are quite different, because education and science policies are horizontal policies and, therefore, full synchronization is not expected.

For example, for Germany, the applicable criteria for research assessment as part of general accreditation include individual achievements in teach-

ing, writing proposals or adequately recognised publications. Performance evaluation is not limited to merely counting the number of publications or comparing index factors. Performance evaluation should primarily be based on qualitative standards. The assessment of a researcher's achievement must be carried out in its entirety and must be based on substantive qualitative criteria. In addition to the publication of articles, books, data and software, other dimensions can be taken into account, such as involvement in teaching, academic self-administration, public relations or knowledge and technology transfer. Details of quantitative metrics such as impact factors and h-indices are not required and are not to be considered as part of the review. Accreditation focuses on curricula (assessed for quality), research is not an explicit object of this assessment, although it is presented as a criterion.

For other countries – for example Bulgaria – the number of publications in indexed journals is a leading criterion for assessing the quality of research activity.

Another factor that strongly influences the evaluation process in Eastern European countries is the Tempus program – conceived as a program for the modernization of higher education. Initially, it was identified as part of the PHARE program. This program started as targeted aid to Poland and Hungary, then expanded to other countries in Central and Eastern Europe. Subsequently, Tempus was distinguished as a separate program (it has three execution cycles).

During the implementation of the Tempus program in the CEE countries, almost all projects involved old member countries, which were often also leading a given project, and their good practices were easily transferred to the new member countries, the same applies to the evaluation process.

In support of these assertions, we also offer the case study of the establishment of an accreditation agency in Bulgaria. The project under PHARE-BG 95.06 – 05.01.001.: the first phase "Preliminary study for accreditation of higher education institutions in Bulgaria" was implemented with the consulting support of the Center for Quality Support at the Free University of London (Quality Support Centre, Open University-London-QSC) with long-term experts Prof. William Callaway (November 1996 – May 1997) and Dr. Hugh Glenville, and its second phase "National Assessment and Accreditation Agency" was implemented over the course of one year by a British Council team led by Dr. D. Billing. Pilot accreditations of higher education institutions were also carried out. Thus, PHARE was an instrument for the early transfer of the British audit culture in Bulgaria.

The superimposition of crisis factors determines a number of peculiarities in the introduction of institutional accreditation. The first and perhaps the most essential feature is the shift of the focus of evaluation from development to accountability by limiting the procedure to seek compliance with state requirements. Another feature is the emphasis on accreditation instead of the process of self-evaluation by the institution and evaluation by external experts. Although the decision to accredit an institution is the result of an assessment, by the very nature of accreditation for both assessees and assessors the focus is on the outcome (i.e., recognition of compliance with laws and government requirements and the granting of a license to continue the activity) rather than on the process itself (i.e., the quality of the assessment). This becomes even more important due to the fact that the refusal to accredit an institution according to the regulations leads to severe sanctions, including closure, which happens very rarely.

While major industry evaluations (i.e. ISO-related) are process-oriented, many research evaluations are centred around the outcomes of the system. Even when process evaluation is immanently a part of the overall evaluation it has a somewhat lower priority than the must-have outcomes.

In some cases, the development of institutional accreditation has been dictated by the widely shared perception that the unsatisfactory state of higher education is primarily due to its structural inefficiency. Therefore, it is assumed that with the improvement of the general structure of the system, more favourable conditions should be created for improving the quality of teaching and scientific research. In order to achieve this, a number of national, European and trans-European financial instruments have been introduced to help solve this problem. This process is not new at all. At the very beginning of the transition to a market economy in some of the countries of central Europe – former satellites of the USSR, grant schemes were awarded through the World Bank to solve some of the problems of the research system and, more precisely, of higher education. Subsequently, almost all EU candidate countries implemented similar projects with the financial support of the World Bank. In a very large part of them, the emphasis was placed on the modernization of the higher education process, its assessment and convergence with good international practices.

All such tools have an effect on the research environment to varying degrees. But in all cases, it (research environment) is influenced and respectively responsive to intervention and leads to behavioural changes.

However, the changes in behaviour as a result of the changes in the research environment have led to a lack of trust in it and, in turn, in

the research guild. (Dis)trust and (dis)respect are considered among the most important factors for research update (Oliver et al. 2014). Therefore, it can be assumed that one of the reasons for introducing new formal criteria is to improve the image of researchers, and for them to use it as a "label" or sign of quality, implying some kind of prestige and the possibility of comparability with other renowned researchers. In a similar context, research institutions proposing similar indicators of comparability can also be compared.

Education and research are constantly being marketised. They are being considered as market products. This in turn results in a massification of education, leading to a decrease in its quality and an absence of a research component. As a consequence, this negatively affects the reputation of researchers involved in the education process. Therefore, a certain kind of "recognition" of researchers is also necessary in order to restore their reputation. The formal evaluation process could contribute to this end, if it includes international benchmarking and popular media.

In the social comparison theory, social competition is assumed to be an element of the framework of these comparisons. In a sense, the evaluative nature of research corresponds to this statement. Based on trivial criteria, certain institutions are divided into groups. They are typified by certain characteristics. In addition, the theory postulates that social motivation is the result of 5 factors, one of which is "affirmation in society". Therefore, the categories into which a given research structure falls, as a result of the evaluation process, contribute to its "appropriateness" and ensure a "respectable" place in society.

There is a global unabated debate on which universities do better than the others, which researchers are better than others, where to publish and, at the end of the day, how to evaluate and fund the research systems on a national level.

The answer to the latter has important consequences for the research behaviour of organisations and their members. Policymakers influence research output through the research evaluation systems they adopt, due to the fact they are strongly linked to the financial support provided to any given research organisation.

So, what is the best research evaluation system then? Does it exist at all? Is Europe converging or diverging on how countries evaluate their research systems? What are the contextual factors which will determine the institutional suitability of a given research evaluation system to a given

national context? What is the subject of the evaluation – outcomes or the assessment process itself?

The anchor of this book or the underlying question is how and to what extent research evaluation practices are interrelated with the national innovation ecosystems. Would there be differences in small open economies, such as Bulgaria, the Czech Republic, Hungary, Lithuania, and Slovenia, or bigger ones, such as Poland, Austria, and the Netherlands? Why do some countries focus on qualitative and others on quantitative indicators? Why do some countries use holistic approaches and other use patchwork (copied fragments from different countries)? How can we link policy priorities, changes in the institutional framework, evaluation planning, and impact measuring in such turbulent times?

Furthermore, could we possibly find examples of practices of a research assessment, which is aligned with societal priorities (communicated through civil society organisations, NGOs) and not with political priorities which change every time there is a change in the political infrastructure? Böschen et al. (2020) advocate for the need for participatory research but also explore its challenges related to epistemic control. There are various examples of how civil society participates in the knowledge creation process in the same way as business representatives have been doing so for decades. The vast majority of research assessment literature, however, does not reflect the quality of research from a societal perspective.

We were curious to see if we could find a compromise between the two perspectives of audit culture from the beginning. The in-depth understanding of the academic landscape in Bulgaria and Eastern Europe suggests a possible third way of introducing an audit culture as an instrument of power within the academia.

The book's endeavour was partially motivated by the need to provide a somewhat coherent policy advice to the acceding countries from the Balkans. Nevertheless, we believe the findings, conclusions, and recommendations could be useful to CIS and BRICS countries as well. At the same time, all three authors have deep roots in civil society and we believe that the book could also assist in finding a way to achieve a larger civil society engagement in research assessment as a way to bypass the political control and self-iterating system of accreditation agencies and processes.

The book provides an analysis of the latest trends in research assessment systems worldwide and concrete methodologies applied by comparing eight European Union countries. Of course, in view of the fact that the authors are Bulgarian, their country might be overrepresented in examples, but it is

only because it is most in need of policy actions among the eight national systems which are the subject of the study.

The book argues that the research assessment system and the national innovation system and the overall institutional enforcement are interdependent. Countries with better rule-of-law and a higher level of innovativeness tend to have more qualitative indicators and stronger peer-review, while those with weak governance systems, low public trust and a low level of innovativeness would prioritise quantitative and objective indicators, however with an overall lower quality than their counterparts.

Last but not least, the idea for the book emerged as a result of the excellent work on the *European Network for Research Evaluation in the Social Sciences and the Humanities* (COST Action 15137) project, which allowed for research assessment know-how to be shared and in which the leading author actively participated.

We hope that this study could serve as a powerful mirror for different stakeholders such as policymakers, research organisations, individual researchers who would wish to design new research evaluation initiatives, but also for think tanks and civil society activists.

Although many people have contributed to the book by providing documents, giving interviews, reading parts of the text, and providing comments and suggestions, all errors remain ours.

Böschen, S., Legris, M., Pfersdorf, S., & Stahl, B. C. (2020). Identity politics: Participatory research and its challenges related to social and epistemic control. *Social Epistemology*, *34*(4), 382–394.

European Education and Culture Executive Agency, Eurydice, (2020). The European higher education area in 2020: Bologna process implementation report, Publications Office. https://data.europa.eu/doi/10.2797/756192

Locke, J. (1689) Essay concerning human understanding

Oliver, K., Innvar, S., Lorenc, T., Woodman, J., & Thomas, J. (2014). A systematic review of barriers to and facilitators of the use of evidence by policymakers. *BMC health services research*, *14*, 1–12.

Reichmann, S., & Wieser, B. (2022). Open science at the science–policy interface: bringing in the evidence? *Health Research Policy and Systems*, *20*(1), 1–12.

Shore, C., & Wright, S. (1999). Audit culture and anthropology: Neo-liberalism in British higher education. *Journal of the royal anthropological institute*, 557–575.