Journal of Scientometric Research, 2025; 14(2):617-631.
https://www.jscires.org

Research Article

A Comparative Study of Research Evaluation Systems in the United Kingdom and Australia

Mitra Baghjanati[1], Mehrdad CheshmehSohrabi[1,*], Hamid R. Jamali[2]

[1]Knowledge and Information Science Department, University of Isfahan, Isfahan, Darvazeh Shiraz, IRAN.
[2]School of Information and Communication Studies, Charles Sturt University, Wagga Wagga, NSW 2678, AUSTRALIA.

ABSTRACT

Research Evaluation Systems (RESs) can be divided into "pre-performance or post-performance evaluation" or "retrospective or prospective evaluation". The Retrospective Research Evaluation Systems (RRESs), also known as Performance-based Research Funding Systems (PRFSs), are complex national systems designed to evaluate universities and research centers and allocate public funds based on their outputs and outcomes. This study compares the RRESs of the UK and Australia to gain a better understanding of the structure and components of these evaluation systems. A comparative study method was applied to look for similarities and differences between the RESs of selected countries. The two countries were chosen based on the criteria of transparency, access to credible documents, formality, comprehensiveness, flexibility, and management. Bereday's four-step model consisting of description, interpretation, juxtaposition, and comparison of RESs was utilized for the data analysis. The results showed that two evaluation systems emphasize the components of human resources, finance and infrastructure in evaluation. Quantitative and qualitative approaches prevail in all two systems. In terms of the evaluation unit, the systems have almost the same structure, and the evaluation is done by specialized panels. The quality of research outputs is evaluated in two systems, and in the UK, the two elements of impact and research environment are also evaluated. In general, it can be said that the formation of national systems for research evaluation affects not only the quality of research, but also the purposefulness of research, scientific and technical progress of the country, and people's lives. The unique aspect of this study is the comparison of the input, process, output, and impact components of the RRESs of the UK and Australia. The results can be beneficiary for managing and policymaking for research assessment units on micro/macro levels.

Keywords: Research Evaluation, Research Evaluation System (RES), Research Excellence Framework (REF), Excellence in Research for Australia (ERA), Performance-based Research Funding Systems (PRFSs), Bibliometrics.

Correspondence:
Mehrdad CheshmehSohrabi
Knowledge and Information Science Department, University of Isfahan, Isfahan, Darvazeh Shiraz, IRAN.
Email: mo.sohrabi@edu.ui.ac.ir
ORCID: 0000-0003-1856-4210

Received: 24-02-2025;
Revised: 09-04-2025;
Accepted: 18-07-2025.

INTRODUCTION

Numerous Research Evaluation Systems (RESs) have already been designed for various purposes. These systems generally fall into two main categories. The first group involves systems which, as put by Whitley (p6),[1] "are organised sets of procedures for assessing the merits of research undertaken in publicly-funded organisations that are implemented on a regular basis, usually by state or state-delegated agencies." Research evaluation approaches might involve "pre-performance or post-performance evaluation"[2] or "retrospective or prospective evaluation".[1] The recently-developed retrospective research evaluation systems are known as "Performance-based Research Funding Systems (PRFSs)" and considered "complex, dynamic systems, balancing peer review and metrics, accommodating differences between fields, and involving lengthy consultation with the academic community and transparency in data and results".[3] Hicks and Geuna and Piolatto[4] considered REFs as complex national systems that are created to evaluate public research institutions (e.g. Universities) and to help allocate public money to them based on their outputs and outcomes instead of structures and processes."

Based on the literature, efforts have been made to set up PRFSs in a number of developed countries over the past decade. PRFS was first established in UK (RAE/REF), and different forms of this system were used, by 2010, in 14 countries: UK, Australia, Belgium, Denmark, Finland, Hong Kong, Italy, the Netherlands, New Zealand, Norway, Poland, Slovak Republic, Spain, and Sweden.[4] The PRFS include three evaluation approaches including qualitative (peer review), quantitative (bibliometric analysis), and hybrid (peer review-bibliometric analysis).[4] One important question in relation to REFs is that if they have resulted

OPEN ACCESS

Manuscript TECHNOMEDIA

DOI: 10.5530/jscires.20251111

Publishing Partner : Manuscript Technomedia. [www.mstechnomedia.com]

in any changes in the research performance of universities and research centers, and some studies have tried to answer this question.[5-17] We will discuss some of the consequences of REFs in the related studies section.

In prospective RES, compared to retrospective RES, pre-performance evaluation is considered. As suggested by Cozzens,[2] referring to the American system, the U.S. funding system puts a lot of weight on pre-performance evaluation (i.e., peer review system) and puts very little weight on the evaluation of programs. As a result, the American funding bodies focus mostly on giving grants and not so much on the outcome of the grants. Whitley[1] studied the investment levels in both prospective and retrospective RESs worldwide and found that, "in most countries, state investment in this prospective evaluation of applications for funds and guiding resource allocation procedures towards the achievement of public policy goals has been considerably greater than that provided for the retrospective evaluation of the results."

In the second category, research funding is not designated to public sectors, and rather research units, document, scholars and authors, disciplines, etc. are assessed, validated, and ranked locally (RES in a province or university), nationally (e.g., systems ranking universities, journals, or scientists in a country) and internationally (e.g., SCImago Journal and Country Rank, Scival, JCR, Google Scholar, Scopus, Web of Science, Shanghai Ranking, and QS World University Rankings). These systems aim at research excellence through controlling and ensuring quality. Note that prospective systems, while trying to finance public research institutes and universities, have the ultimate aim of research prestige. Hicks [3] has pointed that the intention of prospective systems to distribute research funds to universities seems to be an illusion because the real incentive for universities appear to be competing for prestige.

Many RESs in the past performed evaluations on the basis of quantitative measures. However, in recent decades, evaluations have made more use of qualitative methods and peer review. They aim to measure the quality of research, and now more attention is paid to measuring and demonstrating research excellence through enhanced national evaluation systems.[18] Moreover, assessment systems in education, like the "publish or do not graduate" model, have faced criticism for potentially increasing stress levels among students.[19] Due to this problem, certain countries consider only Web of Science indexed articles, which poses an issue. This issue shows itself more in the field of social sciences and humanities, especially in some book-oriented disciplines. Numerous studies have investigated this issue.[20-22] National systems for evaluating research, with varying degrees of formalization, transparency and standardization of procedures are often very complex and dynamic.[3] Such national systems strengthen the link between government policy and academic research and increase public accountability for government budgets.[23] Thus, national research evaluation systems can be introduced as policy-making tools for allocating research budgets and improving the quality of research, in accordance with scientific developments.[24] The purpose of these systems is to create scientific outputs with a high level of accountability and functionalism. This potential can continuously change the organizational structure and culture and individual scientific activities.[25] Comparing evaluation systems will help us develop an understanding of research evaluation and its elements. We will understand different approaches to research evaluations and what different approaches to evaluation can achieve. There has been little comparison of such evaluation systems in the past and this study aims to contribute to this area.

The RESs are different in different countries. This study aims to analyze and compare two evaluation systems which are Research Excellence Framework (REF) in the UK and Excellence in Research for Australia (ERA). The ERA and REF fall into the retrospective RES category. The UK system was chosen because it is one of the first RES in the world and is well-developed after several iterations. It is an example of a PRFSs that helps make research be as accountable as possible. Over the past few decades, the worldwide reputation of this system has inspired countries such as Spain, Australia and New Zealand to adopt this system and have their own RESs.[25-27]

Australia's ERA has been in operation since 2010. The data collected during the various iterations of the system provide valuable information on the research performance of Australian higher education institutions to government, academia, industry and students. The Australian Research Council also collaborates with higher education institutions in strategic planning and decision-making and research activities to assist them in their development.

Overall, the UK and Australia have specific RESs and have developed their RESs in recent decades. This study aims to compare them, identify their components and structure and discuss their advantages and disadvantages. This analysis will provide new knowledge about these systems that will be needed for understanding how these systems conceptualize and evaluate different aspects of research such as output and impact. It will also help countries that intend to develop a new or adopt an existing system, to make informed decisions.

LITERATURE REVIEW

Studies related to national research evaluation systems can be divided into four categories as follows:

Comparative research in the field of RESs

Comparison of evaluation approaches, methods, or indicators within a research evaluation system or in two or more RESs is another research front. The good examples of research in this category are the studies by Coryn *et al.*,[28] Jonkers and Zacharewicz,[29] Zacharewicz *et al.*,[30] and Ochsner and Peruginelli.[31] Coryn *et*

al.,[28] classified countries according to three types of funding 1) large-scale performance-based exercises, 2) bulk funding, and indicator-driven performance-based models. Jonkers and Zacharewicz[29] classified Research Performance Based Funding Systems (RPBF) into two main categories including funding allocation formula based on quantitative indicators (education and PhD awards-based formula, RPBF based on journal-based impact assessments, RPBF based on citation-based impact assessments) and RPBF based on peer review assessments. They compared also the countries based on this typology. Zacharewicz *et al.*,[30] compared the performance-based research funding in EU member states and found two funding allocation formulas based on quantitative metrics and peer review. Ochsner and Peruginelli[27] introduced different classifications of national evaluation exercises. They present a typology of national research evaluation systems into seven groups (including accreditation, national evaluation: formative, national evaluation: performance-based, national excellence initiative, national career promotion, project funding by government, and evaluation for academies of science). Finally, they evaluated the countries based on this typology.

Abramo, D'Angelo, and Di Costa[32] also compared the peer review and bibliometric rankings in the first Italian evaluation exercise. The results indicated "great differences between peer review and bibliometric rankings for excellence and productivity".[32-34] Additionally, many researchers have investigated the methods of bibliometric and peer review, among which we can mention Horrobin,[33] Moxam and Anderson,[34] Moed,[35] Van Raan,[36] Pendlebury,[37] and Abramo and D'Angelo.[38] Furthermore, Sandströma and Besselaar[39] examined also the performance of research evaluation systems. They found large differences in efficiency between national science systems based on funding systems, level of competition, level of university autonomy, and level of academic freedom.

Evaluation of outputs, outcomes, and research impact in RESs

Evaluating the type of research outputs such as books, articles, patents, software, etc. is one of the most challenging issues in research evaluation systems. Among the research outputs, the most challenging is how to evaluate the book in research evaluation systems. This issue shows itself more in the field of social sciences and humanities, especially in some book-oriented disciplines. Numerous studies have investigated this issue. Among these studies, we should mention the research by Zuccala and Robinson-García.[40] They focus on examining the role of experts in improving book evaluation conditions in research evaluation systems. Their research shows where there is the greatest potential for the development of quantitative and qualitative indicators in book evaluation systems. Giménez-Toledo *et al.*,[41] also compared different approaches for assessing books in Spain, Denmark, Flanders, Finland and Norway.

Research outcomes are "the achievements of the research activity, whether conceptual (a new theory), practical (a new analytical technique) or physical (a new device or product - although some authors regard this as an output); research outcomes are potentially available for use".[42] A main challenge in research evaluation is the selection of measuring mechanisms and approaches. For instance, Garrett-Jones[42] examined the university research outcomes evaluation methods in the US, Canada, the Netherlands, etc., concluding that Australia was following the common practice somewhat similar to the other studied countries and in some aspects, it was more advanced. Grigson and Stokes[43] studied the application of peer review in the evaluation of the outcomes of "research grants" by the "Australian Research Council." Lillis,[44] referring to the case of New Zealand, argued that the assessment outcomes are material benefits that improve the quality of life, including the growth of industrial competitiveness, environmental quality, social regulation effectiveness, and skill and learning improvement. Garrett-Jones and Aylward[45] focused on "some tension between criteria of excellence and socioeconomic benefit in valuing university research outcomes."

In view of the focus of RESs in recent years on the impact component, some studies have tried to identify and review existing models and frameworks for evaluating research impact. They review different definitions of research impact and its evaluation methods and investigate the challenges of impact assessment[46] including in specific disciplines, such as the scientific, social and political implications of social sciences and humanities research.[47] They also study the role of systems that may be used in the future to link research and its implications and the requirements that exist for these systems[25] According to the findings of these studies, evolving systems that focus only on recording impact information must consider the possibility of any interactions between researchers, institutions, and external stakeholders. A review of research in this area shows that among the methods used to measure the research impact, there is still no well-established standard method that is agreed upon by all[46] In this respect, the value of bibliographic indicators for science, technology, engineering and mathematics as an objective and low-cost method is more prominent than other methods, but the possibility of using such indicators in evaluating the impact of social sciences and humanities research is highly questionable.[47] Therefore, it can be said that the combined approach of case studies is an excellent way to discover all the available information, data and evidence, which provides a comprehensive summary of the contextual impact. Quantitative impact analyses analyze program effects over a period of time from a fixed point in time, but qualitative studies are often performed during the actual program to provide information to improve executive performance. On the other hand, if a composite approach is accepted as an impact assessment method, the limitations of its complex implementation must also be understood.[48]

Research assessment methods, criteria and indicators in RESs

A basic issue in RESs is the choice and use of the best method, criteria, and indicators to assess the effectiveness of scholars, research unit, discipline, etc., which has been examined from different angles. The relevant studies can generally be divided into two main categories:

Studies that discuss bibliometrics in research evaluation and in RESs

Over the past decade, national RESs that have traditionally been conducted using peer review methods have begun to use bibliometric indicators. In this case, it was assumed that the publications that received the most citations were of higher quality and that metrics were a good judgmental tool for identifying top research in various fields.[49] However, due to the fact that citations in different fields of research behave differently due to different rates of literature decay (for example, research in medical fields become obsoletes fast and its citation drops), before using such an assumption in comparative evaluations between organizations or even between individual scientists, citation data must be standardized.[38] Thus, these studies describe the various aspects of the term "frequency" among citations, define average citations,[50] and introduce dual function as an indicator for evaluating research based on the types of low-cited or high-cited publications.[51]

Studies that compared peer review and bibliometric approaches in RESs

Some of these studies show agreement and correlation between bibliometric and peer review and some argue that bibliometric is not suitable for certain fields. A big factor in advocating for the use of bibliometric is of course the high cost of peer review. Wallmark and Sedig[52] showed that bibliometric was almost 206 times less expensive and less time consuming in the case of Swedish system. However, there are others such as Jacsò[53] who argued that it would be premature to think that citation indicators can replace peer review as a cost-effective alternative.

Examples of studies that looked for agreement between two evaluation techniques are Thomas and Watkins[54] who found high agreement between peer-review and citation analysis results in the case of the old RAE system in the UK. Nederhof and Van Raan[55] also found similar results in the Netherlands as the two evaluation approaches were mutually supportive and complementary. Aksnes and Taxt[56] found weak but positive correlation between peer review and bibliometric results in Norway. Checchi *et al.*,[57] compared ranking of publications based on the UK REF system with bibliometric indicatros used in Italy and found high correlation between the two approaches.

However, such agreements between peer review and bibliometric in many other studies are either non-existent or are discipline specific. Abramo and D'Angelo[38] showed that bibliometric was "by far preferable to peer-review in the natural and formal sciences." Abramo, D'Angelo and Di Costa[32] in a study of the Italian system compared bibliometric and peer review and concluded that these two are very different in terms of research excellence and productivity. Fedderke[58] studied South Africa's evaluation system and maintained that peer review and bibliometric measured different things. Mryglod *et al.*,[59] found different correlation between bibliometric and peer review for hard and soft sciences. Abramo, Cicero and D'Angelo[50] studied the evaluation of Italian universities and concluded that using peer review for hard sciences would be "a complete waste of money". Their results were aligned with that of Wallmark and Sedig's.[52] Baccini and De Nicolao[60] also found week agreement between peer review and bibliometric at individual article level. Rodríguez-Navarro and Brito[61] looked REF results in the UK and suggested for peer review could be replaced with bibliometrics only for top citation percentiles of papers. King[62] looked at avian virology research and showed that bibliometric indicators should not be used by themselves. Nederhof and Van Raan,[63] Rinia *et al.*,[64,65] looked at some sub-fields of physics and found some agreements between per review and bibliometric Oppenheim and Summers[66] showed that citation indexes are not suitable for arts and humanities fields such as music. But Norris and Oppenheim[67] showed bibliometric and peer review results had correlation for the field of archaeology (which is another humanities sub-field). Similar correlation was found in economics in Italy by Bertocchi *et al.*,[68] and in Australia by Bruns and Stern.[69]

Consequences of RESs

This group of studies specifically examine and analyze the gradual institutionalization and the favorable and unfavorable consequences of RESs in specific countries such as the UK[27,70] or Finland,[71] over time. Additionally, we can refer to the research by Chatterjee *et al.*,[72] which reflects the views and reactions of academics to the UK and New Zealand RESs and the impact that these systems have had on their careers in both countries. Whitley[1] classified the main consequences of RES: "a) increase of research organization stratification, b) intensification of reputational competition and coordination of research goals, c) strengthening of central disciplinary standards and priorities, d) reduction of intellectual diversity and pluralism, and e) increase of constraints on establishing new fields and approaches."

In some studies, the development of different concepts and dimensions of research quality and the development of research quality evaluation systems based on bibliometric analysis or peer review or based on a composite evaluation method (for example, Italy) were examined and their special advantages and disadvantages were criticized.[73] On the other hand, some studies about designing RESs have focused specifically on research outputs, such as articles, books, dissertations, patent reports, software or special products, and have examined different

methodologies for measuring and evaluating these resources-for example, measurement-based systems for evaluating the quality of research outputs in Australia.[74] Italy[73] and Finland.[71] It should be noted that in addition to quantitative methods, qualitative evaluation methods, which are mainly performed by presenting case studies and peer review methods, are also accepted by the academic community and administrators in some countries (UK). Studies in this field in other countries are mainly looking for methods to reduce bias and prejudice related to the measurement of existing indicators,[75] because existing methods for evaluating scientific products are often affected by limitations and prejudice. For example, the results of periodic evaluations might be in favor of larger universities, and allocating a joint government budget based on research evaluations might cause unfair consequences. Some studies in the field of designing RESs have also focused on measuring and evaluating inputs or the basic requirements for starting research such as manpower, financial resources, facilities and equipment; However, due to the demand for more accountability of academic research, studies in this area are more focused on research credits and budgets.[71] These studies show that performance-based budget allocation evaluation systems are more popular than other financing approaches. Emphasizing the link between performance and resources and encouraging research excellence, research input evaluations have shown that the results of performance-based evaluations include creating a strong incentive to improve individual and organizational performance and thus increasing productivity through competition,[23] increasing public accountability for government funding invested in research, and establishing a mechanism for linking academic research with government policy.[4] Additionally, funds distribution based on RESs is another issue. Among the researches, we can mention the study by Greco and Scarcello[76] that looked into the laws of funds distribution in the Italian Research Assessment Program (VQR) during 2004 to 2010. Moreover, some researchers have addressed the implications and imperfections of RESs from another perspective. Among these researches, the research by Xiaochun and Dan[77] should be mentioned. They found a major connection between corruption in many of China's colleges and universities and the scientific research evaluation system.

Butler[78] studied the impact of the application of "raw publication counts" as a performance criterion under an output-based funding policy in Australia in the 1990s. She found a dramatic increase in the volume of publications that was achieved at the expense of quality because the increase was mostly in journals ranked low in terms of impact. However, later Van den Besselaar *et al.*,[79] used a longer time series and argued that her main conclusion was not correct. Schneider *et al.*,[80] compared the Norwegian publication-based funding model that used differentiated publication counts to the Australian model and found that the adverse impact that Butler found in Australia, did not happen in the case of Norway. Similar to output based evaluation, citation based evaluation can result in some behavioral changes in researches. For instance, Abramo *et al.*,[81] showed that the use of a citation-based incentive scheme in Italy resulted in 9.5% increase in self-citation rate.

Some experts such as Edwards and Roy[82] warned that too much competition for funding and performance measurement using metrics create perverse incentives for researches which might result in unethical behavior and this in turn might ultimately render the entire science enterprise untrustworthy in the eye of the public. Some studies of unethical research behavior such as plagiarism (e.g., Honig and Bedi)[83] have shown that such behaviors are not uncommon. Therefore, to avoid undesirable consequences due to implementation of an evaluation system, some researchers such as Hazelkorn[84] have suggested that evaluation systems should include both quantitative and qualitative data, and include the assessment of both impact and benefit as well as self-evaluation.

Finally, it can be said that since monitoring the scientific performance of countries is a high priority for research managers and government organizations involved in the research process of each country, and given that the methodologies used in different systems vary in terms of structure, content and use of quantitative and qualitative indicators and their consequences, it is of great importance to determine the extent to which analyses based on different indicators can help better identify and understand the effects and efficiency of different research evaluation systems. Therefore, in the present study, we tried to identify the similarities and differences between the RESs of England and Australia, in order to determine the structural and content elements and components necessary for research evaluation.

METHODOLOGY

The study used a comparative approach to analyze and compare the structural and content elements of the RESs of the UK and Australia.

To identify the systems for comparison a few criteria were considered including: 1) transparency of documentation of the evaluation system, 2) existence of a website that provides accessible, reliable and accurate information about the system, 3) formality (duration of formation and implementation of selected systems), 4) comprehensiveness (e.g., covering a wide range of potential research outputs and fields), 5) flexibility (e.g., the existence of guidelines compatible with different objectives and approaches such as evaluation of researchers, groups or institutions), and 6) management of the evaluation system (leadership of the evaluation framework by a central authority). On this basis, the RESs from the following countries were selected: The UK and Australia.

Bereday's four-step model[85] was used for the comparison process. Based on this model, the required information about RESs in

selected countries was collected and reviewed and were subjected to four steps of description, interpretation, juxtaposition and comparison as explained below:

1. **Description of research evaluation systems:** In this step, by using the websites of RESs in selected countries, other reliable information sources and in a few cases by requesting information from responsible authorities of these systems through e-mail, sufficient information was collected about each system to be able to understand and describe each system.
2. **Interpretation of the content and structural elements of research evaluation systems:** This includes verification of the information described in the first step. At this stage, by focusing on the content and structural elements of each of the studied systems, the specific characteristics of each system were summarized.
3. **Juxtaposition of content and structural elements of research evaluation systems:** In this stage, information that has been interpreted in a systematic way is classified and juxtaposed to reveal the similarities and differences of such information.
4. **Comparison of content and structural elements of research evaluation systems:** In this stage, similarities and differences between structural and content elements were compared. It is noteworthy that this research stage is based on a descriptive approach focusing on the *how*, rather than *why*, of the systems.

In this article, we reviewed academic literature and materials published on the official website of these research evaluation systems up to early 2023.

RESULTS AND DISCUSSION

Here, reliable resources and websites dedicated to RESs in selected countries and valid information about the structure and content of these systems were systematically collected.

The UK research evaluation system

The UK has a long history of evaluating the research performance of its universities and took the first step towards establishing a national evaluation system in 1985. It was almost a decade later that other countries, following the UK, began to nationally evaluate the quality of their research.[27] The first explicit evaluation of research quality took place in 1986. Evaluations were carried out by thematic subcommittees of the university grants committees. Then in 1989, for the first time, a research evaluation activity was conducted nationally in the UK under the title of Research Evaluation Exercise (RAE). However, concerns about the huge costs of RAE soon increased, especially by the UK Treasury.[86] As a result, after the 2001 exercise, steps were taken to replace it with a system that relied more on metrics such as citations, research income, and the number of graduate students rather than specialists. Nevertheless, skepticism from the outset between the system's beneficiaries, including the Higher Education Funding Council for the UK and individual institutions, led to the rejection of the metrics system. Another concern of the existing method was that large universities could still be ranked five (top) despite less work, and all academics at this level could receive funding.[87] Therefore in 2008, RAE introduced quality profiles for each assessment unit and performed it for the last time. In 2014, the Research Excellence Framework (REF) replaced RAE.[88]

The REF is co-sponsored by the Research of England (RE), the Scottish Finance Council (SFC), the Higher Education Funding Council of Wales (HEFCW) and the Northern Ireland Ministry of Economy's Department for Education (DfE). The REF is managed by the REF team based in RE on behalf of the four UK funding bodies and overseen by the REF steering group consisting of representatives of the four funding agencies.[89]

REF is based on the study of specialists and is done using their specialized knowledge and experience. This evaluation framework was expanded to include, in addition to the quality of the research results, the evaluation of the impact of the research and the environment in which the research was conducted. It is important to note that REF is a single framework for evaluation in all disciplines, with a common set of data in all submissions, definitions, and standard procedures.[90] Evaluation is performed by expert panels with general criteria, supported by specific criteria for each main panel and sub-panel group, if possible. Expert panels consistently apply evaluation standards and operate under the supervision of the four main panels. In this unified framework, differences in the nature of research in different disciplines justify differences in the exact approach to evaluation. There is flexibility for panels to create specific aspects of evaluation criteria to ensure that evaluation is sensitive to these interdisciplinary differences. Panels consult with subject groups and research institutes.[91] REF 2021 was developed through an evolutionary and consultative process, based on the successes of previous REFs and RAEs and fundamental changes in response to research institute feedback. In implementing the recommendations, financing institutions have sought to strike a balance between continuity and development.[92]

The Australian Research Evaluation System

Australia's higher education budget and evaluation policies were revised in the 1990s and in the first decade of the 21st century. Following the development of evaluation systems in the UK, a committee was set up to provide a framework for evaluating the research performance of Australian universities based on critiques of the evaluation of research in the UK; however, the Australian Government's attention to research evaluation was

beyond a budget distribution.[93] Assessing the quality of research has become a vital issue in the Australian scientific community, and proposing a Research Quality Framework (RQF) was a fresh start in discussions about how to evaluate the quality of research in the best possible way. Following the move towards measuring research excellence, the Australian Research Council (ARC) substituted the Excellence Research of Australia (ERA) for the RQF. Accordingly, the ERA stated its primary goal of identifying and promoting excellence in a wide range of research activities, including the discovery and application of research in Australian higher education institutions.[94] The first ERA pilot evaluation was performed in 2009 and then continued in 2010, 2012, 2015, and 2018.[74] The ERA evaluation process was developed after extensive consultations between the ARC and the Australian Higher Education Department. These developments included the formation of working groups in a wide range of disciplines and the creation of comprehensive benchmarks for domestic and international experts.[95] Minor corrections were also made to various aspects of the submission and evaluation process, but the set of indicators introduced in the evaluation framework discipline matrix remained largely unchanged.[96] The indicators used in these assessments were developed by the Australian Research Community. This approach ensures that the indicators used also minimize the workload of government and academia, while ensuring that ERA results are robust and acceptable. In 2018, the Australian Research Council (ARC) conducted the fourth Comprehensive Assessment of Australian Research Excellence (ERA). In this way, information on the quality of research activities recognized as eligible in higher education institutions was evaluated by eight Research Evaluation Committees (RECs), at a level of clustering of 147 prominent domestic and internationally recognized researchers in their subject areas.[95] ERA thematic clusters comprise a structure that has been developed primarily to help balance workload across different fields of research.[97]

After extensive consultation in the higher education sector and a thorough review of the strengths and the weaknesses of other systems, particularly the UK Research Evaluation System, it was decided that ERA evaluations in 2023 would focus on disciplines rather than segments or groups. In other words, instead of creating a single evaluation method, the ERA created a set of indicators that differ at the discipline level (known as the discipline matrix).[98] Citation indexes are mainly used to evaluate disciplines in the field of science, whereas peer review is mostly used to evaluate the outputs submitted in the fields of social sciences, humanities and arts. Universities report all of their staff's research outcomes in positions classified as 'researcher only' or 'researcher-lecturer.' Under citation analysis, all university publications in the relevant field are considered in the evaluation process. In peer review, however, the evaluation process comprises only 30% of the university's publications, which are selected by the institutions themselves.[96] Universities are given an overall rating between 1 (bottom) and 5 (top) for each major and sub-major.[99] Australian universities rely heavily on international students to generate revenue, and these students pay close attention to accurate published information on ERA results, in addition to international rankings; therefore, it can be said that while the ERA has little direct impact on the allocation of government research funding, it does have significant financial implications for universities.[74]

In addition to what was mentioned previously, evaluations in Research Excellence Framework (REF) focus on three distinct elements: the quality of research outputs, the research impact, and the environment of the research unit. The first element, the quality of research outputs, refers to the evaluation and selection of the best quality outputs to identify the level of employee activity during the assessment period. All types of research outputs, including basic research that brings new insights or outputs that have not been made public, such as confidential reports to government or industry, software design, plans, demonstrations, and even artifacts, are eligible for presentation. All activities presented should include evidence of the research process and brief statements about the aggregation of their potential audience requirements, both inside and outside the academic community. The quality of research outputs is evaluated based on the criteria of "originality, significance and rigor." For the purposes of REF, citation data related to the outputs of the disciplines of medical sciences, health, biology, physical sciences, psychology, engineering and computer science are prepared for consensus with the opinions of panel experts in the assessment units. Given the limitations of such data in disciplines such as the arts, humanities, or social sciences, citation information in not used for these disciplines.

After evaluating the quality of the outputs, the REF emphasizes the evaluation of the impact element, which shows the tangible effects of research activities during the evaluation period. There is a broad definition of impact (covering economic, social, public policy, cultural, and quality of life considerations) which recognizes intellectual and practical impact beyond academia. In impact statements, research units make a plausible claim about the effectiveness of their research, which should include a wide range of impact indicators as supporting evidence. Some indicators such as research income index, rate and scope of cooperation with a wide range of research users or other social, economic, political, cultural, health and quality of life indicators, apply to all research units in general. There are other indicators that may be specific to certain assessment units.

The third element, the environment, assesses vitality and sustainability of the research unit and the extent to which a research unit covers research infrastructure and supporting activities that lead to a continuous flow of production, dissemination, and effective application of research. Assessing the research environment can provide strong evidence for identifying

research units with a history of producing research outputs with broad impacts and can also help to define research activities that foster the sustainability and viability of disciplines at a national level - instead of particular advances at the level of higher education institutions. Institutions should provide strong and reliable evidence of their research environment using a general model including research groups as research units, research councils, how to structure and manage research and research revenues, the critical volume of infrastructure and facilities and overall financial and human resources of the unit, strategic goals and objectives, how to advance and manage the unit, how the research unit interacts with the community, publication of research findings, maintaining discourse on a wide range of related issues, and participating in research databases.[89]

The general principles of REF are equity, equality, transparency. Panels are instructed to set criteria and adopt evaluation processes that enable them to realize fundamental superiority in a wide range of research (e.g., applied, action-based, fundamental, strategic, etc.,) and determine excellence in various forms of research, including interdisciplinary and collaborative research, while not placing more emphasis on one format than another. In other words, different types of research must be evaluated with fairness and equity (Technopolis 2010).

Australia's ERA is based on peer review and some evaluation indicators. The set of ERA indicators for evaluating research behaviors is specific to each discipline, i.e., for some disciplines citation analysis and evaluation by indicators are used, while for some other disciplines, peer review is used to evaluate research outputs. It should be noted that peer reviews and citation analyses by indicators are not used in combination at the four-digit level. This principle, however, applies to the two-digit level.[86]

There are three broad categories of indicators in the ERA

1. Indicators of research quality: Evaluation in this category of indicators is based on publication specifications, citation analyses, peer review and review of domestic and international research revenues.
2. Indicators of research activity: Evaluation in this category of indicators is based on the results and output of research and other research items within the framework of the characteristics of qualified researchers.
3. Indicators of research application: Evaluation in this category of indicators is based on income from commercialization of research, patents, plant breeders' rights, registered designs and guidelines approved by the National Health and Medical Research Council (NHMRC).[96,98]

Regarding the ERA indicators, it is noteworthy that in 2018, some other measures, such as the evaluation of publication profiles and assessment of impact and engagement of research activities that can provide valuable information about the application of research in society, were considered. These measures were generated by the interaction of quantitative and qualitative elements. The data used for the ERA indicators are largely quantitative, but the impact component added to the evaluation process involves a qualitative impact assessment and the explanatory statement describes research impact and research approaches for the efficiency of the assessment units.[96,100]

The similarities and differences between the structural and content elements of the research evaluation approaches of the UK and Australia are compared in Table 1. While these systems use different methodologies for their evaluations, bibliometric analyses are used in all of them, either as the main evaluation method or as a component to inform the expert review. For example, REF in the UK uses bibliometrics only as an element to inform the main assessment process (i.e., expert review). In some sub-panels, the number of citations to an output is considered additional information about the scientific significance of the output. Panels that do this continue to rely on expert review as the primary method for evaluating output to judge a wide range of evaluation criteria (authenticity, importance, and accuracy). They will also recognize the importance of non-traditional outputs[1] and evaluate all outputs on an equitable basis, regardless of the availability or non-availability of citation data. In using citation data, the panels consider a variety of citation patterns for different areas of research, the possibility of "negative citations," and the limitations of this data for outputs in languages other than English. There is a strong belief in the UK that citation analyses (sets of metrics) cannot be applied uniformly and are therefore inappropriate.[93] The UK does not appear to be ready or willing to take the Australian path to develop a set of specific indicators for each discipline, as in the ERA, quantitative indicators play a central role in research evaluation. Using this information, however, goes beyond simply counting the number of citations. Therefore, using the Relative Citation Impact (RCI), distribution of papers based on the global percentile threshold and the average of Australian universities and distribution of papers versus RCI, three profiles of bibliometric indicators are presented to the ERA evaluation panels. These three indicators are designed to be considered as a complementary set and are not used separately. It should be noted that if citation data is considered appropriate for evaluation, expert review is no longer used for those unique outputs. A number of evaluation activities are carried out using a combination of bibliometric measures and case studies on the relative strength of research. For example, the National Science

1 . Non-traditional research outputs consist of original creative works, live performance of creative works, recording and presentation of creative works, organizing and producing significant exhibitions and public events, and research reports for foreign organizations.

Foundation has studied the commercialization of innovations as well as a range of bibliometric data such as counting patents, articles and citations (Table 1).

Another important point in reviewing systems is the evaluation unit. While most evaluation systems in selected countries, and in particular the British REF and Australian ERA, use subject-based expert panels, there are significant differences between these countries in terms of what outputs are evaluated. In the UK, only selected top research outputs are submitted to assessment units, while in Australia, unlike UK REF practice, Australian universities are evaluated with all their publications (although only a sub-set is subjected to peer review in peer-reviewed disciplines). However, in a sense, ERA and REF are similar, meaning that selected publications are not limited to university research, but may include publications that were written before an employee became a member of the institute. Thus, it can be said that institutional activities in the UK and Australia are based on data submitted by the institutions themselves, with one major difference: unlike Australia, where information on extensive catalogs of publications is considered, REF evaluation in the UK uses selective collections.

The next point to consider is research funds based on the results of evaluations conducted in selected countries. While 65% of REF-linked funding is tied to the results of UK research evaluations, research budgeting is not based on the performance of universities evaluated by the ERA in Australia, or rather the impact of these evaluations on research budgeting is very small.

Note that UK and Australia compared in this study are aware of the need to measure the impacts of research results as a significant contribution to the economy, society, culture, national security, public policy and services, health, environment, and quality of life. This awareness is completely evident from the systems' main objectives and their description of the interactions between researchers and research organizations and society and industry, for the exchange of knowledge, resources and mutual cooperation. However, only in the UK, the social impact of research is assumed as a major component accounting for 20% of the results of research evaluation. In Australia, impact is being gradually introduced into the ERA system.

In general, the Research Evaluation System (RES) of UK and Australia has pros and cons, some of which are mentioned below.

United Kingdom

The UK RES merits are as follows. A standard scale for ranking and publishing is usually utilized in the UK evaluation based on peer review to conveniently define the status of different groups and universities. Funds designation, particularly research institution incomes, and hence the university and other organizations management are directly and significantly affected.[87] In the UK system the performance criteria are decided and established by a group of science elites.[1] For any active staff, a select output is emphasized by focusing on quality. Assessment panels attach equal importance to all basic or applied research, and quality is highlighted.[23] To assess interdisciplinary works, academies are invited to present interdisciplinary research

Table 1: The status of the structural and content elements of RESs in the UK and Australia.

Elements		Components	Country	
			United Kingdom	Australia
Structural	Input	Strategies	✓	-
		Logistics	✓	-
		Funds	✓	✓
		Human resources	✓	✓
		Infrastructures	✓	-
	Process	Evaluation type	Quantitative and qualitative with more emphasis on qualitative indicators	Qualitative and quantitative with more emphasis on quantitative indicators
		Evaluation criteria	Originality, importance, and accuracy	-
		Evaluation unit	Four main panels (A, B, C and D) and 34 sub-panels	Eight research evaluation committees
		Ranking scale	5-star scale	5-star scale
		Evaluation element	- Research outputs quality- Impact- Research environment	- Research outputs quality
		Evaluation method	- Expert review- Citation analysis (count of citations per publication)	- Citation analysis- Review of ERA experts- Review of international experts- International research income

Elements		Components	Country	
			United Kingdom	Australia
Content	Output	Publications (journal articles, monographs and book chapters)	✓	✓
		Patents	✓	✓
		Designs	✓	✓
		Compositions	✓	✓
		Exhibitions	✓	✓
		International visibility and networking	-	-
		Prizes and awards	-	-
		Research Income	-	✓
	Impact	Impact on trade and economy	✓	-
		Impact on production	✓	-
		Impact on society	✓	-
		Impact on culture	✓	-
		Impact on public policies or services	✓	-
		Impact on health and hygiene	✓	-
		Impact on the environment	✓	-
		Impact on quality of life and social welfare	✓	-
		Impact on understanding, learning and participation	✓	-
		Impact on national development	-	-
		Impact on the progress of basic research	-	-
		Impact on human resource development	-	-
		Impact on creating a knowledge repository	-	-

to the most fitting panel and offer additional panels for parallel review of papers submitted.[23] Funding is based on peer-review via "scientific excellence and social relevance" criteria,[23] (which measure research impact and presence in society, economy, industry, etc.[24] The quality assessment is of great importance.[101] The importance of UK RES - as a transparent and well-known RES[101] - for scholars and staffs is due to both financial issues and public standing.[1] However, REF shortcomings include its expensiveness complexity, heaviness.[4,24] Assessment in REF are done the constant participation of strongest sectors and "achieve top positions in the rankings" leading to their further ability to adsorb external funds.[7,27]

Australia

The Australian RES system is public, consequential, standardized, and transparent.[102] The ERA funding is "unique in its exclusive reliance"[102] and is used in all universities in Australia. They demonstrate "neither the freedom of entry nor the freedom to specify their product",[101] and it is reported that, "a significant

feature of the ERA is the high level of dependence of scientists on a single agency, the Australian Research Council".[1,102]

Some ERA demerits are as follows. It assesses academics performance using mostly quantitative indicators.[101] The growing outputs assessment by quantitative indicators and scientists' dependency on the Australian Research Council resulted in the short-run choice of research preferences to simplify research funding.[102] Based on the total of publications, substantial funds were distributed between and within academies, irrespective of the publications quality or impact. In this regard, a significant increase has been observed in research productivity, while impact diminished.[78] The Australian Research assessment process is extensive and complex although quantitative indicators are used. In ERA, the complete research outputs of a university are evaluated, and specific outputs quantity conducted in any particular period is viewed as a proxy of quality.[103] In recent years, Australia introduced another component into its ERA to measure impact and engagement. Currently, the Australian Research Council has decided to pause ERA and review it as some argue that ERA has achieved its initial objectives and there is no point in running it again.[104]

Through the review of the literature on research evaluation and RESs, the following points can be extracted

First, experts and scholars in the field of research evaluation vary in their opinions and viewpoints. Peer review is assumed by many scholars and evaluators to be a quality control. Peer review is also a driver in the system of science organization for "quality control and innovation encouragement"[33] and "scholarly standing of researchers".[58]

Second, evaluation is in the direction of research purpose and shows the degree of achieving the intended goals of researchers, research units, universities, and disciplines. Besides, the evaluation has specific purpose. For example, REF in the UK focuses on resource allocation level, while the system of Netherlands highlights quality assurance.[84] Accordingly, the bibliometric analyses alone might not satisfy all the requirements of "quality control and assurance and resource allocation" in a RES, and a mix of quantitative and qualitative methods is essential.

Third, bibliometric analyses play a supportive role for peer review".[36,37,55,62,64,105] However, the problems of human resources, the needed time and costs, and bureaucracy level have caused some policy-makers and research managers to further use bibliometric indicators or indicator-based systems [84] to assess research performance. Some scholars have argued that bibliometric analysis is much more economic in terms of time and cost, particularly in hard sciences.[38,52,106] Bibliometric analysis cannot yet be regarded as an alternative for peer review.[32,53,60,64,107] The peer review is still preferred to bibliometric analysis[108] and it is important to consider the recent recommendations of the scientific authorities in using bibliometric indicators, such as the "Declaration on Research Assessment (DORA)".[109]

Fourth, bibliometric analysis has been reported to be an appropriate tool for defining the degree of scientific impact and production, particularly in natural sciences and in certain fields of humanities and social sciences.[32,38,54,58,59,106] Naturally, bibliometric analysis also includes negative consequences like the increased number of citations[81] and increased publications in the network of science. For instance, to obtain a high H index, we should publish more papers and get more citations. To attain a high level of productivity, documents need to be increasingly published. It is important to realize if the indicators practically measure what they are designed to measure, as individuals are trying to "adjust to the indicator value system by optimizing their indicator rather than their performance".[110] In addition, rankings such as defining the productivity of scholars and showing the effect of a university by citations are fundamentally one-dimensional and hence in opposition to the spirit of research as a quest for prosperity, health, comfort, and life betterment by answering the unknown. The research evaluation is indeed a multi-dimensional appraisal of performance, like REF in the UK.[84] Accordingly, to evaluate research impact and productivity, it is not enough to merely use bibliometric indicators such as the number of citations and the number of publications.

Note also that research assessment systems, methods, and indicators can strongly affect the status and achievement of an individual, center, discipline, or higher education system in a country, by defining research direction and pathway. This direction can motivate people to get more citations, publish further document, and provide a higher H index. It can also ensure science quality, development and innovation, and enhance welfare, comfort, and life of people. The mission of academics is teaching, conducting research, and providing service to society.

CONCLUSION

In this study, two RESs, the national-level performance evaluation systems of the UK and Australia, were reviewed and compared. By selecting these countries, research evaluation processes were examined based on different attributes as follows:

1. A system with highly selective data collection related to research budgeting (UK).
2. A system based on a comprehensive set of data provided by institutions but not related to research budgeting (Australia).

The comparison showed that in terms of input, almost two evaluation systems emphasize the components of human resources, finance and infrastructure in the evaluation. In terms of type and method of evaluation, it was found that there are both quantitative (citation and publishing analysis and domestic

and international research income) and qualitative (survey of domestic or international experts) approaches in two systems with differences. In terms of evaluation criteria, each system is based on different criteria. In UK, the main indicator for evaluating research output is the established frameworks for peer reviwers to determine the quality of the output, its impact, etc. Of course, the count of citations per publication criterion is used as an aid criterion. In Australia, tow indicators of peer review and citation (including world citations per paper (cpp) benchmarks and Australian higher education provider cpp benchmarks, distribution of papers based on world centile thresholds, and distribution of papers against relative citation impact classes) use to evaluate research. In terms of the evaluation unit, the two systems have almost the same structure, and the evaluation is done by specialized panels. In terms of ranking scale, three different scales were identified: the 5-star scale (the UK and Australia), the quantitative scale based on resource data or research tools, and the descriptive scale. In terms of the evaluation element, the quality of research outputs is evaluated in two systems, and in the UK, the two elements of impact and research environment are also evaluated. In terms of evaluated research outputs, publications (journal papers, monographs, and book chapters) and patents are evaluated in two systems, and items such as designs in these systems are evaluated.

With a national system that covers the whole country and is associated with a significant level of budget, it seems the UK evaluation system performs well. The country's assessments provide accountability for public investment in research and provide evidence of the benefits of these investments so that research funders can use the results of evaluations to selectively allocate their grants to research institutes. In addition, these evaluations can create a rich evidence base for informing strategic decisions about national research priorities and motivate research institutes and individual researchers to show a strong performance. Australia has a systematic evaluation framework that is not related to research budgeting. However, this evaluation system is considered a key indicator of performance in contracts between the Australian Government and higher education institutions and has already developed the strategy of the Australian Government Research Workforce. The results of the Australian ERA also provide valuable information on allocating financial resources by creating sustainable excellence in universities, developing new quality management standards and training standards for education and research, designing a national scale to identify research areas and disciplines that might lead to the development of Australian higher education institutions, and defining the emerging research areas. Thus, it can be said that while the two evaluation approaches examined in this study are different in their objectives and results, there are not many structural and content differences in these systems at the national and institutional level, and the main system components (i.e., input, process, output, and impact) can be extracted more or less from these systems.

One limitation of this study is that these RESs are not the same and each has characteristics that distinguish it from other systems. The analysis and comparison of concrete effects, outcomes, and consequences of these RESs (in terms of scientific, social, economic, technological, and industrial, etc.,) were beyond the scope of this study. Despite these limitations, this study conducted a relatively comprehensive review of RESs.

The structural and content elements of RESs identified in this study can inform the development of such systems in other countries. The results can be beneficiary for managing and policymaking for research assessment units on micro/macro levels. Our study of centralized in England and Australia assessment systems provides further insight on the types and selection of research assessment designs. Moreover, this study can help to identify the necessary processes for constant monitoring of the progress of research activities and guide and refine research processes so that research budgets are spent on research that can lead to the country's development, economic growth, and prosperity. The formation of national systems for evaluating research will affect not only the quality of research but also the purposefulness and significance of research and its impact on scientific progress and people's lives.

In this study, we compared two retrospective national research evaluation systems. Our future work will focus on examining prospective research evaluation systems, which involve pre-performance evaluation. Comparing the research results of both retrospective and prospective approaches can help guide the selection of these systems and assess their efficiency in research excellence. We also evaluated two national systems in this study, but it will be necessary to compare countries with national and state research evaluation systems in future studies. These comparisons can aid in the selection of research evaluation systems. Additionally, it's important to investigate the scientific and technological development in countries with national research evaluation systems and state systems to determine the efficiency and effectiveness of these systems. Furthermore, since some research evaluation systems rely on peer review while others emphasize bibliometric quantitative indicators, and some use both methods, comparing the quality of outputs and achievements and the scientific and technological development of countries with these systems can be effective in choosing and using these systems.

ACKNOWLEDGEMENT

This research was supported by the University of Isfahan and the Iran National Science Foundation (INSF) under Project No. 98018007.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest.

REFERENCES

1. Whitley R. Changing governance of the public sciences. In: Whitley R, Gläser J, editors. The changing governance of the sciences. Sociology of the sciences yearbook. Dordrecht: Springer; 2007; 26: 3-27. doi: 10.1007/978-1-4020-6746-4_1.
2. Cozzens SE. Options for the future of research evaluation. In: Cozzens SE, Healey P, Rip A, Ziman J, editors. The research system in transition. Dordrecht: Kluwer Publishers; 1990. p. 281-93. doi: 10.1007/978-94-009-2091-0_22.
3. Hicks D. Performance-based university research funding systems. Res Policy. 2012; 41(2): 251-61. doi: 10.1016/j.respol.2011.09.007.
4. Geuna A, Piolatto M. Research assessment in the UK and Italy: costly and difficult, but probably worth it (at least for a while). Res Policy. 2016; 45(1): 260-71. doi: 10.1016/j.respol.2015.09.004.
5. Butler L. What happens when funding is linked to publication counts? In: Moed HF, Glanzel W, Schmoch U, editors. Handbook of quantitative science and technology. Dordrecht: Kluwer Academic Publishers Publishers; 2004. p. 340-89.
6. Butler L. "Impacts of Performance-based Research Funding Systems: A Review of the Concerns and the Evidence", performance-based funding for public research in tertiary education institutions: workshop proceedings. Paris: OECD; 2010. p. 127-65.
7. Adams J, Gurney K. Funding selectivity, concentration and excellence-How good is the UK's research? Oxford: Higher Education Policy Institute; 2010.
8. Wang J, Hicks D. Detecting structural change in university research systems: A case study of British research policy. Res Eval. 2013; 22(4): 258-68. doi: 10.1093/reseval/rvt016.
9. Aagaard K, Bloch C, Schneider JW. Impacts of performance-based research funding systems: the case of the Norwegian Publication Indicator. Res Eval. 2015; 24(2): 106-17. doi: 10.1093/reseval/rvv003.
10. Good B, Vermeulen N, Tiefenthaler B, Arnold E. Counting quality? The Czech performance-based research funding system. Res Eval. 2015; 24(2): 91-105. doi: 10.1093/reseval/rvu035.
11. Aagaard K, Schneider JW. Some considerations about causes and effects in studies of performance-based research funding systems. J Inf. 2017; 11(3): 923-6. doi: 10.1016/j.joi.2017.05.018.
12. Engels TC, Guns R. The Flemish performance-based research funding system: A unique variant of the Norwegian model. J Data Inf Sci. 2018; 3(4): 45-60. doi: 10.2478/jdis-2018-0020.
13. Checchi D, Malgarini M, Sarlo S. Do performance-based research funding systems affect research production and impact? Higher Educ Q. 2018; 37(1): 45-69. doi: 10.1111/hequ.12185.
14. Krog Lind J. The missing link: how university managers mediate the impact of a performance-based research funding system. Res Eval. 2019; 28(1): 84-93. doi: 10.1093/reseval/rvy038.
15. Buckle RA, Creedy J. The evolution of research quality in New Zealand universities as measured by the performance-based research fund process. N Z Econ Pap. 2019; 53(2): 144-65. doi: 10.1080/00779954.2018.1429486.
16. Söderlind J, Berg LN, Lind JK, Pulkkinen K. National performance-based research funding systems: constructing local perceptions of research? In: Pinheiro R, Geschwind L, Foss Hansen H, Pulkkinen K, editors. Reforms, organizational change and performance in higher education. Cham: Palgrave Macmillan; 2019. p. 111-44. doi: 10.1007/978-3-030-11738-2_4.
17. Abramo G, D'Angelo CA. The different responses of universities to introduction of performance-based research funding. Res Eval. 2021; 30(4): 514-28. doi: 10.1093/reseval/rvab022.
18. Nicholson J, Howard K. A study of core competencies for supporting roles in engagement and impact assessment in Australia. J Aust Libr Inf Assoc. 2018; 67(2): 131-46. doi: 10.1080/24750158.2018.1473907.
19. Uludag K. Rethinking the publish or do not graduate paradigm: balancing graduation requirements and scientific integrity. In: Morris GR, Kozuch S, editors. Engaging higher education teachers and students with transnational leadership. IGI Global; 2024. p. 165-77. doi: 10.4018/979-8-3693-6100-9.ch009.
20. Nederhof AJ. Bibliometric monitoring of research performance in the social sciences and the humanities: a review. Scientometrics. 2006; 66(1): 81-100. doi: 10.1007/s11192-006-0007-2.
21. Prins AA, Costas R, van Leeuwen TN, Wouters PF. Using Google Scholar in research evaluation of humanities and social science programs: a comparison with Web of Science data. Res Eval. 2016; 25(3): 264-70. doi: 10.1093/reseval/rvv049, Google Scholar.
22. Norris M, Oppenheim C. Comparing alternatives to the Web of Science for coverage of the social sciences' literature. J Inf. 2007; 1(2): 161-9. doi: 10.1016/j.joi.2006.12.001.
23. Geuna A, Martin BR. University research evaluation and funding: an international comparison. Minerva. 2003; 41(4): 277-304. doi: 10.1023/B:MINE.0000005155.70870.bd.
24. Martin BR. "The Research Excellence Framework and the 'impact agenda": are we creating a Frankenstein monster? Res Eval. 2011; 20(3): 247-54. doi: 10.3152/095820211X13118583635693.
25. Williams K, Grant J. A comparative review of how the policy and procedures to assess research impact evolved in Australia and the UK. Res Eval. 2018; 27(2): 93-105. doi: 10.1093/reseval/rvx042.
26. Karlsson S. Evaluation as a travelling idea: assessing the consequences of Research Assessment Exercises. Res Eval. 2017; 26(2): 55-65. doi: 10.1093/reseval/rvx001.
27. Marques M, Powell JJ, Zapp M, Biesta G. How does research evaluation impact educational research? Exploring intended and unintended consequences of research assessment in the United Kingdom, 1986-2014. Eur Educ Res J. 2017; 16(6): 820-42. doi: 10.1177/1474904117730159.
28. Coryn CL, Hattie JA, Scriven M, Hartmann DJ. Models and mechanisms for evaluating government-funded research: an international comparison. Am J Eval. 2007; 28(4): 437-57. doi: 10.1177/1098214007308290.
29. Jonkers K, Zacharewicz T. Research performance based funding systems: A comparative assessment. Luxembourg: Publications Office of the European Union; 2016.
30. Zacharewicz T, Lepori B, Reale E, Jonkers K. Performance-based research funding in EU Member States—a comparative assessment. Sci Public Policy. 2019; 46(1): 105-15. doi: 10.1093/scipol/scy041.
31. Ochsner M, Peruginelli G. Handbook on research assessment in the social sciences; 2022. 26. National research evaluation systems and the social sciences. p. 416.
32. Abramo G, D'Angelo CA, Di Costa F. National research assessment exercises: a comparison of peer review and bibliometrics rankings. Scientometrics. 2011; 89(3): 929-41. doi: 10.1007/s11192-011-0459-x.
33. Horrobin DF. The philosophical basis of peer review and the suppression of innovation. J Am Med Assoc. 1990; 263(10): 1438-41. doi: 10.1001/jama.1990.03440100162024, PMID 2304222.
34. Moxam, H, and Anderson, J. 1992 'Peer review. A view from the inside', Science and Technology Policy; 5(1): 7-15.
35. Moed HF. The impact-factors debate: the ISI's uses and limits. Nature. 2002; 415(6873): 731-2. doi: 10.1038/415731a, PMID 11845176.
36. Van Raan AF. Fatal attraction: conceptual and methodological problems in the ranking of universities by bibliometric methods. Scientometrics. 2005; 62(1): 133-43. doi: 10.1007/s11192-005-0008-6.
37. Pendlebury DA. The use and misuse of journal metrics and other citation indicators. Arch Immunol Ther Exp (Warsz). 2009; 57(1): 1-11. doi: 10.1007/s00005-009-0008-y, PMID 19219526.
38. Abramo G, D'Angelo CA. Evaluating research: from informed peer review to bibliometrics. Scientometrics. 2011; 87(3): 499-514. doi: 10.1007/s11192-011-0352-7.
39. Sandström U, Van den Besselaar P. Funding, evaluation, and the performance of national research systems. J Inf. 2018; 12(1): 365-84. doi: 10.1016/j.joi.2018.01.007.
40. Zuccala A, Robinson-García N. Reviewing, indicating, and counting books for modern research evaluation systems. In: Glänzel W, Moed HF, Schmoch U, Thelwall M, editors. Springer handbook of science and technology indicators. Cham: Springer; 2019. p. 715-28. doi: 10.1007/978-3-030-02511-3_27.
41. Giménez-Toledo E, Mañana-Rodríguez J, Engels TC, Ingwersen P, Pölönen J, Sivertsen G, et al. Taking scholarly books into account: current developments in five European countries. Scientometrics. 2016; 107(2): 685-99. doi: 10.1007/s11192-016-1886-5.
42. Garrett-Jones S. International trends in evaluating university research outcomes: what lessons for Australia? Res Eval. 2000; 9(2): 115-24. doi: 10.3152/147154400781777278.
43. Grigson D, Stokes T. Use of peer review to evaluate research outcomes. Res Eval. 1993; 3(3): 173-7. doi: 10.1093/rev/3.3.173.
44. Lillis DA. Towards a new science evaluation system for New Zealand. Res Eval. 2000; 9(2): 145-50. doi: 10.3152/147154400781777287.
45. Garrett-Jones S, Aylward D. Some recent developments in the evaluation of university research outcomes in the United Kingdom. Res Eval. 2000; 9(1): 69-75. doi: 10.3152/147154400781777430.
46. Penfield T, Baker MJ, Scoble R, Wykes MC. Assessment, evaluations, and definitions of research impact: a review. Res Eval. 2014; 23(1): 21-32. doi: 10.1093/reseval/rvt021.
47. Reale E, Avramov D, Canhial K, Donovan C, Flecha R, Holm P, et al. A review of literature on evaluating the scientific, social and political impact of social sciences and humanities research. Res Eval. 2018; 27(4): 298-308. doi: 10.1093/reseval/rvx025.
48. Osiander CH. Lessons from mixed-method evaluations, An example from labor market research. Res Eval. 2021; 30(1): 90-101. doi: 10.1093/reseval/rvaa016.
49. Langfeldt L, Reymert I, Aksnes DW. The role of metrics in peer assessments. Res Eval. 2021; 30(1): 112-26. doi: 10.1093/reseval/rvaa032.
50. Abramo G, Cicero T, D'Angelo CA. Revisiting the scaling of citations for research assessment. J Inf. 2012; 6(4): 470-9. doi: 10.1016/j.joi.2012.03.005.
51. Brito R, Rodríguez-Navarro A. Research assessment by percentile-based double rank analysis. J Inf. 2018; 12(1): 315-29. doi: 10.1016/j.joi.2018.01.011.
52. Wallmark JT, Sedig KG. Quality of research measured by citation method and by peer review-A comparison. IEEE Trans Eng Manage. 1986;EM-33(4): 218-22. doi: 10.1109/TEM.1986.6447682.

53. Jacsò P. Pragmatic issues in calculating and comparing the quantity and quality of research through rating and ranking of researchers based on peer reviews and bibliometric indicators from Web of Science, Scopus and Google Scholar. Online Inf Rev. 2010; 34(6): 972-82. doi: 10.1108/14684521011099432.
54. Thomas PR, Watkins DS. Institutional research rankings via bibliometric analysis and direct peer review: A comparative case study with policy implications. Scientometrics. 1998; 41(3): 335-55. doi: 10.1007/BF02459050.
55. Nederhof AJ, Van Raan AF. A bibliometric analysis of six economics research groups: A comparison with peer review. Res Policy. 1993; 22(4): 353-68. doi: 10.1016/0048-7333(93)90005-3.
56. Aksnes DW, Taxt RE. Peers reviews and bibliometric indicators: A comparative study at Norvegian University. Res Eval. 2004; 13(1): 33-41. doi: 10.3152/147154404781776563.
57. Checchi D, Ciolfi A, De Fraja G, Mazzotta I, Verzillo S. Have you read this? An empirical comparison of the British REF peer review and the Italian VQR bibliometric algorithm. Economica. 2021; 88(352): 1107-29. doi: 10.1111/ecca.12373.
58. Fedderke JW. The objectivity of national research foundation peer review in South Africa assessed against bibliometric indexes. Scientometrics. 2013; 97(2): 177-206. doi: 10.1007/s11192-013-0981-0.
59. Mryglod O, Kenna R, Holovatch Y, Berche B. Comparison of a citation-based indicator and peer review for absolute and specific measures of research-group excellence. Scientometrics. 2013; 97(3): 767-77. doi: 10.1007/s11192-013-1058-9.
60. Baccini A, De Nicolao G. Do they agree? bibliometric evaluation versus informed peer review in the Italian research assessment exercise. Scientometrics. 2016; 108(3): 1651-71. doi: 10.1007/s11192-016-1929-y.
61. Rodríguez-Navarro A, Brito R. Like-for-like bibliometric substitutes for peer review: advantages and limits of indicators calculated from the ep index. Res Eval. 2020; 29(2): 215-30. doi: 10.1093/reseval/rvaa002.
62. King J. The use of bibliometric techniques for institutional research evaluation: A study of avian virology research. Scientometrics. 1988; 14(3-4): 295-313. doi: 10.1007/BF02020081.
63. Nederhof AJ, Van Raan AF. Peer review and bibliometric indicators of scientific performance: A comparison of cum laude doctorates with ordinary doctorates in physics. Scientometrics. 1987; 11(5-6): 333-50. doi: 10.1007/BF02279353.
64. Rinia EJ, Van Leeuwen TN, Van Vuren HG, Van Raan AF. Comparative analysis of a set of bibliometric indicators and central peer review criteria. Res Policy. 1998; 27(1): 95-107. doi: 10.1016/S0048-7333(98)00026-2.
65. Rinia EJ, Van Leeuwen TN, Van Vuren HG, Van Raan AF. Influence of interdisciplinarity on peer-review and bibliometric evaluations in physics research. Res Policy. 2001; 30(3): 357-61. doi: 10.1016/S0048-7333(00)00082-2.
66. Oppenheim C, Summers MA. Citation counts and the Research Assessment Exercise, part VI: Unit of assessment 67 (music). Inf Res Int Electron J. 2008; 13(2).
67. Norris M, Oppenheim C. Citation counts and the Research Assessment Exercise V. J Doc. 2003; 59(6): 709-30. doi: 10.1108/00220410310698734.
68. Bertocchi G, Gambardella A, Jappelli T, Nappi CA, Peracchi F. Bibliometric evaluation vs. informed peer review: evidence from Italy. Res Policy. 2015; 44(2): 451-66. doi: 10.1016/j.respol.2014.08.004.
69. Bruns SB, Stern DI. Research assessment using early citation information. Scientometrics. 2016; 108(2): 917-35. doi: 10.1007/s11192-016-1979-1.
70. Barker K. The UK Research Assessment Exercise: the evolution of a national research evaluation system. Res Eval. 2007; 16(1): 3-12. doi: 10.3152/095820207X190674.
71. Himanen L, Puuska HM. Does monitoring performance act as an incentive for improving research performance? National and organizational level analysis of Finnish universities. Res Eval. 2022; 31(2): 236-48. doi: 10.1093/reseval/rvac004.-.
72. Chatterjee B, Cordery CJ, De Loo I, Letiche H. The spectacle of research assessment systems: insights from New Zealand and the United Kingdom. Acc Aud Acc J. 2020; 33(6): 1219-46. doi: 10.1108/AAAJ-01-2019-3865.
73. Franceschini F, Maisano D. Critical remarks on the Italian research assessment exercise VQR 2011-2014. J Inf. 2017; 11(2): 337-57. doi: 10.1016/j.joi.2017.02.005.
74. Diezmann CM. Understanding research strategies to improve ERA performance in Australian universities: circumventing secrecy to achieve success. J Higher Educ Policy Manag. 2018; 40(2): 154-74. doi: 10.1080/1360080X.2018.1428411.
75. Calabrese A, Capece G, Costa R, Di Pillo F, Giuffrida S. A 'power law' based method to reduce size-related bias in indicators of knowledge performance: an application to university research assessment. J Inf. 2018; 12(4): 1263-81. doi: 10.1016/j.joi.2018.10.005.
76. Greco G, Scarcello F. Fair division rules for funds distribution: the case of the Italian Research Assessment Program (VQR 2004-2010). Intell Artif. 2013; 7(1): 45-56. doi: 10.3233/IA-130042.
77. Xiaochun W, Dan J. Scientific and ethical reflections on academic corruption in universities: on the science research evaluation system in China's universities. Chin Educ Soc. 2007; 40(6): 67-76. doi: 10.2753/CED1061-1932400610.
78. Butler L. Modifying publication practices in response to funding formulas. Res Eval. 2003; 12(1): 39-46. doi: 10.3152/147154403781776780.
79. Van den Besselaar P, Heyman U, Sandström U. Perverse effects of output-based research funding? Butler's Australian case revisited. J Inf. 2017; 11(3): 905-18. doi: 10.1016/j.joi.2017.05.016.
80. Schneider JW, Aagaard K, Bloch CW. What happens when national research funding is linked to differentiated publication counts? A comparison of the Australian and Norwegian publication-based funding models. Res Eval. 2016; 25(3): 244-56. doi: 10.1093/reseval/rvv036.
81. Abramo G, D'Angelo CA, Grilli L. The effects of citation-based research evaluation schemes on self-citation behavior. J Inf. 2021; 15(4): 101204. doi: 10.1016/j.joi.2021.101204.
82. Edwards MA, Roy S. Academic research in the 21st Century: maintaining scientific integrity in a climate of perverse incentives and hypercompetition. Environ Eng Sci. 2017; 34(1): 51-61. doi: 10.1089/ees.2016.0223, PMID 28115824.
83. Honig B, Bedi A. The fox in the hen house: a critical examination of plagiarism among members of the academy of management. Acad Manag Learn Educ. 2012; 11(1): 101-23. doi: 10.5465/amle.2010.0084.
84. Hazelkorn E 2010. Pros and cons of research assessment, in World Social Science Report. Knowledge divides 2010. UNESCO Press.
85. Bereday GZ. Comparative method in education. NY: Holt, Rinehart & Winston; 1964.
86. Technopolis; 2010. 'REF research impact pilot exercise lessons-learned project: feedback on pilot submissions'. [cited Dec 5 2021] Available from: http://www.ref.ac.uk/media/ref/content/pub/refresearchimpactpilotexerciselessons-learnedprojectfeedbackonpilotsubmissions/re02_10.pdf.
87. Schiene C, Schimank U. Research evaluation as organisational development. In: Whitley R, Gläser J, editors. The changing governance of the sciences. Sociology of the sciences yearbook. Vol. 26. Dordrecht: Springer; 2007. p. 171-90. doi: 10.1007/978-1-4020-6746-4_8.
88. HEFCE. The nature, scale and beneficiaries of research impact. An initial analysis of Research Excellence Framework (REF) 2014 impact case studies. King's College London and Digital Press Science; 2015. Available from: https://www.kcl.ac.uk/policy-institute/assets/ref-impact.pdf.
89. UK-REF. Guidance on submissions; 2021. [cited Dec 11 2021] Available from: https://www.ref.ac.uk/publications/guidance-on-submissions-201901.
90. UK-REF. Building on Success and Learning from Experience: an independent review of the Research Excellence Framework. London: Department for Business, Energy and Industrial Strategy. Available; 2016.
91. UK-REF. Panel criteria and working methods; 2021. Available from: https://www.ref.ac.uk/publications/panel-criteria-and-working-methods-201902/.
92. Marjanovic S, Hanney S, Wooding S. A historical reflection on research evaluation studies, their recurrent themes and challenges. Rand Europe technical report; 2009. Available from: http://www.rand.org/pubs/technical_reports/2009/rand_tr789.pdf.
93. Glänzel W, Moed HF, Schmoch U, Thelwall M, editors. Springer handbook of science and technology indicators. Springer handbooks; 2019. [cited Jan 19 2022] Available from: https://doi.org/10.1007/978-3-030-02511-3.
94. Australian Research Council. Submission guidelines; 2018. Available from: https://www.arc.gov.au/file/3781/download?token=Wq9o-CbM.
95. Crowe SF, Prado C. Excellence in research in Australia: the souffle keeps on rising. Aust Psychol. 2020; 55(5): 468-87. doi: 10.1111/ap.12453.
96. ERA. 'Submission Guidelines'; 2018. [cited Jan 23 2022] Available from: https://www.arc.gov.au.
97. Kochenkova A, Grimaldi R, Munari F. Public policy measures in support of knowledge transfer activities: a review of academic literature. J Technol Transf. 2016; 41(3): 407-29. doi: 10.1007/s10961-015-9416-9.
98. ERA. Discipline matrix; 2018. Available from:https://www.griffith.edu.au/__data/assets/excel_doc/0025/294406/ERA-2018-Discipline-Matrix.xlsxandved=2ahUKEwii_uyHp472AhWeppUCHYw5BoAQFnoECAkQAgandusg=AOvVaw0-hXI6v0vWSggu2hBNqD3h.
99. Perry LB. Assessing the performance of educational research in Australian universities: an alternative perspective. Higher Educ Res Dev. 2017: 1-16. doi: 10.1080/07294360.2017.1355893.
100. ERA. 'Peer Review Handbook'; 2018. [cited Jan 24 2022] Available from: https://www.arc.gov.au.
101. Gläser J. The social orders of research evaluation systems. In: Whitley R, Gläser J, editors. The changing governance of the sciences. Sociology of the sciences yearbook. Vol. 26. Dordrecht: Springer; 2007. p. 245-66. doi: 10.1007/978-1-4020-6746-4_12.
102. Gläser J, Laudel G. Evaluation without evaluators. In: Whitley R, Gläser J, editors. The changing governance of the sciences. Sociology of the sciences yearbook. Vol. 26. Dordrecht: Springer; 2007a. p. 127-51. doi: 10.1007/978-1-4020-6746-4_6.
103. Woelert P. "The logic of escalation" in performance measurement: an analysis of the dynamics of a research evaluation system. Policy Soc. 2015; 34(1): 75-85. doi: 10.1016/j.polsoc.2015.03.004.
104. Ross J. Australian research assessment exercise 'has achieved purpose'. Times Book Company Higher Education; 2022, Nov 11. Available from: https://www.timeshighereducation.com/news/australian-research-assessment-exercise-has-achieved-purpose.
105. Petersohn S, Heinze T. Professionalization of bibliometric research assessment. Insights from the history of the Leiden Centre for Science and Technology Studies (CWTS). Sci Public Policy. 2018; 45(4): 565-78. doi: 10.1093/scipol/scx084.

106. Abramo G, Cicero T, D'Angelo CA. National peer-review research assessment exercises for the hard sciences can be a complete waste of money: the Italian case. Scientometrics. 2013; 95(1): 311-24. doi: 10.1007/s11192-012-0875-6.

107. Abramo G, D'Angelo CA. The VQR, Italy's second national research assessment: methodological failures and ranking distortions. J Assoc Inf Sci Technol. 2015; 66(11): 2202-14. doi: 10.1002/asi.23323.

108. Brinn T, Jones MJ, Pendlebury M. Measuring research quality: peer review 1, citation indices 0. Omega. 2000; 28(2): 237-9. doi: 10.1016/S0305-0483(99)00048-1.

109. San Francisco declaration on research assessment; 2012. 'Declaration on research assessment'. Available from: https://sfdora.org/.

110. Retzer V, Jurasinski G. Towards objectivity in research evaluation using bibliometric indicators - A protocol for incorporating complexity. Basic Appl Ecol. 2009; 10(5): 393-400. doi: 10.1016/j.baae.2008.09.001.

Cite this article: Baghjanati M, CheshmehSohrabi M, Jamali HR. A Comparative Study of Research Evaluation Systems in the United Kingdom and Australia. J Scientometric Res. 2025;14(2):617-31.