

Developing and Testing a Method to Measure Academic Societal Impact

Paul Phillips¹, Luiz Moutinho² and Pedro Godinho^{3,*}

¹Kent Business School, University of Kent, Canterbury, CT2 7PE, UK;

²University of Suffolk, UK and University of South Pacific, Fiji;

³CeBER and Faculty of Economics, University of Coimbra, Av. Dias da Silva, 165, 3004-512, Coimbra, Portugal.

*Corresponding author. E-mail: pgodinho@fe.uc.pt; phone number: +351 239790571

Abstract

This paper aims to extend understanding of the business and societal impact of academic research. From a business school perspective, it has taken stock of the role of academic research and relevance in business and society. The proposed conceptual framework highlights the forces influencing the pursuit of academic rigour and relevance in scholarly outputs. A theoretical model for measuring the societal impact of academic journal articles - the Academic Rigour and Relevance Index (AR2I) – was developed. This index comprises six key parameters, which are assessed by three stakeholder groups connected with academic research into business issues, these groups being: business practitioners, society, and academics.

The behaviour of the AR2I model was evaluated using the Monte Carlo simulation model. Taking into account the relationships between the standard deviations and the differences of classification between articles with different levels of rigour and relevance, it is demonstrated that the AR2I model is an effective tool.

Keywords: Societal impact; Academic research; Business impact; Science–society interactions.

Introduction

Universities are increasingly expected to pursue a third mission in addition to undertaking research and providing education. This third obligation concerns engagement with society through the application and exploitation of knowledge, and universities must demonstrate this relationship (de Jong *et al.*, 2014). Nowadays, however, academics are under increasing pressure to publish and deliver outcomes for their areas of specialization, and this pressure can lead to a focus on precision in non-relevant areas. In business and management research, the dynamics are characterized by the associated disciplines and boundaries. And in respect of those boundaries, it has been the preserve of scholars to decide what they deem to be legitimate research, and therefore, what to investigate (Guthrie & Parker, 2014). Wiek *et al.* (2014) have introduced a framework and methodological scheme for capturing the societal effects of participatory sustainability research. Such effects include the provision of quality products, knowledge gains, increased decision-making capacity, enhanced networks, and transformational changes.

To date, academic journal editors and reviewers remain the boundary setters, yet there is an increasing influence of governments, publishers, educational and research institutes being brought to bear in this respect, and this has led some academics to question their roles in terms of scholarship.

At the extreme, some academics may feel that by metricising impact, the government and academic managers may acquire additional power, which could hasten the devaluing or deskilling of the profession (Sutherland *et al.*, 2015; De Vita & Case, 2016). Nevertheless, it is believed that such impact is here to stay, and UK academics need encouragement to meld their research with the needs of the outside world through the introduction of a measurement of the effects, or 'impact', of academic research on wider society. Regarding support, the UK's Economic and Social Research Council emphasizes that research impact includes:

enhancing economic performance, increasing effectiveness of public services and policy, and enhancing the quality of life, health and creative output. The type of research impact can be broken down into academic, economic, and societal impact.

The purpose of this paper is to develop a theoretical model that intersects business and society, by measuring the societal impact of academic journal articles. This model is entitled the Academic Rigour and Relevance Index (AR2I), and it is also being presented in the context of the UK Research Excellence Framework (REF) which is increasingly emphasizing relevance under the rubric ‘need for impact’. The study seeks to make a number of contributions. Firstly, it highlights the importance of legitimacy within a business school context, which can enhance the successful management of the institution. Secondly, by delving further into what constitutes valid evidence of impact beyond traditional bibliometrics, it is possible to help to advance the debate in another direction – societal impact. The business and management literature appears to have reached an impasse about how to define and demonstrate impact. Moreover, given the current funding model of UK higher education, business schools need to become more accountable for the money they receive for, and spend on research. Thirdly, for those business schools that are struggling to produce internationally-rated research, alternative strategies exist such as focusing on the production of research for societal stakeholders. The following section briefly considers some of the projects that have sought to delve deeper than traditional quantitative bibliometric approaches.

The remainder of this paper is organized as follows: Section 2 draws attention to business school relevance and societal impact, and considers potential political and cultural issues. Section 3 develops a conceptual academic rigour and relevance index (AR2I) model within a conceptual framework. Section 4 briefly presents the Analytic Hierarchy Process method, which was used to assess the weights of the variables in the AR2I model. Section 5 outlines a

simulation model for an analysis of some aspects of the developed index, and also presents and briefly discusses the results of its application. Finally, the conclusion indicates the implications and limitations of the current work and appropriate avenues for future research.

Business school relevance, and societal impact

Relevance

Motivated in part by rankings, business schools flout their own rules to the effect that their research outputs should be impacting business and society in general, as opposed to having academic outputs which include oceans of paper with scant practical relevance. As long ago as the Gordon and Howell (1959) Ford and Carnegie Foundation Report, business schools have taken criticism for being too academic. And in the inaugural edition of the *Academy of Management Learning & Education*, Pfeffer and Fong (2002) question the success of the United States (US) business schools.

More than eighteen years ago, Tranfield and Starkey (1998) wrote an article entitled “The Nature, Social Organization and Promotion of Management Research: Towards Policy”. In this piece, the authors use the vocabulary of Mode 1 and Mode 2 research (Aken, 2005). The former approach relates to the production of academic and mono-disciplinary research, whereas the latter is of a more greater practical focus and transdisciplinarity. These differences in knowledge production has led to a lineage of the rigour-relevance debate (e.g. Hodgkinson and Starkey, 2011). They consider the periods of the 1990s and 2000s to 2011, and observe the infatuation with Mode 2 research, reminding the reader that Mode 1 has not been universally outmoded. In line with this, Willmott (2012) provides a commentary and concludes that it is high time to raise the sights of business schools beyond a myopic notion of relevance. While business schools pride themselves on high level executive education as well as prestigious research, Tushman *et al.* (2007) point to a tendency for executive

education to be underutilised as a lever in shaping practice and research. Encouragingly, Aguinis *et al.* (2014) propose a pluralist concept of scholarly impact, where the focus is also on other stakeholders and, echoing this view, Flickinger *et al.* (2014) introduce the concept of legitimacy to the rigour-relevance debate arguing, for instance, that legitimacy is the missing piece in the field of entrepreneurial research.

Identifying effective strategies to translate research evidence is a priority. Multiple channels exist to disseminate, translate, and communicate research evidence. Some thought leaders have specifically advocated for researchers to play a direct role in research dissemination, particularly through social media. There is a certain level of public demand for scholars to embrace new modes of research dissemination, taking into account, for instance, the rapid growth and reach of social media to disseminate information. There is also some scepticism and confusion within the academic community about how best to use social media to disseminate relevant research. One conclusion is that while scholars will need to be engaged in evidence translation to inform key stakeholders, they may be best served by connecting with trusted intermediaries and knowledge brokers to promote efficient use of the best available evidence to answer the most timely society questions. Journals, traditional media and universities may be well positioned to invest in this capacity to curate research evidence and disseminate it using social media and other technologies. The public, civil society and other relevant stakeholders should be provided with timely information about all aspects of the national OGP process, including feedback on how their inputs are taken into account.

It seems that there is not an absolute concept of academic rigour – but rather there are degrees of it depending of the closeness of research to known theory, whether or not the research is supported by known theory and the credibility of the data from which inferences are drawn which clearly implies a very robust array of measurement and methodological procedures.

For this study, we concur with Gulati (2007), who succinctly explains academic rigor as being about methodological soundness and relevance being relevance to managers. Obviously, in the latter instance managers can operate in all areas of societal impact. This point will be explained in the following section.

Societal impact

Numerous terms describe societal impact, such as third stream activities, societal benefits, societal quality, usefulness, public values, knowledge transfer and societal relevance (Bornmann, 2012a). Globerman (2011) states that the social responsibility of business has been debated in academic and practitioner literatures. Olmos-Penuela *et al.* (2014) conclude in their study that measures promoting a focus on the societal impact of research could enhance engagement in knowledge transfer activities. Clearly, business schools have a role to play in bringing business and society back together (Dyck *et al.*, 2011). From a definitional standpoint society can have a broad meaning. In the REF2014, UK exercise, which was a new system for assessing the quality of research in UK higher education institutions, societal impact was defined as a focus on: societal attitudes to and impacts of education, government directives and employment opportunities, lifestyle changes, changes in populations, distributions and demographics, the societal impact of different cultures. For this study we would specifically include government, public sector, non-profit sector and community-based organisations.

Consequently, a body of literature continues to grow, drawing attention for the need to review how research can be more effectively connected to practice and policy setting (Alvesson & Sandberg, 2013; Grey, 2010; Oswick *et al.*, 2011; Starbuck, 2006). This rigour-relevance gap (Baldrige *et al.*, 2004; Daft & Lewin, 2008; Kieser & Leiner, 2009), and attempts to address it (Hodgkinson & Rousseau, 2009; Starkey *et al.*, 2009) have reached stalemate. Editors of influential journals such as the *Journal of Business Research*

(Woodside, 2009), *Journal of Management Studies* (Clark & Wright, 2009), *Academy of Management Journal* (Bartunek *et al.*, 2006), and *British Journal of Management* (Wood & Budhwar, 2014) have all joined the debate. A common thread echoed by these editors is the absence of ground-breaking, competently executed and really interesting papers. Encouragingly, the new editor of the *Academy of Management Journal* states that the *AMJ* seeks to test, extend, or build strong theoretical frameworks, while examining issues with importance for management practice (George, 2014). Yet, incremental research rather than innovation and creativity seems to dominate the efforts of academics (Alvesson & Sandberg, 2013).

Academic rigour and relevance index (AR2I)

The term ‘academic relevance index’ can incorporate traditional academic impact metrics, but the focus of this study is the creation of a novel approach to generate theory and research that is relevant to business and society. The role of business in society and the accompanying responsibilities that transpire from that role is a highly contentious and debated topic. A business system is a methodical procedure or process that is used as a delivery mechanism for providing specific goods or services to customers. The economist Milton Friedman famously contended that the “business of business is business” and thus it has only one responsibility and that is to generate profit for shareholders (while adhering to the law). Contrary to that argument is thinking that recognizes business as a system in society that is affected by and affects other systems in society (such as the surrounding community, government bodies, other types of organizations, the natural environment, etc.). Thus business needs to work with these systems to attain its economic goals in a way that will also benefit the system (society) as a whole. This reasoning would recognise and include influences on the government and the public sector, the not-for-profit or philanthropic sectors, unions or community

organisations, although the latter are more clearly seen as part of the role of societal stakeholders.

Given the paucity of prior social science-led measurement studies, it starts by adopting a medical perspective. This could consider reducing social inequalities, which disadvantage the long-term health of populations (Smith, 2001). However, it has proved much harder to develop reliable and meaningful measures to assess the societal impact of research (Bornmann, 2012b). Patent data can be used to measure the business impact, but there appears to be no societal impact equivalent. As Niederkrotenthaler *et al.*, (2011) state, the social impact of research has been postulated rather than demonstrated. Specific studies that assess societal impact include Mansfield (1991, 1998) and Beise and Stahl (1999). Mansfield's (1991) survey asked respondents to estimate the proportion of their products or processes that could not have been developed in the last ten years without academic research. The results revealed a figure of approximately 10%. When followed up in Mansfield's later study in 1998, the corresponding figures were 15% for new products and 11% for new processes. Beise and Stahl (1999), applying a similar approach found that less than 10% introduced innovations between 1993 and 1995 that would not have been made without public research.

Civic participation is a core component of open government. This policy of open government is carried through a multi stakeholder process, with the active engagement of citizens and civil society. The collaboration of citizens, civil society, political and official champions and other stakeholders, like business systems, is essential to developing, securing and implementing lasting open government reforms that would benefit society as a whole. The public, civil society and other relevant stakeholders should be provided with timely information about all aspects of the academic research dissemination process, including feedback on how their inputs are taken into account. Spaces and platforms for dialogue and

co-creation should be planned - An inclusive and ongoing dialogue should be facilitated using a variety of spaces and platforms appropriate to the particular research thematic context. In this research study , the role of government can be taken as a proxy entity due to its main functioning in this context being indirectly relevant.

The conceptual framework

The AR2I Index is supported by a conceptual framework which is depicted in Figure 1. It includes seven major constructs that encapsulate the philosophy, ethos, and articulation of thoughts behind the conception of the index. They are the political system, the typology of the business school (Ivory *et al.*, 2006), the cultural system, the academic reward system, tangible research outputs, academic rigour, and academic relevance.

(Please insert Figure 1 around here)

Universities' efforts to enhance the commercial value of scientific research is causing increased politicization of government research funding, a growing winner-take-all contest, and subtle, but potentially profound, changes in the culture of academic research. Politicization may contribute to the fragmentation of knowledge claims, adding to varying discourse coalitions in which knowledge claims are aligned with political claims. Politicization clearly also influences knowledge production in terms of support to specific research areas as well as for specific research paradigms.

A business school education no longer guarantees access to lucrative careers in the best firms (Datar *et al.*, 2010), and given this scenario, it is believed that if business schools are to survive and prosper in today's post 2014 REF environment, new strategies are required. The work of Ivory *et al.* (2006) suggests four potential typologies for business schools to pursue, and provides guidance for the business school dean. For example, the possibilities are raised regarding positioning within the spectrum of social sciences research, or taking more of a leaning towards a professional school vision, towards the liberal arts, or indeed being more

knowledge-based. Ultimately, this is a critical indication of the school's contributory roles to academia, business systems, and society.

Societal impact has the potential to benefit individuals within the greater community who are in a position to be affected by it as its evaluators and judges. This will mean that some types of business schools are more likely to prosper under the new regime. Our logic of argument is that the impact agenda will strengthen those business schools that aspire to be professional schools rather than social science schools. Interestingly, governments have moved towards investing in research proposals “not for its intrinsic worth, nor to win esteem and still less to indulge researchers' curiosity, but for contribution to economic development” (Moodie, 2006). Thus, the AR2I can help social science business schools too.

The cultural and political systems, along with the typology of the schools, form the basis of the framework, given their role in shaping academic research. In particular, the typology and scholarly orientation of a business school is a paramount concern. The role of business schools and their output in the areas of teaching, research and community service is seen as a fundamental asset and ‘feeder’ to the contemporary knowledge economy. Therefore, there is a clear link in our cultural system that reinforces science-society interaction research.

The academic reward system will then have a decisive influence on the academic outputs. Although the four typologies of business schools are not mutually exclusive, there will be a differing focus on the recruitment need for each category. This will affect the retention strategy, which will be built into the academic reward system. For example, the social science school will need PhD qualified staff, who have been well trained in scholarship. Hence academic rewards for these staff will be dependent upon the attainment of publication in 4* academic journals. However, a newer university business school may pursue a professional school model and will need staff who can make impact through their teaching and outputs geared towards the societal market. This suggests that there will be differing emphasis on

academic rigour and relevance with a corresponding reward system. The very academic research process and its major pillars of scholarly robustness – research philosophy, theoretical development and nomological validity, research modelling and questions, measurement scaling, data analysis, research bias and contamination minimisation – currently remain the basic guarantors of academic rigour. The outputs of the research process will be dependent upon the contextual issues raised in the conceptual framework.

The last construct, relevance, is a dominant premise behind the construction of the whole index. Academic research and intellectual contributions in many fields have the potential to influence policy-making, policy evaluation and public understanding of policy. Society relevance of scholarly research can clearly influence academic credibility. The micro-sociological analysis of the general public's understanding of science is increasingly a factor to be considered. However, the paramount element rests in the role of academic research for improving the lives of human beings and the ultimate wellbeing of society. This discussion has considerable ideological overtones, where the contribution of the market to welfare and well-being is at stake. Welfare is usually conceptualized in material terms, and we surmise that both market and society can contribute to welfare and well-being. A market society is a way of life where market relations and market incentives and market values come to dominate all aspects of life. The public-understanding-of-science movement in the UK has not spent much time articulating its own understanding of science: science is taken as given. Nor has the public been examined too closely – the assumption is that we all know who they are. Understanding has also been neglected, though 'understanding' has clearly if tacitly been equated with 'knowledge' – factual knowledge of the content of science. Understanding is also equated with 'appreciation' – of the scientific enterprise, and of particular innovations. Putting these two equations together gives 'knowledge = appreciation'. It is clear that many in

the scientific and policy communities in the UK who want to further the public understanding of science are really concerned with increasing the public's appreciation of science.

The role and influence of the media in modern society contributes to, creates, directs and represent that society's 'culture' Consideration of the role, influence and development of the Media is embedded in the research construct “ Society “.

We integrate the nature and scope of the "mass media" itself, its influence as a powerful force for maintaining social cohesion, for constructing or reproducing social consciousness, as a mediator of values and beliefs, and as the major sources of information and means of information control within modern societies.

The approach

In developing AR2I, a two-level hierarchical approach was followed. Firstly, the major factors that drive the rigour and relevance of an article were identified, ending up with six factors, referred to as the parameters of the model. Two of these parameters are related to the significance of the contribution and academic scholarly intelligence. Measures for these parameters include the robustness of academic works regarding the article classification and the impact assessed by the reviewer, as well as the broader assessment concerning the timeliness of topics and methods. The other four remaining parameters are related to the impact on business systems; society and citizens; implications and recommendations; and citations. An analysis was then conducted to determine what variables were best suited to measure each parameter. Finally, there were sixteen variables: Classification by the Reviewer; Perception by the Reviewer; Pattern Recognition of the topic; Growth, Decline, Stability of the Topic; Read by Executives; Executive Interest; Change Mindsets; Take Action; Impact on Society by Executives; Interest in the Topic; Could make an improvement to Personal Productivity and Solutions Assembly; Relevance to the Whole Community; Judgment Perceptions by Business Systems; Judgment Perceptions by Citizens; Number of

Citations/Impact Factor of the journal. The hierarchy of parameters and variables is illustrated in Figure 2 and in Table 1, which also provides the references in which each parameter is supported.

(Please insert Figure 2 around here)

(Please insert Table 1 around here)

The framework conceptualizes the measurement of societal impact and provides a basis for empirical analyses and subsequent policy recommendations. Table 1 illustrates that AR2I: (i) can provide a concept to capture direct relevance to business and society, (ii) allows for business and society actors' perspective on relevance, and (iii) indicates a reliable picture of the precise societal impact gaps in current research articles. The approach does not conflate business with society, but both remain very important actors. Academic research in the domain of management scholarship, though steeped in scientific and methodological rigour, is generally found to be of little relevance to practice. The rigour-relevance debate has been revisited in light of recent developments and with special reference to the management research scenario. The central thesis of the argument is that the gulf between rigour and relevance needs to be bridged to make academic research more relevant to business organizations and practitioners. They can usefully offer some inputs and suggestions to enhance the relevance of academic research to the whole spectrum of business/commercial activities. A practitioner is a person who practice business. This means that this practitioner is currently in business. Practitioners are people that have current practical experience in business, e.g., owners ,consultants, managers. Just as better management practices can create incredible value for society, ineffective management can have a devastating impact on firm performance and make thousands or, potentially, millions of people worse off.

AR2I has been designed to respond to a pressing concern that has been mounting in recent years, that being the need to measure the societal impact derived from academic research in management. The proposed model goes well beyond the traditional yardsticks of academic citations and journal impact factors. Societal impact has been postulated and discussed in academic and policy-making circles, but very seldom demonstrated and tackled with an accurate and encompassing measurement approach. AR2I reflects this project's major goal in attempting to foster productive interactions between the scientific community in management and two key stakeholder cohorts - business systems, and society at large.

In order to properly integrate the rigour and impact of academic research, the perspectives of three groups of stakeholders were considered. Firstly, the opinion of academics was taken into account, since they are the main actors and the people in the best position to judge the rigour of research. Secondly, the perspective of businesspeople was introduced into the index because of the impact of research on the economy and on companies. And finally, the relevance to society as a whole was incorporated through the perspective of the citizens. It is recognized that the inclusion of citizens in an assessment of academic research may be controversial, because they will usually be laypersons, and hence, individuals without specific knowledge in a particular field (in our case, without knowledge about the scientific process). However, they are fundamental stakeholders of academic research and they will usually be well aware of the benefits to society that could be sought and derived, from the pursuit and conduct of academic research in many different fields and domains of science. Also, quantified summaries of constituents' feedback can provide credible performance data to policy-makers and funders on the academic outcomes that matter most.

In this approach, the perspectives of the stakeholders are summarized in a composite measure, an index. There were several reasons for resorting to a composite measure. One is that complex or multi-dimensional issues can be summarized in a simple manner, making it

possible for policy-makers to obtain a tractable and representative sense of the situation as it stands in comparison with others. Because they provide a single estimate, composite indices are easier to interpret than the use of multiple benchmarks. It is believed that the commitment to regularly produce and update quantitative ratings facilitates communication with ordinary citizens, showing the commitment of an academic organization. Finally, and because of the simplicity of its presentation, such a composite index is also an important starting point for debate.

There are, however, other possible approaches for the analysis and presentation of the stakeholders' perspectives. For example, the REF incorporates case studies to ascertain research impact. However, after a careful analysis, it was noticed that qualitative case studies would be less useful than an index. In fact, they are limited by the sensitivity and integrity of the investigator (the researcher/evaluator) who is the primary instrument of data collection and analysis. In this situation, the investigator is left to rely on his/her instincts and abilities throughout most of the research effort. Both the readers of case studies and the authors themselves need to be aware of biases that can affect the final product. Further limitations involve the issues of reliability, validity, and generalizability. As Hamel (1993, p. 23) observes, "... the case study has basically been faulted for its lack of representativeness and its lack of rigour in the collection, construction, and analysis of the empirical materials that give rise to the study".

Elicitation of the values of the variables

The quantified modelling of the AR21 Index can be encapsulated in this stage-by-stage framework:

Conceptual framework which underpins the quantitative analysis.

1. Choice of parameters/variables

2. We had parameters and then variables in two different levels, and we wanted to synthesize all their values into a final index.

3. In order to do that, we had to choose a multicriteria method that allowed hierarchical preferences, in order to consider the lower level variables and the upper level parameters. AHP is probably the most utilized of such methods, as it provides a nice and simple protocol to select the weights by answering simple questions, therefore, this was our choice.

Apart from the need to collect traditional baseline numeric data, such as impacts and citations, it was also considered important to include information about the original assessment by the papers' referees and to gather opinions from the three stakeholder groups. Consequently, the values of thirteen variables were elicited from members of these three groups of stakeholders. The use of an interval scale with seven levels was considered adequate for the detail with which it was expected that individuals would be able to express their opinions regarding each variable, leading to the selection of a Stapel scale ranging from -3 to +3. In order to obtain a sufficient number of responses to gauge the practicality of the AR2I, it was considered reasonable to use a sample of 30 respondents. Hence, responses were elicited from ten members from each group of stakeholders, thereby providing opinion from ten academics, ten citizens, and ten business practitioners. Monte Carlo simulation analysis was used to assess the possible impact and distortions of using a limited sample of respondents. We are going through a time of profound change in our understanding of the ethics of applied social research.

The principle of voluntary participation requires that people not be coerced into participating in research. Closely related to the notion of voluntary participation is the requirement of informed consent. Essentially, this means that prospective research participants must be fully informed about the procedures involved in research and must give their consent to participate. There is also standard that is applied in order to help protect the

privacy of research participants. Almost all research guarantees the participants' confidentiality -- they are assured that identifying information will not be made available to anyone who is not directly involved in the study.

The sampling procedure is based on non-probabilistic approaches- judgemental, quota and purposive.

The study participants were not recipients of any type of reimbursement , incentive or prize as a pre-condition for their recruitment and involvement in the research .

The process for eliciting responses involved meetings with each group of stakeholders. Each participant received the papers in advance, and was asked to read a sub-set of papers and skim through the rest. Each paper was allocated to exactly one participant, meaning that each participant received one tenth of the papers. In the meeting, participants worked initially in dyads, discussing the most relevant aspects of the papers allocated to the elements of the dyad, and preparing a succinct presentation of these papers. The dyads presented the papers to the remaining participants, after which discussions about each paper were held. Thereafter, each participant rated the paper on each of the variables corresponding to the respective stakeholder group, using a 7-point Stapel scale. An average classification on each variable was calculated for each paper, and this classification was then rescaled to a 0-1 scale.

Aggregation of variables into an index

An important issue in building such an index was how best to aggregate the variables' values into a final index that synthesizes the relevance and rigour of an article. A detailed analysis of the variables led us to conclude that the citations and impact factors should be aggregated into a single variable, defined as the ratio between citations and impact factor. The reason for using such a ratio was to cope with the influence of the journal reputation in the number of citations. The number of variables to aggregate was thus reduced to 15. A

weighted sum was used for the aggregation of these variables. The mathematical representation of the AR2I model is, therefore:

$$\begin{aligned}
 AR2I = & \alpha_1 CR + \alpha_2 PR + && \text{(Significance of the contribution)} \\
 & \alpha_3 P_{AR} + \alpha_4 GDST + && \text{(Academic scholarly intelligence)} \\
 & \alpha_5 RE + \alpha_6 EI + \alpha_7 CM + \alpha_8 TA + \alpha_9 JSE + && \text{(Relevance to business systems)} \\
 & \alpha_{10} IT + \alpha_{11} PPSA + \alpha_{12} PWC + && \text{(Perceived content by society/citizens)} \\
 & \alpha_{13} JP_{BS} + \alpha_{14} JP_{CIT} + && \text{(Implications and recommendations)} \\
 & \alpha_{15} (CIT/IF) && \text{(Citations and impact factor)}
 \end{aligned}$$

In this expression, it is assumed that the variables are already in a 0–1 scale (for example, all of them have values between zero and one) and the α_i 's are the weights to be used. Given the hierarchical nature of the model, the Analytic Hierarchy Process (AHP) was identified as a suitable method for defining the weights.

The Analytic Hierarchy Process (AHP) and AR2I weights

The hierarchical structure underlying AR2I leads to a natural application of the AHP framework, considering first the parameters and then the variables composing each parameter. This natural fitting of the AR2I structure to the AHP framework, along with the simplicity of the judgments required by AHP, allows the definition of the required weights by addressing a limited number of relatively simple questions to experts. Additionally, there are several methodologies for judgment aggregation in AHP (Dong *et al.*, 2010), which make it easy to combine the opinions of different experts in a set of weights for AR2I.

In order to assess a set of weights for the AR2I model, the AHP was performed on the basis of judgments made by the authors of this article. All parameters were pairwise compared, and then all variables composing a parameter were similarly compared. These comparisons were initially made by each author individually. Instead of simply using a method for aggregating

different opinions, the differences in our comparisons were analysed, and the reasons for these discussed. From this discussion, it was possible to reach a consensus for each individual comparison. These comparisons were then used to calculate the weights of the overall parameters and those of the individual variables. The consistency indexes were calculated for all sets of comparisons, and found to be well within the acceptable levels for AHP application (the level of inconsistency was below 4% in all cases). The resulting weights are shown in Table 2.

(Please insert Table 2 around here)

A simulation-based analysis of the AR2I index

Simulation has been used to analyse the behaviour of mathematical models under certain conditions, or the impact of different policies (Reikard, 2005; Jääskeläinen *et al.*, 2007; Ensthaler & Giebe, 2014). The application of AR2I is based on a limited sample of respondents, who represent the whole population, and it is important to assess whether this sample is enough for ensuring a good level of confidence in the representativeness of the results. Additionally, it was agreed to gauge whether the results were very sensitive to the weights used. If they are very sensitive, small changes in the weights may have a large impact in the classifications. So, similarly to other authors (Ensthaler & Giebe, 2014), a first assessment of the model, based on Monte Carlo simulation, was performed.

The simulation model

In a model based on weights that incorporate some subjectivity, a large sensitivity of the final results to these weights is undesirable, since that would mean a lack of robustness in those results. A simulation model was applied to perform the first assessment of the sensitivity of the model to the weights, and also to assess how a limited sample of

respondents might introduce excessive variability in the results, or lead to difficulties in differentiating among articles with different rigour and relevance.

To tackle these issues, simulation was used to compare the classifications of articles with different characteristics, and to compare the impact of using different weights. The general flow of the simulation model is depicted in Figure 3.

(Please insert Figure 3 around here)

Each simulation run was used to determine the probability distribution of the classifications of an article with given characteristics, for a determined set of weights. After careful assessment, it was agreed to that the article characteristics comprised two major dimensions: an article profile, and a level of rigour/relevance. It was decided not to consider articles related to pure R&D and applied research, and the definition of the article profile was focused on how fashionable the subject of the article was, and whether it used real or simulated data. The alternative levels considered for each dimension are depicted in Figure 3. For assessing the impact of using different weights chosen for analysis, two alternative sets of weights apart from the AHP weights in the previous section were used. One of these sets gives equal weight to each variable, and the other one gives equal weight to each parameter, dividing the weight of each parameter by all the variables composing it.

To simulate article classifications, it is necessary to consider a behavioural model for the respondents, with stochastic components. This behaviour, along with the article characteristics, defines the simulated answers from respondents. The behavioural model used is structured according to a two-layer approach, with each layer introducing randomness in the respondents' answers. The first layer defines the general opinion of a respondent concerning an article, and the second layer establishes the specific rating given by a respondent to the article, in each question. The logical flow of this model, illustrated for the answers of a member of a specific group of stakeholders – a citizen – is presented in Figure 4.

(Please insert Figure 4 around here)

In the absence of an objective reference for the dispersion of opinions in each group of stakeholders, it was agreed to consider different levels for the variance introduced in each layer of the behavioural model, and perform independent simulations for these different levels. The dispersion of values in the first layer, which represents the variance of the general opinion about an article in the population of each group of stakeholders, is termed “dispersion of the general opinion”, and it may be ‘Small’, ‘Medium’ or ‘Large’. The dispersion of values in the second layer, which represents the variance of ratings given by an individual in different variables concerning the same article, is termed “specific perturbation” and it may be ‘Small’ or ‘Large’.

To operationalize this behavioural model, the first step was to define, for each article profile and for each level of rigour/relevance, the average rating that the members of each stakeholder group would give to an article with such characteristics. These values were termed the ‘central classification’ of each article according to each variable, and they are defined in the -3+3 scale used in this study but, in order to gain some flexibility, it was considered that this central classification could consist of either one or two contiguous values from the scale. Table 3 presents these central classifications.

(Please insert Table 3 around here)

The central classification of an article may be different for different members of the same group. In order to take this fact into account, randomness is incorporated in the central classification of an article. This randomization is included through a probability distribution for each variable, assuming that the general opinion of a respondent can be represented as a Percentile[1] of this distribution, this percentile being the same for all variables. By randomly generating a percentile for each respondent, it is possible to simulate preliminary values for all variables used in his/her assessment of the article. Therefore, as illustrated in the top of

Figure 4, a preliminary value of each variable can be defined by the article characteristics, by a randomly generated percentile, and by the probability distribution concerning the variable. The definition of these probability distributions is presented in Table 4.

(Please insert Table 4 around here)

In the second layer of the behavioural model, specific perturbations[2] were applied to the preliminary values obtained in the first layer, in order to consider differences in perception concerning each variable. The specific perturbations are defined as probability distributions, with their values being added to the preliminary values of the variables. This is depicted in the lower half of Figure 4. In this second layer, when the average of the specific perturbations, rounded to the nearest integer, is different from zero, it is subtracted from the simulated specific perturbations, in order to ensure that the introduction of specific perturbations does not significantly change the simulated general opinion. Additionally, whenever the application of the perturbation leads to a value outside the scale (+4), the nearest value of the scale is used instead (+4 is changed to +3).

To clarify, the simulation process is exemplified by using the ratings given by a citizen (case depicted in Figure 4) for an average article of profile 1 with medium dispersion of the general opinion, and small dispersion of the specific perturbations. The general opinion is initially simulated as a percentile, and the specific perturbations as samples from a discrete distribution. For the first layer of the behavioural model, it is assumed that the simulated percentile of the general opinion is 71%. The simulated percentile of the general opinion, taken into consideration simultaneously with the article profile, dispersion of general opinion and the level of rigour/relevance, leads to preliminary classifications of +2, +1, +1 and +1 for parameters IT, PPSA, RWC and JP_{CT} , respectively (see Tables 3 and 4). For the second layer, it is assumed that the specific perturbations are +1, 0, -1 and +1 for IT, PPSA, RWC and JP_{CT} , respectively. These perturbations are added to the preliminary values of the

parameters being assessed, leading to ratings of +3, +1, 0 and +2 for IT, PPSA, RWC and JP_{CIT} , respectively.

In order to determine a classification for each article, the classification of each question is rescaled to a 0–1 range. The classification by the reviewer is naturally in a 0–1 scale, and CIT/IF was truncated at a maximum of 2 and then rescaled to a 0–1 range. After this rescaling, the classification of each article is calculated. A simulation was performed for each possible combination of an article profile, a level of rigour/relevance, a set of weights, and a scenario concerning the dispersion of the general opinion and the specific perturbation. The results are presented in the next section.

Results and discussion

The most important results of the simulation exercises are presented in Tables 5-6. For reasons of space, and since the results were very similar for large and small specific perturbations, only those results concerning large specific perturbations are shown. For above and below average articles, only results concerning AHP weights are presented.

(Please insert Table 5 around here)

(Please insert Table 6 around here)

It is possible to see that the set of weights has some influence in the average classification of each type of article, with different sets leading to slightly different average classifications. However, this influence can be described as limited, without significant changes in classification taking place. This means that some care must be taken in defining the weights to be used in the model, but some perturbation in the weights is not expected to cause significant differences in the classification of the articles.

The number assumed in the simulations (ten respondents for each group of stakeholders) seems to be enough to achieve quite small standard deviations for the classifications, even when the considered perturbations are large. These standard deviations are particularly small

when the differences in average classification for articles with different rigour/relevance are taken into account, meaning that the model can differentiate between articles with different levels of rigour/relevance (at least for the levels considered in this simulation model).

For the same level of rigour/relevance, articles with different profiles sometimes obtain quite different average results leading to a conclusion that some care must be taken when comparing articles with different profiles. Looking at the way the dispersion of general opinion influences results, it is concluded that an increase in this dispersion always increases the standard deviation of the results, as was to be expected at the outset. Summarizing, the testing, validation and operationalization of the AR2I Model through this simulation research study seems to indicate that the Index and its make-up both work, which was the ultimate purpose of this stage of research.

Conclusions and future research

Academic research has a vital role in enhancing economic performance, the quality of life, and the well-being of society. AR2I has been developed to make a contribution in respect of the measurability of the attainment of these goals. A direct implication of this analysis is the measurable effect of the value and impact of academic research on government-based organizations, funding bodies and higher education institutions. An important link to be assessed is the demonstration of the effectiveness of public money invested in scholarly research.

It is clear that only limited data is available related to the measurement of societal impact derived from academic research. AR2I will make a contribution in measuring the impact of theory-led research and its published outputs on business systems in regard to changing mindsets, applicability and future effective practice. This will not only benefit commercial organizations but also, and more importantly, the human beings who are the recipients and

users of their products and services. These impacts will reflect on society in general in terms of improving human lives.

AR2I could also play an important role in the timely and critical issue related to the emerging need for Business Schools to raise their sights beyond the myopic notion of relevance. It is acknowledged that various contextual issues surround the implementation of metricising societal impact, these being the political system, the typology of the business school, the cultural system, the academic reward system, tangible research outputs, academic rigour, and academic relevance. The proposed conceptual framework outlines a differentiated-focus strategy that could be applied by Business School deans. The legitimacy perspective is also worthy of consideration by Business School Deans as it can assist in positively influencing the alignment between the Business School and society.

Within the measurability attributes of AR2I, it was also possible to calibrate the direction and intensity of opinions by using a Stapel scale. The Analytic Hierarchy Process (AHP) as a structured technique allowed for the weights required by the AR2I to be defined. It facilitated the understanding of the research problem as well as the intricacies of group decision-making. With an analysis based on Monte Carlo simulation, it is concluded that AR2I is able to differentiate articles of varied stature in regard to academic rigour and relevance. Articles based on empirical research with real data and research studies tackling more ‘fashionable’ themes might be preferred and, therefore, rated highly. Hence, it is recommended that clusters and typologies of academic articles should be compared within the specific aggregate nucleus/band/chunk or collection of academic output sorted by theme, area of research specialization, type or research philosophy or methodology, research context or any other research classificatory domain.

The variations of opinion-making associated with the three specific typologies of individual, academics, and other stakeholders concerning the key evaluation variables for a particular

article, have proven to be of little influence and have limited impact on the results. However, it is believed that small opinion perturbations by very influential scholars, authoritative business people, and society leaders might produce a larger dispersion in the range of opinions. It has also been found that the limited sample of respondents does not lead to very disparate classifications for similar articles, and the results indicate that the AR2I model is able to differentiate between different levels of academic rigour and relevance.

In the current paper, the weights of the variables were defined using AHP, based on the pairwise judgments provided by the authors. In future applications, the assessment of the weights could be based on the opinion of several stakeholders, hopefully leading to more robust results. Finally, the validation of the representativeness achieved by the selected respondents is another avenue for future research. After a real application, alternative analyses – for example, based on bootstrapping – could be performed, allowing a more robust validation.

[1] The $x\%$ percentile of a distribution is the value below which $x\%$ of the values of that distribution can be found.

[2] While the random percentile can be seen as the ‘average’ perception of a respondent concerning the article, these specific perturbations represent the variance of ratings given by the respondent in different variables. For example, a respondent may consider that a given article is ‘median’ (50% percentile), but that it is above median in what concerns changing mindsets and below median in what concerns how it may lead to taking action.

References

- Aguinis, H., Shapiro, D. L., Antonacopoulou, E. P., & Cummings, T. G. (2014) 'Scholarly Impact: A Pluralist Conceptualization', *Academy of Management Learning & Education*, 13/4: 623–39.
- Aken, J.E.V. (2005) 'Management research as a design science: Articulating the research products of mode 2 knowledge production in management', *British Journal of Management*, 16/1: 19–36.
- Alvesson, M. & J. Sandberg (2011) 'Generating research questions through problematization', *Academy of Management Review*, 36/2: 247–71.
- Alvesson, M. & J. Sandberg (2013) 'Has management studies lost its way? Ideas for more imaginative and innovative research', *Journal of Management Studies*, 50/1: 128–50.
- Antonacopoulou, E.P., E. Dehlin & M. Zundel (2011) 'The challenge of delivering impact: Making waves through the ODC debate', *The Journal of Applied Behavioral Science*, 47/1: 33–52.
- Baldrige, D.C., S.W. Floyd & L. Maroczy (2004) 'Are managers from Mars and academicians from Venus? Towards an understanding of the relationship between academic quality and practical relevance', *Strategic Management Journal*, 25: 1063–74.
- Bartunek, J.M. & S.L. Rynes (2014) 'Academics and practitioners are alike and unlike: The paradoxes of academic-practitioner relationships', *Journal of Management*, 40/5: 1181–201.
- Bartunek, J.M., S.L. Rynes & D.R. Ireland (2006) 'What makes management research interesting and why does it matter?', *Academy of Management Journal*, 49: 9–15.

Baum, J.A.C. (2011) 'Free-riding on power laws: questioning the validity of the impact factor as a measure of research quality in organization studies', *Organization*, 18: 449–66.

Beise, M. & H. Stahl (1999) 'Public research and industrial innovations in Germany', *Research Policy*, 28/4: 397–422.

Bornmann, L. (2012a) 'What is societal impact of research and how can it be assessed? A literature survey', *Journal for the American Society for Information Science and Technology*, 64/2: 217–23.

Bornmann, L. (2012b) 'Measuring the societal impact of research', *EMBO Reports*, 13/8: 673–76.

Bornmann, L. & R. Mutz (2011) 'Further steps towards an ideal method of measuring citation performance: The avoidance of citation (ratio) averages in field-normalization', *Journal of Informetrics*, 5/1: 228–30.

Clark, T. & M. Wright (2009) 'So farewell then.... Reflections on editing the Journal of Management Studies', *Journal of Management Studies*, 46: 1–9.

Collins, D. (2000) *Management fads and buzzwords: Critical-practical perspectives*. Oxon, UK: Routledge.

Corley, K.G. & D. Gioia (2011) 'Building theory about theory building: What constitutes a theoretical contribution', *Academy of Management Review*, 36/1: 12–32.

Daft, R.L., R.W. Griffin & V. Yates (1987) 'Retrospective accounts of research factors associated with significant research outcomes', *Academy of Management Journal*, 30/4: 763–85.

Daft, R.L. & A.Y. Lewin (2008) 'Rigor and relevance in organization studies', *Organizational Science*, 19: 177–83.

Datar, S., D.A. Garvin & P.G. Cullen (2010) *Rethinking the MBA: Business Education at a crossroads*. Boston, Massachusetts: Harvard Business Press.

Davis, M.S. (1971) 'That's interesting!', *Philosophy of Social Science*, 1: 309-44.

De Jong, S., K. Barker, D. Cox, T. Sveinsdottir & P. Van den Besselaar (2014) 'Understanding societal impact Through Productive Interactions: ICT Research as a Case', *Research Evaluation*, 23: 89–102.

De Vita, G. & P. Case (2016) 'The smell of the place: Managerialist culture in contemporary UK business schools', *Culture and Organization*, 22/4: 348–64

Dong, Y., G. Zhang, W.C. Hong & Y. Xu (2010) 'Consensus models for AHP group decision making under row geometric mean prioritization method', *Decision Support Systems*, 49/3: 281–9.

Duncan, W.J. (1974) 'Transferring management theory to practice', *Academy of Management Journal*, 17/4: 724–39.

Dunn, W.N. (1980) 'The two-communities metaphor and models of knowledge use', *Knowledge Creation, Diffusion, Utilization*, 1/4: 515–36.

Dyck, B., K. Walker, F.A. Starke & K.L. Uggerslev (2011) 'Addressing concerns raised by critics of business schools by teaching multiple approaches to management', *Business and Society Review* 116/1: 1–27.

Ensthaler, L. & T. Giebe (2014) 'A Dynamic Auction for Multi-Object Procurement under a Hard Budget Constraint', *Research Policy*, 43: 179–89.

Flickinger, M., A. Tusche, T. Gruber-Muecke & M. Fiedler (2014) 'In search of rigor, relevance and legitimacy: What drives the impact of publications?', *Journal of Business Economics*, 84: 99–128.

Garfield, E. (1972) 'Citation analysis as a tool in journal evaluation', *Science*, 178: 471–9.

George, G. (2014) 'Rethinking management scholarship', *Academy of Management Journal*, 57/1: 1–6.

Globerman, S. (2011) 'The Social Responsibility of Managers: Reassessing and Integrating Diverse Perspectives', *Business and Society Review* 116/4: 509–32.

Gordon, R.A. & J.E. Howell (1959) *Higher education for business*. New York: Columbia University Press.

Grey, C. (2010) 'Organizing studies: publications, politics and polemics', *Organization Studies*, 31: 677–94.

Gulati, R. (2007) 'Tent poles, tribalism, and boundary spanning: The rigor-relevance debate in management research'. *Academy of Management Journal*, 50(4), pp.775-782.

Guthrie, J. & L.D. Parker (2014) 'The global accounting academic: what counts!', *Accounting, Auditing & Accountability*, 27/1: 2–14.

Hamel, J. (1993) *Case Study methods. Qualitative Research Methods, Vol. 32*. Thousand Oaks, CA: Sage.

Hodgkinson, G.P. & D. Rousseau (2009) 'Bridging the rigour-relevance gap in management research: it's already happening!', *Journal of Management Studies*, 46: 534–46.

Hodgkinson, G.P. & K. Starkey (2011) 'Not simply returning to the same answer over and over again', *British Journal of Management*, 22: 355–69.

Ivory, C., P. Miskell, H. Shipton, A. White, K. Moeslein, & A. Neely (2006) *The future of business schools in the UK*. London: Advanced Institute of Management Research.

Jääskeläinen, M., M. Maula & G. Murray (2007) 'Profit distribution and compensation structures in publicly and privately funded hybrid venture capital funds', *Research Policy*, 36/7: 913–29.

Kaur, J., F. Radicchi & F. Menczer (2013) 'Universality of scholarly impact metrics', *Journal of Informetrics*, 7/4: 924–32.

Kieser, A. & L. Leiner (2009) 'Why the rigour-relevance gap in management research is unbridgeable', *Journal of Management Studies*, 46: 516-33.

Mansfield, E. (1991) 'Academic research and industrial innovation', *Research Policy*, 20/1: 1–12.

Mansfield, E. (1998) 'Academic research and industrial innovation: An update of empirical findings', *Research Policy*, 26/7-8: 773–6.

Miller, D. & J. Hartwick (2002) 'Spotting management fads', *Harvard Business Review*, 80/10: 26–7.

Mingers, J. (2009) 'Measuring the research contribution of management academics using the Hirsch-Index', *Journal of the Operational Research Society*, 60/9: 1143–53.

Mingers, J. & E.A. Lipitakis (2014) 'A bibliometric comparison of the research of three UK business schools', *Proceedings of the International Multiconference of Engineers and Computer Scientists, Vol. II*, IMECS, March, 12–14, Hong Kong.

Mizzaro, S. (2003) 'Quality control in scholarly publishing: A new proposal', *Journal of the American Society for Information Science and Technology*, 54/11: 989–1005.

Moodie, G. (2006) 'Vocational education institutions role in national innovation', *Research in Post-Compulsory Education*, 11/2:131–40.

Niederkrötenhaler, T., T.E. Dormer & M. Maier (2011) 'Development of a practical tool to measure the impact of publications on the society based on focus group discussions with scientists', *BMC Public Health*, 11: 588.

Olmos-Penuela, J., E. Castro-Martinez & P. D'Este (2014) 'Knowledge transfer activities in social sciences and humanities: Explaining the interactions of research groups with non-academic agents', *Research Policy*, 43/4: 696–706.

Oswick, C., P. Fleming & G. Hanlon (2011) 'From borrowing to blending: rethinking the process of organizational theory-building', *Academy of Management Review*, 36: 318–37.

Palmer, D., B. Dick & N. Freiburger (2009) 'Rigor and relevance in organization studies', *Journal of Management Inquiry*, 18/4: 265–72.

Pedersen, E. R. (2010) 'Modelling CSR: How managers understand the responsibilities of business towards society', *Journal of Business Ethics*, 91: 155–66.

Pfeffer, J., & Fong, C. T. (2002) 'The end of business schools? Less success than meets the eye', *Academy of Management Learning & Education*, 1(1): 78–95.

REF (2014) 'REF2014 Excellence Framework', Available at <http://www.ref.ac.uk>

Accessed on 26th June 2017.

Reikard, G. (2005) 'Endogenous technical advance and the stochastic trend in output: A neoclassical approach', *Research Policy*, 34/10: 1476–90.

Scherer, A.G. & G. Palazzo (2011) 'The new political role of business in a globalised world: A review of a new perspective on CSR and its implications for the firm, governance and democracy', *Journal of Management Studies*, 48/4: 899–931.

Shrivastava, P. & I. Mitroff (1984) 'Enhancing organizational research utilization: the role of decision makers' assumptions', *Academy of Management Review*, 9/1: 18–26.

Smith, R. (2001) 'Measuring the social impact of research', *British Medical Journal*, 232: 528.

Starbuck, W.H. (2006) *The productivity of knowledge: The challenge of social science research*. Oxford University Press.

Starkey, K., A. Hatchuel & S. Tempest (2009) 'Management research and the new logics of discovery and engagement', *Journal of Management Studies*, 46: 547–58.

Sutherland, I., J. Gosling & J. Jelinek (2015) 'Aesthetics of power: why teaching about power is easier than learning for power, and what business schools could do about it', *Academy of Management Learning Education*, 14/4: 607–24.

Tranfield, D. & K. Starkey (1998) 'The nature, social organization and promotion of management research: Towards policy', *British Journal of Management*, 9: 341–53.

Tushman, M. L., C. O'Reilly, A. Fenollosa, A.M. Kleinbaum & D. McGrath (2007) 'Relevance and rigor: Executive education as a lever in shaping practice and research'. *Academy of Management Learning & Education*, 6/3: 345–62.

Weiss, C.H. & M. Bucuvalas (1977) 'The challenge of social research to decision making'. In: H.W. Carol (ed), *Using Social Research in Public Policy Making*, pp. 213–34. Lexington, MA: Lexington Books.

Wiek, A., S. Talwar, M. O'Shea & J. Robinson (2014) 'Toward a methodological scheme for capturing societal effects of participatory sustainability research', *Research Evaluation*, 23(2):117–32.

Willmott, H. (2012) 'Reframing relevance as "social usefulness": A comment on Hodgkinson and Starkeys "not simply returning to the same answer over and over again"', *British Journal of Management*, 23: 598–604.

Wood, G. & P. Budhwar (2014) 'Advancing theory and research', *British Journal of Management*, 25/1: 1–3.

Woodside, A.G. (2009) 'Journal and author impact metrics: an editorial', *Journal of Business Research*, 62/1: 1–4.