

## Title

“Oh it was big, I don’t know how big but it was big”

Characteristics of metrics that influence researcher adoption

## Authors

Gemma Elizabeth Derrick, Instituto de Políticas y Bienes Públicos, Consejo Superior de Investigaciones Científicas, C/Albasanz 26-28, 28037Madrid, Spain.

E-mail: gemma.derrick@csic.es

James Gillespie, Menzies School for Health Policy, The University of Sydney, Sydney, NSW, Australia, 2006.

E-mail: james.gillespie@sydney.edu.au

## Abstract

This article analyses why researchers decide to adopt Research Evaluation Metrics (REMs) as part of their role and also investigate other factors that lead them to adopt some REMs and not others. This research adapts the approach of Rogers (1983) and combines qualitative (interview) with quantitative (survey) methodological approaches to analyse two university departments regarding individual researcher adoption, use and attitudes of REMs.

A wide range of REMs were shown to have been adopted and used in various ways to promote individual researchers and their research. A peer-based understanding and faith in the Journal Impact Factor and h-index was identified. This fuelled bottom-up organisation change but also facilitated the propagation of myths about REM-shortcomings and disadvantages. However, a majority of researchers who had already adopted REMs demonstrated an understanding of REM shortcomings and continued to use them as a yardstick pertaining to the quality of a researcher. The results show that metrics are perceived as beneficial in their simplicity and ease of understanding but they are still far from being widely accepted as a replacement for peer review (relative advantage, Rogers (1983)). Researchers were also more predisposed to adoption of REMs if the REMs of researchers, articles and journals reflected their already existing opinions of the value of these actors.

This research represents one of the first, in-depth, mixed methods analyses of the current use and -opinions of primary-users towards REM use.

## Keywords

Research evaluation, diffusion of innovations, research policy.

## **1.0 Introduction**

The social value of all research is the extent with which it has been adopted by the community. As such, for research evaluation metrics (REMs), the value of any metric indicator produced by the scientometric research community, lies in its utility. If it is not used by the community it serves, then there is little value in the investment used to create it. An REM will continually fail to offer value to the research community as long as it does not reflect what is valued in research evaluation for that community. Therefore, a measure of excellence for an REM is the extent with which it is disseminated and understood by the research community.

This research acknowledges that there are different categories of REM users. Primary users have already been acknowledged in the literature and include research managers, governments and specialist REM researchers (Jonkers & Derrick, 2011). However, the way that these users learn and apply REMs is different to secondary REM users such as researchers. Indeed the attitudes, understanding and opinions of secondary REM users will affect the calculation and provision of these REMs for evaluation by these secondary users. It is therefore important to investigate the behaviours and attitudes that are influential for researchers when deciding whether to adopt REMs.

This research investigates why researchers decide to adopt REMs as part of their role and also investigates other factors that lead them to decide to adopt specific REMs but not others. This research adapts the approach of Rogers (1983) by combining qualitative (interviews) with quantitative (survey) methodological approaches to analyse the current situation within two university departments regarding the adoption, use and attitudes towards REMs. This research represents one of the first, in-depth analyses of the current use of REMs and the opinions of secondary-users towards their use.

## **2.0 Methodology**

This research adopts a mixed-methods approach combining Likert scale surveys with in-depth semi-structure interviews. The components of each methodological approach and how they were combined, are described below.

## 2.1 Participants

*The participants were researchers with at least 3 years post-PhD from the Departments of Physics and Medicine at the University of Sydney, Australia. The departments of Physics and Medicine were selected to participate as both fields are well represented by current REMs. It is therefore valid to assume that the researchers within these departments would have had more exposure, discussion and time to experiment with various REMs. Permission from each Departmental Head was received prior to data collection took place. This permission also increased the legitimacy of the research project and was used to increase response rates.*

## 2.2 Mixed methods design

*The research evaluation field has traditionally used quantitative methods to examine the visibility, use, diffusion and influence of journal publications. Despite their prominence, quantitative methods are not as well suited to measure other complex aspects of the diffusion of research such as the motivations of the adopter, organisational change, research leadership and the process of implementing evaluation guidelines, which are also critical issues in research evaluation management. These more nuanced aspects of the value of REMs and their diffusion within research communities require additional non-traditional methods. These include the application of qualitative alongside quantitative methods within a mixed methods framework.*

## 2.3 Survey

*The survey collected a combination of basic demographic information of participating researchers, information about current use of REMs, as well as a 5-point Likert scale survey of researchers' opinions of REM use.*

*The questions were based around a number of factors influencing adoption as specified by Rogers (1983) such as Relative Advantage; Compatibility; Complexity; Trialability; Observability; Ease of Use; and Social Approval. Questions were divided into five separate parts asking questions regarding different aspects of REM adoption including: Characteristics of adoption; My use of REMs; My attitudes towards REMs; My research and REMs; and My REMs and my peers.*

*A total of 200 researchers completed the survey with a further 49 agreeing to take part in a follow-up, semi-structured interview (described below). Surveys were analysed using frequency analysis using STATA and were cross referenced with the themes identified in the interview analysis also described below.*

## **2.4 Interviews**

*A total of 26 researchers took part in a follow up semi-structured interview about their use of nominated REMs and opinions about the current use of REMs during evaluation. During the survey, researchers were asked to nominate the REMs they currently use and describe where and why they used them. The interview questions and related prompts were modelled around the characteristics influencing adoption (Relative advantage, Compatibility, Complexity, Trialability and Observability); the process of adoption (Knowledge, Persuasion, Decision, Implementation and Confirmation) as described by Rogers (1983) as well as the possible causes for non-adoption of other metrics. Finally, questions exploring the role of organisational and/or peer influence in the decision to adopt (or not) were also included.*

### **2.4.1 Interview analysis**

*Interviews were fully transcribed and analysed using Atlas-ti. Interview analysis was based on a semi-structured, cognitive-based, grounded theory method that involved two separate rounds of coding. The first round of coding included the identification of responses in line with Rogers (1983) five main characteristics of adoption (described above) as well as other characteristics identified in previous diffusion research (Ease of Use, Status/Social approval, Image, Usefulness, Divisibility, Voluntariness and Communicability). Once this round of coding was complete, in-vivo codes were further identified within and between first round codes in order to explore themes and interlock related themes. Coding used but was not restricted to the following domains as a guide (Table 1).*

*Throughout data analysis, a process of constant comparison was employed, with codes being continually refined, enriched and reorganised. Enough material was analysed to ensure that categories were saturated, that is, that all codes appeared to fit under on or more existing categories and all concepts were fully described and well-understood. Our categories and concepts were then organised under a number of general headings that form the structure of the results section below.*

**Table 1: Coding domains and descriptions for 2nd round interview analysis**

<b>Domain</b>	<b>Description</b>
<b>Strategies/methods/use or REMs and the professional behaviours that researchers use REMs.</b>	Ways of doing things; Communicative practices and the strategies that are central to these.
<b>Judgements/values/meanings/beliefs/discourses about REMs and research evaluation.</b>	Ways of assessing and representing research excellence, ideological/paradigmatic constructions, contested viewpoints about what is valued, legitimisations, value judgements about use and researchers who use REMs. Judgements about benefits/limitations of peer review and of REMs.
<b>Interactions/Relationships</b>	Networks, collaborations, supports, adversaries. Includes influences, hierarchies, leverage, inequalities, vested interests. Conflicts/battles eg. Between researchers, evaluators and organisations.
<b>Cultural Practices</b>	This includes social processes, rules, norms and conventions, expectations.

### **3.0 Results**

#### **3.1 Adoption and current use of REMs**

*Basic information was collected about the main REMs adopted by researchers and where these are currently being used. The results are shown in Table 2. The Journal Impact Factor, despite widespread discussion regarding its misgivings and inadequacy as a quality measure (Derrick et al submitted, 2013), was still a popular measure used by researchers. Not surprisingly, researcher used this metric about all other REMs when deciding where to publish (99.0%), a consideration reflective of the reward-maximising model of journal submission (Luukkonen, 1992). This model outlines how researchers make a hierarchical list of potential journals and submit their article to each journal on the list, starting from the top, until publication acceptance is achieved.*

*The results below suggest that the JIF is a major consideration when constructing this hypothetical list. The JIF was also shown to be used in a number of different professional situations including for competitive purposes such as part of grant applications (80.0%) and promotion applications (79.7%), as well as for self-promotion activities such as on their Curriculum Vitae (71.1%) and media purposes (81.3%). Interestingly, researchers also indicated that they used the JIF to represent themselves in Research Evaluation*

Frameworks (90.9%) despite clear government policies stating that the JIF should not be used for these evaluation purposes (NHMRC, April, 2010). Other REMs used widely included the H-index and Citations per paper. These REMs were used widely in a number of competitive and self-promotion purposes. In addition, a smaller number of alternative REMs were also indicated, however, the three metrics: JIF; Citations per paper; and the H-index, were shown to constitute the majority of REMs currently adopted by these researchers.

**Table 2: Types of REMs adopted and current use by researchers (%)**

	<b>Journal Impact Factor (JIF)</b>	<b>Citations per paper</b>	<b>H-index</b>	<b>Other</b>
Curriculum vitae	71.1	66.7	72.2	8.8
Grant applications	80.0	73.3	63.8	10.6
Promotion applications	79.7	82.3	73.4	11.4
Position reviews	78.9	64.9	66.7	7.1
Media purposes	81.3	12.5	18.8	25.1
Biography information	60.5	52.6	55.3	7.9
Professional webpage	61.5	42.3	61.5	11.5
Deciding where to publish	99.0	5.7	2.9	2.0
Research Evaluation frameworks	90.9	60.6	51.5	3.0
Evaluating other researchers	68.9	65.6	65.6	7.8
Evaluating research quality	75.4	72.3	41.5	15.3

## **3.2 Interview results**

### **3.2.1 Peer-generated adoption of REMs**

*The interview results also showed that researchers have voluntarily adopted a number of REMs and use them in a variety of ways including on their C.Vs and grant applications, despite the absence of a formal requirement to provide them. Researchers independently introduced themselves to a variety of new metrics and adopting those metrics that were perceived as: vital for successful grant applications (58.7%); or increased the success of promotion applications (62.0%). These metrics were felt to provide a simple representation of complex information about a researcher's experience or, in other words, their "value" as a researcher.*

*"...when I want to sell myself, you know, when I have written promotion applications in the past, on grant applications, you try to make it clear upfront what my value is...."*

*Relative to the belief that "everybody uses metrics" was the widespread belief that it was essential to include REMs in grant and promotion applications. Indeed, researchers increasingly used REMs for these purposes and researchers expressed their belief that providing metrics was a key factor in guaranteeing success of an application. The survey results indicated that 62% of researchers agreed or strongly agreed with the statement "I include REMs in funding or promotion applications in order to increase my chances of success." This belief that REMs were widely used by the majority of researchers, only served to accelerate REM adoption among research peers.*

*"It really latched on and then people were really – continued to use it as a benchmark. And it kind of surprised me to be honest. And it's really because everybody does it, that I do it. I mean, I think that's the critical thing is this sort of widespread acceptance of something as a benchmark."*

*This bottom up, peer-generated pressure to include REMs on applications was further fuelled by reports from researchers who had also acted as evaluators. These researchers reflected how applications that included REMs, by the merit of REM simplicity and supposed, objectivity, were viewed more favourably during evaluations.*

*"...if they put – how many papers they published, how many citations for each of those papers, what their h-factors are [then] I feel more positively predisposed to them because they provided me with information in a quick, easy to understand format that I can, that I can get at..."*

*In addition to including REMs, to make the task of the evaluators easier and include REMs to be comparable to their competitors, was the risk researchers felt when not including them. This was felt to be the equivalent of potentially hiding information from evaluations and further emphasised the perceived necessity in including REMs on grant and promotion applications,*

“If you haven’t got it and you’re not putting it in there, and I just think ‘what have you haven’t got?’ and therefore you are going to get marked down.”

*Indeed, a large amount of peer influence that forced researchers to use REMs on grant and promotion applications was related to the belief that providing metrics was an accepted norm and that by not providing REMs, their absence sad more about the researcher’s value than when a researcher provided them.*

“I put it there [REMs on grant applications] because, you know, somehow if you don’t put it there then it might make people wonder, ‘why didn’t you put it there’.”

*This was exacerbated by researchers stating that, when they acted as evaluators, they specifically searched for metrics on applications to guide their evaluations. The reasons stated for this behaviour ranged from time restrictions, to a need for a “yardstick” to measure and compare researcher value.*

“...if they put how many papers they published, how many citations for each of those papers, what are their h-factors. I instantly feel more positively predisposed to them because they provided me with information in a quick, easy to understand format that I can – that I can get at, that I think has – personally, I think has some validity.”

*The same applies to specific metrics, such as the h-index;*

“...it is a good metric, because it gives you a bit of a yardstick to be able to judge people on in terms of their track record and so on....”

*In some circumstances, researchers remarked how that in the absence of a provided metric in applications, they would calculate the REMs they favoured for that application. This practice is alarming and further research is needed to determine how widespread this practice is among research evaluators. A more detailed exploration of the characteristics that appealed to researchers is outlined below in regards to the relative advantage of REMs.*

### **3.2.2 The contentious relative advantage of REMs**

*In terms of metrics, the relative advantage of REMs refers to the advantage the new innovation (REMs) has over the old innovation (peer review). Specifically, whether there was an agreement that REMs could replace peer review. Perhaps not surprisingly, of all of the potential identified characteristics, the perceived relative advantage of REMs was the most contentious. The results showed an incomplete resolution as to whether REMs presented a complete advantage over the more traditional, peer review.*

*REMs were favoured by researchers because they were perceived as simple, objective measures that provided beneficial summary information about a researcher and/or research. When discussing the benefit of simplicity of REMs, researchers described the*

*benefit of REMs of providing a quick “approximation” of a researcher’s professional background.*

“And it’s not a perfect metric by any stretch but it’s a simple one. I mean, that gives you sort of a first approximation I suppose, of somebody’s track record.”

*No distinction between REMs that were used to describe research as separate from researchers. Instead, researchers referred to how it didn’t matter what the REM meant, but if in the case of rankings, what mattered was the position of the researcher and/or research and how they looked relative to their colleagues.*

“Well trying to be honest, [it] doesn’t really, you know, whether you’re number one or whether you are number twenty, if you are up there with that, you know big US researchers and institutions, that’s only where they need to go....”

*REMs were also perceived to be “objective” measures of research value. The definition of objective in this regard does not include a measurement that is free from tampering or misuse, rather REM appearance as quantitative measures and the ability of researchers and evaluators to rank according to these numbers, appealed to researchers as “objective”. In addition, compared to peer-review, a naturally subjective evaluation tool, REMs were perceived as free from the biases influencing peer-review generated evaluations.*

“...for me again, the quantitative background, I guess they impressed me. You know, they’re trying to do something here in a rather more objective way is a bit less of, you know, personal view of that and less subjective....”

*And;*

“It’s a good, quantitative piece of evidence that you can include as to your track record.”

*Despite a number of REM characteristic favoured by researchers, there were a number of characteristics that made researchers cautious about their overuse and the potential loss of peer review. These included, especially for medical researchers, an understanding that REMs did not reflect the wider importance of research or their activities as a researcher.*

“So these metrics are only useful to a certain extent and they’re kind of a broad brush stroke.”

*However, researchers also gravitated away from using specific REMs and towards others, because of the way that the alternative metrics represented them. Researchers, though acknowledging that REMs represented a competitive advantage in the research “game”;*

“....and it’s a game where everybody wants to look good, right?”

*And understanding that there was a lack of guidelines about what REMs are to be used in grant and promotion applications, would automatically assume metrics that made them look better and then promote the benefits of these alternative metrics to their peers. The*

*survey results also confirmed with this, 46.1% of researchers agreeing or strongly agreeing with the statement “I would not use an REM if it represented me badly.”*

*“...the person that you often talk to want you to use a different index which they look better in, right?”*

*Researchers were cautious about this, less the new metric represent them poorly, but this was balanced with a curiosity-driven experimentation with new REMs.*

*“...we used it just kind of for fun, to speculate and just to play around with different metrics just to, you know, just for fun basically.”*

*However, this was not widespread, with only 27.4% of researchers indicating that they had “Experimented with different metrics”. In addition, 63.5% agreed or strongly agreed with the statement “REMs encourage researchers to ‘game’ or ‘cheat’ the system”.*

*Out of all the characteristics of REMs that appealed to researchers, the strongest characteristics was when the researcher’s existing impressions of excellent research was reinforced by corresponding excellent metrics. In these cases, researchers reflected on situations where they had come across articles or researchers that they value and had considered their REMs as equivalent to their personal appreciation of that researcher and their research. However, this was the same case with researcher’s own opinions about their research contributions.*

*“...people are, the average person is prepared to sort of just accept the number that’s given, especially if it is higher.”*

*If researchers felt their work was adequately represented by metrics, they felt more predisposed towards the widespread use and relative advantage of REMs. However, the agreement between a researcher’s personal opinion of research excellence and the corresponding metrics did not always occur. A total of 42.8% of researchers agreed with the statement, “I am sceptical about REMs”.*

*“I’ve read some papers which were really useful to me and, of course, I have cited them and I’m surprised to find that they’re not cited very often.”*

*This was for a number of reasons, including a broader definition of research excellence beyond the limitations of REMs;*

*“...I suppose I’ve got my own mental metric that I use to assess whether research is good or not...”*

*Also included was an understanding of the restrictions of REMs as measures of research excellence,*

*“...because it [REM] is not actually an indicator of what you do.”*

*In fact, the corresponding survey results show that only 7.1% of researchers agreed or strongly agreed that “The REMs I use now are better than using peer review to evaluate research.”*

*In these cases, we expected that this mismatch would result in an increased scepticism in REMs and a corresponding decrease in their use by these researchers, but this was not the case. On the contrary, the widespread belief that REMs were essential for success seemingly reinforced researcher use of REMs. However, this mismatch between the personal definition of research value and REMs, was identified as a major factor limiting the perception of the relative advantage of REMs over peer review (old innovation), and therefore affecting their complete adoption.*

*Finally, REMs were perceived by researchers as “....a number you just can’t get away from”, its increased and unregulated use in research evaluations make it a powerful tool with which researchers can engage. Specifically, the motivation behind this engagement is the perception of REMs as a marketing tool rather than as a benchmarking/evaluation tool. In relation to this view, researchers preferred metrics that they could “control” or “talk around”. These tended to be individual metrics such as the h-index rather than fixed-metrics like the journal impact factor, or institutional level metrics such as rankings.*

#### **4.0 Discussion**

*These results have important implications for how universities, their departments and research offices, as well as governments evaluate research in official evaluation exercises. Researchers are yet to be convinced to the complete relative advantage of REMs, a factor that limits more widespread adoption (Rogers, 1983). Researchers were aware of the advantages of REMs and were drawn to their use because of their ability to communicate complex information simply and “objectively”. Despite researcher awareness of REM disadvantages, the belief that REMs are essential for grant research success, combined with the belief of their widespread use between competitors (colleagues), by peers who also act as evaluators, and the preference expressed by these evaluators about the REMs use, fuels REM adoption.*

*Current research use of REMs should illustrate concerns over the perceived objectivity of REMs. Not all provided metrics are free from tampering. Researchers aim to achieve as high a number as possible as part of a personal marketing campaign. Therefore, evaluators on the institutional and government levels should treat such metrics with caution as they are not as “objective” as they are perceived. Instead of promoting metrics as a tool, researchers should engage with, organisations should focus on providing a system for independent calculation including guidelines for the calculation including specified time periods, databases and types of publications.*

#### **Acknowledgements**

*I would like to acknowledge the Department of Physics and The Sydney Medical School at the University of Sydney for supporting me while I conducted this research.*

#### **References**

- NHMRC (April, 2010) “NHMRC removes Journal Impact Factors from peer review of individual research grant and fellowship applications.” NHMRC Media release. Available from [http://www.nhmrc.gov.au/\\_files\\_nhmrc/file/grants/peer/impact%20factors%20in%20peer%20review.pdf](http://www.nhmrc.gov.au/_files_nhmrc/file/grants/peer/impact%20factors%20in%20peer%20review.pdf).*
- Luukkonen, T. (1992) “Is Scientist Publishing Behaviour Reward-Seeking?” *Scientometrics*, 24(2): 297-319.*
- Rogers, E.M. (1983) “Diffusion of Innovations”. Free Press: New York.*