

Impact case study (REF3)

Institution: Coventry University		
Unit of Assessment: 10		
Title of case study: Counting what Counts Confidently (CCC)		
Period when the underpinning research was undertaken: 2010-ongoing		
Details of staff conducting the underpinning research from the submitting unit:		
Name(s):	Role(s) (e.g. job title):	Period(s) employed by submitting HEI:
Ralph Kenna	Professor	Since 2002
Olesya Mryglod	Invited Researcher	Since 2019
Yurij Holovatch	Visiting Professor	Since 2018
Rob Low	Reader	1986-31.01.2021
Period when the claimed impact occurred: From 1 Aug 2013		
Is this case study continued from a case study submitted in 2014? No		

1. Summary of the impact (indicative maximum 100 words)

There has been direct and indirect impact on policies, practices, guidelines, strategies, reviews, and documents pertaining to the REF itself and counterpart frameworks internationally. E.g., direct impacts include on the Metrics Tide review of metrics in research assessment [S1] that informed the development of the UK's REF [S4,S5]. Similar impact in Finland [S2], Norway [S3], Ukraine [S6], and other countries was on the processes themselves and/or on reports that helped frame them (e.g., by RAND [S2b], OECD [S3a] and other organisations). All universities, units, and researchers assessed by such processes, and assessors who assessed them – including you, the reader of this document - were indirectly impacted [S1,S4,S5]. Other quality-measuring organisations such as CAMRA [S6] and Ofqual [S7] were directly impacted through applying the outcomes of the research to their analysis methods.

2. Underpinning research (indicative maximum 500 words)

In 2010, Kenna and B.Berche (Lorraine) developed a mean-field model that explained patterns they observed in RAE2008 results and its French counterpart [R1a]. Research quality in 30 disciplines increases linearly with **unit size** up to saturation points which reflect limits on amounts of meaningful interactions an academic sustains on average. This quantified the notion of **critical mass** in research and other cooperative endeavours. Having been ranked “Best of 2010” in EPL and been discussed by Europhysics News, Times Higher Education and other media outlets, these findings caught (and continue to catch) the interest of policy makers, leaders, managers, etc. worldwide, leading to new research collaborations, including with Y.Holovatch and O.Mryglod (Lviv) [R2-R5], R.Low and R.S.MacKay (Warwick) [R6].

At the start of the current REF cycle, there were “powerful currents whipping up the metric tide” [S1]; replacing peer review by metrics would have reduced the cost of REF2021 and moved evaluation away from academics. Using citation data from seven disciplines supplied by Thomson Reuters and Scopus, Mryglod, Berche, Holovatch and Kenna analysed correlations between peer-review RAE2008 scores and different scientometric indicators [R2,R3]. Results showed correlations are good only for large teams (**size defined by critical masses** identified in [R1]) and only for aggregated measures of unit strength (research power). Correlations were poor for small and medium teams and universally poor for per-head measures of quality (as opposed to power). These results suggest that while citation counts might possibly be used to replicate QR funding for large groups, they **do not correlate with peer review by REF panels for small and medium groups and could not replicate REF rankings of quality for any size categories**.

Despite having established that metrics cannot replicate peer review, the team went on to use metrics to “predict” rankings for four UoAs in REF2014 and published these predictions in [R4] before REF results were released. They revisited these forecasts afterwards in [R5]. “Predictions failed to anticipate with any accuracy either overall REF outcomes or movements of individual institutions in the rankings relative to their positions in the previous RAE.” Conclusions were that “an extended discussion on the role of metrics in national research assessment exercises of the types considered here is warranted” [R5].

As a direct consequence of this research, Kenna, Holovatch and Mryglod were invited as advisors on national and international levels and into new research collaborations [R6]. The algorithm

Impact case study (REF3)

developed in [R6] was inspired by MacKay's experience as a panel member for mathematical sciences at RAE2001 and RAE2008. Called Calibrate with Confidence (CWC), it estimates assessor stringencies and calibrates assessee scores from raw scores and uncertainties when assessment panels comprise assessors with variable stringencies and confidence levels who can't all assess every assessee [R6]. Kenna conceived and led [R1-5]. Mryglod and Holovatch contributed mostly while visiting Coventry from Lviv as part of the L4 programme (see Environment). MacKay invented CWC and led on [R6]. Kenna and Low helped develop and trial the algorithm, interpreted results and, with MacKay, wrote [R6].

3. References to the research (indicative maximum of six references)

[R1] (a) Kenna, R., & Berche, B. (2010) 'The extensive nature of group quality'. EPL (90) 58002; (b) (2011) 'Critical mass and the dependency of research quality on group size'. Scientometrics (86) 527–540; (c) (2011) 'Critical masses for academic research groups and consequences for higher education research policy and management'. OECD Higher Education Management & Policy (23) 9-29; (d) (2012) 'Managing research quality: critical mass and optimal academic research group size'. IMA J. Manag. Math. (23) 195–207

[R2] Mryglod, O., Kenna, R., Holovatch, Yu., & Berche, B. (2013) 'Absolute and specific measures of research group excellence'. Scientometrics (95), 115-127

[R3] Mryglod, O., Kenna, R., Holovatch, Yu., & Berche, B. (2013) 'Comparison of citation-based indicators and peer review for absolute and specific measures of research-group excellence'. Scientometrics (97) 767-777

[R4] Mryglod, O., Kenna, R., Holovatch, Yu., & Berche, B. (2015) 'Predicting Results of the Research Excellence Framework using departmental h-Index'. Scientometrics (102) 2165-2180

[R5] Mryglod, O., Kenna, R., Holovatch, Yu., & Berche, B. (2015) 'Predicting Results of the Research Excellence Framework using departmental h-Index – Revisited'. Scientometrics (104) 1013-1017

[R6] MacKay, R.S., Parker S., Low R., & Kenna R. (2017) 'Calibration with confidence: a principled method for panel assessment', R. Soc. Open Sci. (4) 160760

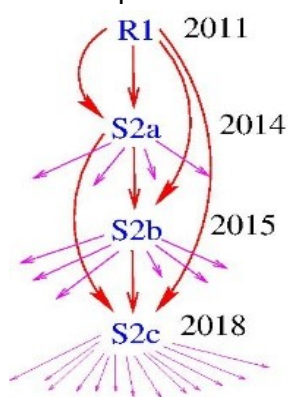
4. Details of the impact (indicative maximum 750 words)

The research changed beneficiaries' usage of **group size** and **critical mass** in research policy, management, and evaluation globally. It also stymied attempts to replace peer review by metrics globally; in essence, it has defined the operation of the current REF as an academic peer review exercise, not one based on metrics alone.

E.g., consultancy firm Economic Insight Ltd were "asked by the UK's Department for Business, Innovation & Skills (BIS) to undertake a study [S2a] to help develop its understanding of the drivers of research excellence in UK institutions". Sentiment analysis indicates [R1b] was most significant of the six papers on which the discussions on **group size** (Section 3.4 "Collaborating with Others") rested; [R1b] was the basis for 15% of the word count in the Evidence section (Sec. 3.4.2) and [R1b] drew the largest word count of the associated part of Annex A. The significance of this impact on enriching, influencing, and informing Economic Insight's service as well as BIS's understanding and awareness is evidenced by **group size** being core to one of the three findings in the 101-page report. Impact via the report is that it was "used to help BIS decide how it should best deliver its [national] responsibility for maintaining and building on current levels of excellence in research" [S2a]. HEFCE commissioned RAND Europe to analyse "the characteristics of high-performing research units within UK higher education institutions" [S2b]. Ch.2 "primarily focused on **departmental size** [and] **critical mass**." In the opening 221-word paragraph, two of the five academic sources cited are [R1b,c] each of which is referenced thrice, the remaining sources referred to once only. [R1c] also forms half the academic base of the 209-word paragraphs on "**Department Size**". Altogether, six of the 39 references to the literature in Ch.2 are to [R1] and two of the 23 journal citations in the entire 85-page report are to [R1]. The significance of this impact on enriching, influencing, and informing RAND's product, service, understanding and awareness is evidenced by this body of research forming the basis of one of the report's eight main findings, namely that on the importance of collaboration as opposed to individual strength. Beside bringing [R1] to BIS via [S2a] and to HEFCE via [S2b] the direct impact of [R1] on [S2c] was reinforced by indirect impact through [S2a,b]. [S2c] is the University of Jyväskylä's (JYU) 2018 Research Evaluation Report (JYU has 1,400 research personnel). [S2c] cites [S2a,b] five times, both of which cite [R1]. They also cite [R1b,d] directly and 20% (133 words) of the "overview"

Impact case study (REF3)

which is Ch.2 (Theoretical Considerations in Developing the Research Environment, 661 words) are about [R1]. Indeed, all discussion of “**group size**” are in this section of the report. Similarly, the double usage of the term “**critical mass**” in the discussions of [R1] is 29% of the term’s usage in the report. Impact on JYU’s strategies and practices are evidenced by [S2c] identifying “**group size**” as one of five factors “of great importance in enhancing research quality.” It is also evidenced by unit-specific conclusions; e.g., for Mathematics and Statistics the paragraph “Future Plans” states “the aim is to recruit so that the **critical mass** needed to support an international profile is retained” while the “research and publication strategy” for an off-campus extension of JYU in collaboration, with the Universities of Oulu and Vaasa, “which has a small pool of research-oriented professors” is “to focus research on one or two selected areas to ensure **critical mass**.”



The above example of simultaneous, reinforced direct and indirect impact is depicted in the figure opposite where curved arrows represent direct citations and the splayed arrows represent indirect impact emanating from [S2]. E.g., the declared purpose of BIS’s report was to “to provide research managers and funders with an overview of strategic approaches to delivering excellent research” [S2a]. HEFCE commissioned [S2b] to “be of interest to anyone involved in managing and funding research, facilitating high performance in research and, more broadly, those in the higher education sector [and] to provide research managers and funders with an overview of strategic approaches to delivering excellent research.” Likewise the JYU report “not only supports internal development but is also important for communication with the large

network of the University’s stakeholders” [S2c]. In each case independent, quantitative sentiment measures indicate [R1] as the academic work with highest direct and indirect impact on the role of **group size and critical mass** on these and other commissioned documents and their beneficiaries.

There are many similar channels of policy impact internationally. The 128-page OECD Economic Survey of Norway examined economic developments, policies, and prospects [S3a]. It has two chapters, one of which (44 pages) covers higher education. [R1b] and [R1c] are the most referenced of the nine academic papers on which that chapter rests. [R1b] and [R1c] alone form the full academic basis of the 199-word paragraph that discusses **critical mass**. The importance of critical mass for the report is evidenced in its featuring in the 200-word abstract. OECD working papers “are published to stimulate discussion”. Indeed, the issue of **critical mass** is also addressed in the 2018 Working Paper *Does size matter?* published by the Nordic Institute for Studies in Innovation, Research and Education (NIFU) [S3b]. The conclusion that absence of “**size effects**” at the level of 210 university departments and institutions (17,117 work years R&D FTEs) did not apply to research groups “that are the functional units of science” was on the basis of a single academic paper, namely [R1b] which was quoted as giving “clear evidence of **critical mass**.” The significance of this impact lies with NIFU’s responsibility for “collection, processing, interpretation and dissemination of national R&D statistics and indicators for the overall Norwegian R&D system [and] development of new indicators for the purpose of designing research and innovation policy in Norway as well as internationally” [S3b].

In 2014, Times Higher Education (THE) (28,000 copies per issue, 60,000 readers per week and over 650,000 online users globally each month) used the headline “The (predicted) results for the 2014 REF are in: Research team hopes that predictions will help to clarify **the value of metrics in assessment**” for their 470-word report [S4a] on [R4]. The 11 online comments (1,500 words) included by high-profile opponents of REF and advocates of metrics. The comment “Brave of you to publish your predictions ...It’s good to get the debate started early” evidences causality of [R4] on the debate on the usage of metrics to assess research quality. A simultaneous report of [R4] in Chemistry World (CW) (sent monthly to 50,000 members of the Royal Society of Chemistry) headlined “**Can research quality be predicted by metrics**” for their 400-word item [S5a]. In 2015 THE used the headline “Hit and miss metrics: ‘Throw of dice would give more accurate REF prediction” in [S4b] and CW used the headline “**Metrics failed to predict REF outcomes**” in [S5c] (see also [S5b]) to report [R5]. In 2015, headlined “Can the research excellence framework run on metrics?” [S4c] THE report it had asked Elsevier, owner of Scopus and data provider for World

Impact case study (REF3)

University Rankings, “to carry out its own analysis of correlations between citations data and quality scores in the REF.” Arguments by Elsevier in favour of metrics were supported by a high profile advocates they were “a good predictor”. [R4,R5] **stood alone** in the 1,400-word report as academic evidence of “another nail in the coffin for the idea of replacing REF by metrics” [S4c].

The debate impacted on ‘The Metric Tide’ (MT, 2015), a Government commissioned Independent Review of the Role of Metrics in Research Assessment [S1]. “Given your extensive work on this whole area” wrote Director of the Higher Education Policy Institute, “we would be delighted to offer you a free ticket [to] ‘Reflections on REF2014 – Where next?’”, conference at the Royal Society in 2015 where the report’s Chair presented. Emails record direct impact as do citations and reviews [S1]. Section 9.1 of MT (seven pages) concerns “Quantitative data and the assessment of research outputs”. The three non-HEFCE works cited include a blog “in favour of using the departmental h-index as an alternative to peer review evaluation for the allocation of research funding” and the Elsevier analysis [S4c]. As the only academic papers in this section, [R4,5] **stand alone** as concluding “that the relationship is **not strong enough** to justify the use of the departmental h-index as a replacement for peer review.” The significance of this is evidenced by one of the 10 “**Headline Findings**” of [S1a] **resting on** (quoting directly from) Sec.9.1 and **agreeing fully with** [R4,5] that metrics “cannot provide a like-for-like replacement for REF peer review”.

Impact on and through MT is backed by impact on and via Supplementary Report I [S1b]. Of eight “studies of correlating indicators and outcomes of peer review” in Ch.5 of [S1b], [R3] is the most extensive (covering most disciplines) and up-to-date analysis. Of the 12 bodies of work that impact the section of [S1b] titled “Predicting the outcomes of the UK RAE and REF by bibliometrics”, [R3-R5] form the basis for the longest discussion, longest quotations, most citations with over one page devoted entirely to [R3-R5]. Indeed, the section culminates quoting [R4]: “Clearly, however, overreliance on a single metric by persons who are not subject experts could be **misleading**, especially in increasingly managed landscapes in which academic traditions are diminished or eroded”. The impact of [R2-5] on MT and the impact of MT on REF was recently independently reviewed in Springer-Nature’s <https://doi.org/10.1057/s41599-019-0233-x> (4522 downloads, altmetrics score 94). Like the evidence provided above, it positions [R2-5] alongside MT as the **primary academic work** against “using metrics instead of peer review” with the above-mentioned blog and Elsevier report in favour. To quote: “The REF 2014 results were also analysed by Mryglod et al., 2015 [who] found that the departmental h-index was not sufficiently predictive, even though an earlier analysis suggested that the h-index might be predictive in Psychology [and] an analysis by Elsevier found that metrics were reasonably predictive of peer review outcomes”. Thus [R2-5] are **the independent academic basis** for MT’s conclusion “that citations should only supplement, rather than supplant, peer review” <https://doi.org/10.1057/s41599-019-0233-x>. Significance of the impact on MT is further supported by written confirmation by the Review Chair [08.07 2015] that Mryglod et al.’s research “**enriched our deliberations and provoked us to ask questions** that might otherwise remain unaddressed” [S1c].

Impact via Metrics Tide is evidenced by its 533 citations (google scholar), many of which are in policy documents. One of these is the Stern Report (July 2016) which “recognise[s] the findings of The Independent Review of the role of metrics in research assessment that it is not currently feasible to assess research outputs in the REF using quantitative indicators alone”. Echoing THE’s “nail in the coffin” comment above [S4c], the Editor of Research Fortnight, (19.07.2016) wrote the halting of “A metrics-heavy REF [would be] a victory for the team behind The Metric Tide report” and indeed, the REF is in line with recommendations of the Stern Report (which itself has 193 citations on Google Scholar).

Thus, while MT rests on 70 academic works, sentiment analysis and independent review, supported by email, point to [R2-5] as most impactful on it. [R2-5] therefore impacted directly and indirectly on non-usage of metrics in REF nationally, on similar exercises internationally, and hence on assessors and objects assessed by **non-“metrics-heavy”** peer review, including REF2021.

An example of direct international impact is on and via the National Academy of Sciences of Ukraine. This has over 43,000 employees of which over 17,000 are researchers who generate over 90% of Ukraine’s scientific outputs. Lviv Oblast is national leader for research and has over

Impact case study (REF3)

4,000 researchers supported by a budget of 260,000,000 UAH. In terms of cost per head of population normalised by average salaries this is the same as the UK's QR funding. Science is managed by the Lviv System of Researchers (LSR) who runs the regional counterpart of REF. [R1-5] impacted directly on the LSR, the researchers funded by it, support staff and students [S6]. [R2] changed LSR's understanding, making them aware for the first time that there is a strong dependency of research quality on **group size**. [R4,5] changed LSR's services, practices and policies by halting sole **reliance on metrics** for research assessment. As a result of [R1-5] they now use a "combination of different assessment components instead" [S6]. Impact of [R2-R5] on LSR was deemed so significant that from 2015 they invited Mryglod and Holovatch "to help formulate the shape of [their] process." To quote [S6], LSR confirms [R1-R5] "was an **important argument** for us not to rely solely on scientometrics (as it is done, for example, in CWTS Leiden Ranking or Shanghai ranking) but to base Ukrainian research assessment processes on the combination of different assessment components instead."

Coventry University engaged Spectra Analytics to develop an Excel platform [S9] based on CWC [R6]. This algorithm had direct impact. The Campaign for Real Ale (CAMRA) is a consumer organisation with 16 regions, over 200 branches and over 192,000 members. It ranks quality of over 55,000 pubs. CAMRA provided 2016 data from six branches to be analysed using CWC. CWC increasing ranking efficiency by at least 100%, and this has been disseminated through CAMRA branch and leadership. Online publication broadened the reach of 2,000 hard copies of Coventry's branch magazine which reported on [R6] and CWC before lockdown interrupted further activities [S7]. Ofqual (Office of Qualifications and Examinations Regulation) regulate qualifications, exams, and tests in England. They have 200 permanent employees, a budget of £17.5 million and a dedicated research team. They held six meetings with Kenna and Low and provided marking data for CWC. A report was presented to senior Ofqual research staff, detailing how CWC could be used in the training of examiners and assessors. Ofqual testify "We were particularly engaged and interested in how the report showed that the CWC system detected that chief examiners tend to mark more generously than less experienced examiners, and that this might have application in examiner training processes." They conclude: "we anticipate CWC will impact on a broad scale" [S8].

5. Sources to corroborate the impact (indicative maximum of 10 references)

[S1] (a) pdf report Wilsdon, J., et al. (2015) 'The Metric Tide: Report of the Independent Review of the Role of Metrics in Research Assessment and Management'; (b) pdf report Wouters, P. et al. (2015) 'The Metric Tide: Literature Review (Supplementary Report I)', HEFCE; (c) email from Author, Science Policy Research (8 July 2015) and email from Director, Higher Education Policy Institute (26 Feb 2015)

[S2] (a) pdf report 'Growing the best and brightest: The drivers of research excellence', March 2014. Ref: BIS/14/689; (b) pdf report Manville, C., et al. (2015) 'Characteristics of high-performing research units', for the Higher Education Funding Council for England; (c) pdf report A. Lyytinen et al. (2018) 'Research Evaluation Report 2018', University of Jyväskylä,

[S3] (a) Report from OECD (2016) 'OECD Economic Surveys: Norway 2016', 128 pages; (b) pdf report Aksnes, D.W., et al. (2018) 'Does size matter?', NIFU, Research and Education

[S4] (a) pdf taken from Jump, P. (2014) 'The (predicted) results for the 2014 REF are in', Times Higher Education (online) [accessed 12/3/2021]; (b) pdf taken from Jump, P. (2015) 'Academic estimates 'real' cost of REF exceeds £1bn', Times Higher Education (online) [accessed 12/3/2021]; (c) pdf taken from Jump, P. (2015) 'Can the research excellence framework run on metrics?', Times Higher Education (online) [accessed 12/3/2021]

[S5] (a) pdf taken from Burke, M. (2014) 'Can research quality be predicted by metrics?', Chemistry World (online) [accessed 12/3/2021]; (b) pdf taken from Burke, M. (2015) 'Time spent assessing research impact was worthwhile', Chemistry World (online) [accessed 12/3/2021]; (c) pdf taken from Burke, M. (2015) 'Metrics failed to predict REF outcomes', Chemistry World (online) [accessed 12/3/2021]

[S6] Testimonial from Ukrainian Adviser to the Lviv City Mayor & Chief coordinator for Lviv System of Researchers (22 March 2020)

[S7] Testimonial from NBBS Co-ordinator / Former National Director of CAMRA et al

[S8] Testimonial from Director of Research and Analysis, Ofqual (5 May 2020)

[S9] <http://www.calibratewithconfidence.co.uk/home>