Pilot Training on the Responsible Use of Research Metrics
WP4: Post-Testing Interview Qualitative Report

Analysis by Dr Hugh Turpin

Project Team

Colleen Thomas, Research Culture & Engagement Manager, UCD
Patrick Phillips, Research Educational Technologist, UCD
Emma Dorris, Engaged Research Manager, UCD
Liam Cleere, Senior Manager, Research Analytics & Impact, UCD
Fiona Brennan, Director of Research Support, DCU
Patrick Murray, Head of Research and Technology Transfer, TUS
David O'Connell, Director of Research Support & Policy, UCC
William Fitzmaurice, Institute Manager, UCD Earth Institute Manager
Michelle Norris, Director of the Geary Institute for Public Policy and Professor of Social Policy, UCD
Grace Mulcahy, Research Integrity Officer; Chair of the Research Culture Initiative, Project Lead, UCD

Contents

Executive Summary

<u>Very high satisfaction</u>: Most are very satisfied overall. As one said, 'Much better than the average training module, but that's a low bar'. One notable trend was that the less someone knew about metrics to begin with, the more enthusiastic they were about the module. Given response biases above, this may mean broader academic population will be even less critical than this group with its high level of 'meta' types. And as discussed below, the sharpest critics probably cannot be won over. Together, this suggests that radical changes are unnecessary. Nevertheless, a number of suggestions for fine-tuning can be made based on the feedback.

Interviews: Content and Structure

First, participants were asked about their role in the university. Then they were asked to report their general attitudes towards research metrics. After this, they were given a general prompt for feedback on the module. Next, specific questions were asked relating to four areas: discipline specifics; role specifics; knowledge confirmation; aesthetics and content. Interviews closed with a prompt for any further thoughts on how to make the module better. The question list is in appendix 1. The structure was deviated from if, for example, specific areas were already covered by the initial general prompt, or if unforeseen avenues of exploration presented themselves during interviewing.

Sample Overview

(N=18) [UCC =4 / UCD =6 / DCU =4 / TUS =4]

<u>Early career academic (Postdoc, ECR)</u>: 6
- *Education: 1*
- *Science: 4*
- *Social science: 1*
- Meta Quotient (2/6): Educational field x 1; 'research' as an object of study x 1

<u>Established academic</u>: 8
- *Humanities: 2*
- *Social Sciences: 1*
- *Science: 3*
- *Education: 2*
- Meta Quotient (7/8): Deans of Research x 2; Educational field x 2; 'research' as object of study x 1; involvement in module x 1; 1 x research integrity champion

<u>Administrative (Research officer, etc)</u>: 4
- *Librarian: 1*
- *Research Officer: 3*
- Meta Quotient (4/4): 'research' as an object of study/instruction x 1; research officer x 3

Sample Limitations

<u>Very high number of 'meta' individuals</u>: Meta individuals have valuable insights on content and design, but their experiences of taking the module may be unrepresentative due to a substantive rather than pragmatic interest in research metrics.

<u>Lack of early career humanities/social sciences participants</u>: A persistent arts/humanities critique of metrics is that they take a reductive mode of evaluation modelled on the sciences and apply this outside of its proper domain. The emphasis on altmetrics was considered an inadequate proxy. These kinds of critiques were made by the established arts academics above. However, we lack entry-level arts and humanities participants and therefore we do not know whether such informants would have felt the same way.

Findings by Theme

Discipline Specifics

- Not much support for having separate modules for disciplinary areas; most liked the overview. 'Good to know how the other side are evaluated'. Also, content is already quite well divided by category in the module. Furthermore, substantial disciplinary differences around metrics exist within master categories too (e.g. physics vs biology), so where would it end?
- Include some more examples that do not fit neatly into the three master categories, such as business, education, art, etc. Could possibly gather these under heading 'interdisciplinary or non-traditional research areas', or have a subsection on same. Note: there may already be too many healthcare examples according to one informant.
- According to one senior research officer, it may be the case that evaluation of researchers within most arts and humanities disciplines is simply not amenable to *any* form of metricisation. Reputation among other scholars in the field may be what matters most, but this can't be measured, or at least no measure currently exists for it. Thus, you may have to adapt to the fact that AHSS researchers will almost inevitably feel the course is an example of STEM norms being foisted upon them, no matter how well meaning. It's not that they want metrics to be improved; their minimisation in favour of qualitative appraisals may go some way, but really they don't want metrics to feature in their evaluation *at all*.
- Illustrative quote: "Until the AHSS side gets some sort of formal research assessment foisted upon them, I don't think you are going to get the kind of engagement you'd like. So you're probably just going to have to put up with people not liking it. Eventually, a research assessment system will be foisted on them, just like in STEM, and you will be able to come in as a hero able to talk to them about reform of this terrible system, but right now, they're not subject to these kind of metrics, so why would they engage?"
- This connects to an issue with the humanities sections: an overemphasis on social media engagement and public discourse influence as primary non-traditional 'metrics'. This catalysed a certain amount of opposition in its own right. For monograph-centric traditional humanities disciplines for instance, consider a section on peer-reviewed books published by not-for-profit prestige publishers (university presses etc).

Role Specifics

- The examples tend to be high level (e.g. social scientists influencing public policy). There need to be some examples of achievements that are more realistic for early career researchers.
- One informant suggested a section on ideal vs real practices, so that ECRs are not misinformed about the current importance of non-traditional metrics. This connects to a tension in the module: is it for practical information on career progression or is it a tool for moulding future research culture?
- Somewhat more support for splitting module according to role, such as ECR vs evaluator, or researcher vs research manager. If you split it, maybe do so to produce a 'lite' version for postgrads and ECRs. Keep the original for managers, heads of research, etc.

- If you split it, one senior research officer emphasises that it would be unfair to expect change to be bottom up. So perhaps the module for evaluators should emphasise their responsibility for cultural change, while the module for ECRs and so on should be more pragmatic. While still emphasising that metrics are flawed and need to change, it could also describe how and why they are so embedded in institutions, and that they must use them strategically all the same, at least until they get to a point of security where they might be able to turn around and influence things.
- Evaluation culture changes over time, so the module will need to be regularly updated so as not to misinform participants, especially ECRs and other junior level types seeking advancement.

Knowledge Confirmation

- There is substantial support for making the knowledge confirmation sections more 'interactive'.
- However, this should not come at the cost of slowing down progress through the module as this will create frustration.
- Quizzes etc ok to sum up or highlight gaps in knowledge, but be cautious about making progression dependent upon passing these. The promise of a badge will not be enough to offset this.
- Only advocates of 'testing' were research managers who saw tests as a method for preventing 'cheating' or 'skipping through'.
- Various ideas were floated, from multiple choice quizzes to integration of a chatbot to evaluate reflection point responses entered into a text box.

Aesthetics

- The most common complaint about the module is that it is **repetitive**; the second most common is that it is **text heavy**
- Charts or visualisations (perhaps especially interactive ones) are the most popular suggestions for reducing text density
- Hyperlinks referring back to earlier definitions or sections rather than repetition within sections
- Judicious use of colour to break up monochrome text walls (e.g. colour code master categories, or subtopic boxes, etc)
- Some support for summary videos, but only as options. Padded out videos may increase frustrations around length, so careful to keep them informative and short. Graphics more popular than videos.
- Illustrations should include people doing non-scientific research (not all people in front of computers)
- DEI concerns around people in pictures mentioned by one or two
- Careful about overuse of AI images: exacerbating problem for critical types of illustrating 'soulless quantification of research' with 'soulless co-optation of human creativity'. Additionally, unexpected elements in the pictures can be weird or unintentionally comic. For instance, one image says 'pubic' instead of 'public'.
- Some said it was too long, but many also appreciated the thoroughness. Might be better to address length by reducing repetition and text density than by trimming out topics
- Overuse of bold type
- Lukewarm support for idea of completion badge

- Altmetrics: How do you boost them? How do you increase social media presence? 'Superficial and not everyone likes social media'
- Narrative CVs: what would these look like for early career researchers? How and where can you publicly display your narrative CV (is there a platform)? How do you evaluate a narrative CV? Problem of 'onwards and upwards, strength to strength' obligatory narrative in such CVs: cultural issues; issues for those with career gaps etc.
- 'Objectives' framed as 'outcomes'
- Discussions of h-index do not take context into account. What is a good h-index score? For what area?
- Make module available as a PDF handbook for reference
- Should have a reference section
- A number of (primarily more critical) respondents suggested there should be more 'problematisation' of metrics as a whole, perhaps a section giving higher-level or philosophical critiques.
- Socio-economic critiques exist, but whether or not they should be broached in any hypothetical critical section is a difficult question. For instance, one insider maintained that Irish universities' fixation on bibliometrics was primarily because these metrics affect university rankings. High rankings then attract the international students upon whom the sector has become financially reliant. Such deep-seated structural critiques call into question whether reform is even possible, so should be approached with caution.
- Could also have greater number of optional branches deep-diving into more specialised subtopics.

The dissatisfied minority:

Sample too small to do this really, but we can attempt an 'ideal typology' of the dissatisfied: 'metrics specialist' versus 'metrics pessimist'.

- <u>Metrics specialist</u>: 1) professional specialists in metrics research, 2) ECR scientists enthusiastic about research reform, 3) senior researchers 'who know it all already'. Core critiques: not up-to-date / deep enough in area X; or 'I know this already'. Method of satisfaction: specialised informational sidequests 'for those with advanced knowledge' in this area; foreshadowing of section content with opportunity to skip ahead

- <u>Metrics pessimist:</u> arts, humanities, or social sciences background. Moral/ideological opposition connects research metrics to threatening projects such as 'scientism' and 'neoliberalism'. Metrics are 'reductive' and 'destructive'. Core critique is that the module does not sufficiently 'problematise' metrics. Probably cannot ever satisfy this kind of participant (see also comments on AHSS in discipline section above). One strategy may be to placate by showing 'recognition'. This could be done through an (optional) section on anti-metrics critiques. Could also try damage limitation strategy by 'covering cosmetic bases' to avert escalation from disgruntlement to indignation. This would involve cosmetic fixes such as more non-science research in pictures; diverse identity groups in pictures; less use of AI images, etc).

- <u>Limits on satisfiability of critical subgroups:</u> Embedded critical stances: specialist seeks distinction through displays of discernment and superior knowledge; pessimist primed for opposition, is less impacted by metrics anyway, and engages with content from a vigilant deconstructive standpoint.

Appendix: Question Schedule for RURM Post-Test Interviews

I. Intro / Warm-up

We'll begin with some general questions about your role in the university and your general feelings about research metrics to better contextualise your subsequent thoughts about the module itself. Please rest assured that your name and institution will be anonymised. Additionally, I have had no role in the design of the module, so please feel free to express yourself as candidly as you wish.

- What is your professional role within the university?
- Which academic discipline(s) are you associated with?
- At a very general level, how do you feel about research metrics?

II. Module: General Questions

Before looking at specific areas of the module, I'd just like to get an overview of how you felt about it.

- Overall, did you find the module useful? Why/Why not?
- What strikes you as:
  A) Something the module did well?
  B) Something that could be improved?

III. Specific Themes

There are four specific areas about which we would like to elicit some more detailed feedback. These are: discipline specifics; role specifics; knowledge confirmation; and aesthetics and presentation of information. I'll ask you some questions about these four areas now:

1) Discipline Specifics
- How are metrics used and misused in your area, and do you feel the module sufficiently addressed all these issues?
- Was the module balanced in its coverage of metrics and their use across different academic disciplines?
- Would you prefer if the module were redesigned to have discipline-specific paths?

2) Role Specifics
- How relevant was the information provided to your specific role within the university? What could be better?
- Is the module equally useful for all roles, from researcher to research manager to research evaluator?
- Is it equally valuable to all levels of seniority within the academy, i.e. from postgraduate students and ECRs to senior faculty and research management?
- Would you prefer if the module were redesigned to have role-specific paths?

3) Knowledge Confirmation
- Do you remember any of the practical exercises the module deployed at the end of each section? Could these have been improved?
- Would you prefer some other method of self-assessing? If so, what?

4) Aesthetics and Content
- How was the module in terms of length?
- How did you find the text/language?
- How were the images? Would you like to see any changes?

- Would additional graphics and visualisations be beneficial? Anything in particular you would like to see?
- Would you prefer a video summary or a text summary at the end of each topic?

IV. <u>Finishing up</u>

Many thanks for your time today. Just one last question:
- Is there anything that comes to mind now that I haven't covered above, and that might make this module (even) better?