

# Peer review and Impact

## *Lessons about evaluation mechanics and their artefacts from REF2014*

Dr Gemma Derrick



# Peer review

---

*“The essential principle of peer review is simple to state: it is that judgments about the worth or value of a piece of research should be made by those with demonstrated competence to make such a judgment.” The British Academy*

- Pillar of academic self governance
- The gold-standard in research evaluation
- Benefit of a discussion by peers and experts.
- Peer review of scientific impact works as there are shared community understandings, interpretations and value of “scientific impact” (Lamont, 2009)
- Peer review outcomes only accepted if the process is perceived to be fair by the research community (Tayler, 2006)
- A number of accepted shortcomings e.g. gender bias, halo effect and cognitive cronyism.

# Evaluation mechanics in Peer review

---

- Peer review is a black box – neither transparent nor accountable
- Rules, structures and/or guidelines that are used in large scale evaluations
  - From Knorr-Centina (1999) and Lamont (2009)
  - Play a vital role in legitimising the process (e.g. selection criteria)
  - Pseudo-transparency of a black box
- Peer review as an “organised evaluation process” (Chubin & Hackett, 1990)
- Helps groups navigate ambiguous evaluation objects – aka “Impact”

# Peer review: Evaluation by groups

---



# How do groups work?

---

- Behavioural economics and group-psychology (Janis, 1982; Levi, 2011; Thaler & Sunstein, 2008; Sunstein & Hastie, 2015)
  - Consensus forming
  - Polarisation
  - Groupthink
  - Nudge
  - Satisficing (Lamont, 2009) or Pragmatic evaluative practice
- Guided by a negotiated definition of “excellence”
  - Not always the same – achieved through deliberation
- Succeeding in peer review a matter of “luck of the reviewer draw), and not just excellence (Cole et al, 1981)

# Claire Oliver

(1981-2007)

---



# Peer review for societal impact is different

---

- Under researched due to lack of opportunity (ex ante and ex post)
- Who is considered an “expert” and who is a “peer”
  - The inclusion of user-evaluators
  - Reaching a common understanding through discussion difficult
- Evaluators have little prior experience to base decisions
  - Bring in new, different biases and tendencies to evaluation
  - Change in group dynamics
- Risk of time poor process – no time for experimentation and for things to go wrong
  - Adopt a pragmatic approach to evaluation
  - Potential misuse/use of proxies

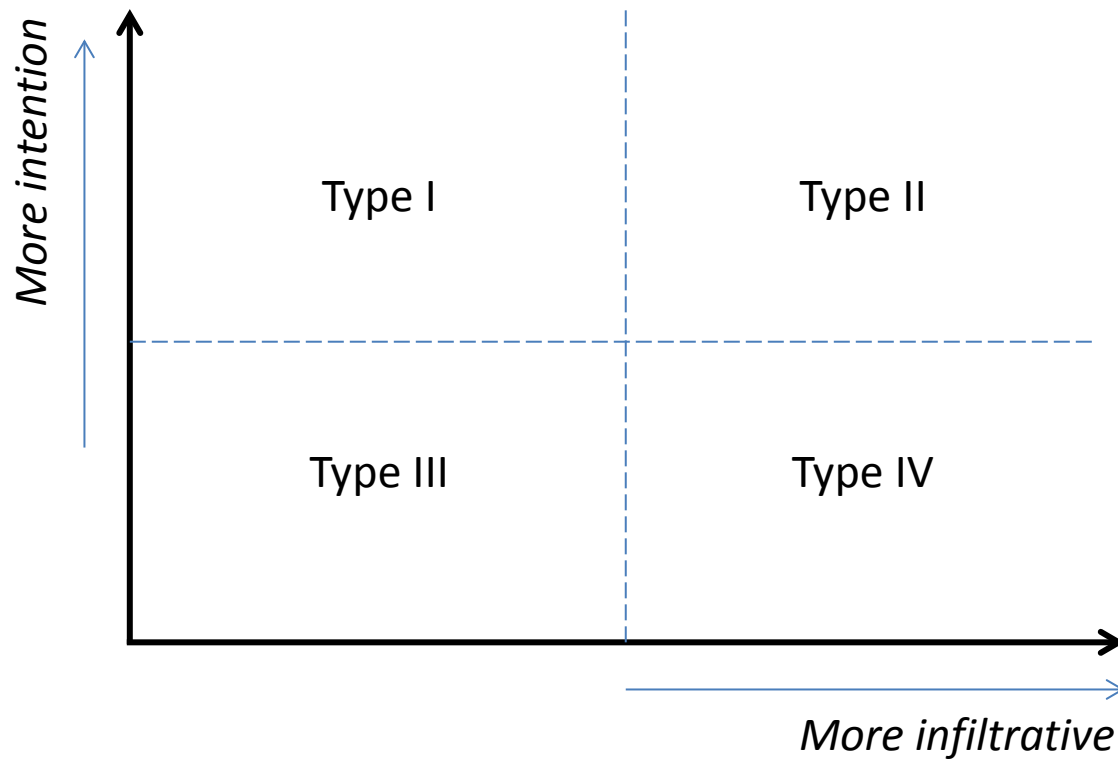
# Typologies of Evaluation mechanics

	<i>Infiltrative (<math>I_1</math>)</i>	<i>Facilitative (<math>F</math>)</i>
<i>Intended (<math>I_2</math>)</i>	<b>Type I</b> $I_1 - I_2$	<b>Type II</b> $F - I_2$
<i>Unintended (<math>U</math>)</i>	<b>Type III</b> $I_1 - U$	<b>Type IV</b> $F - U$

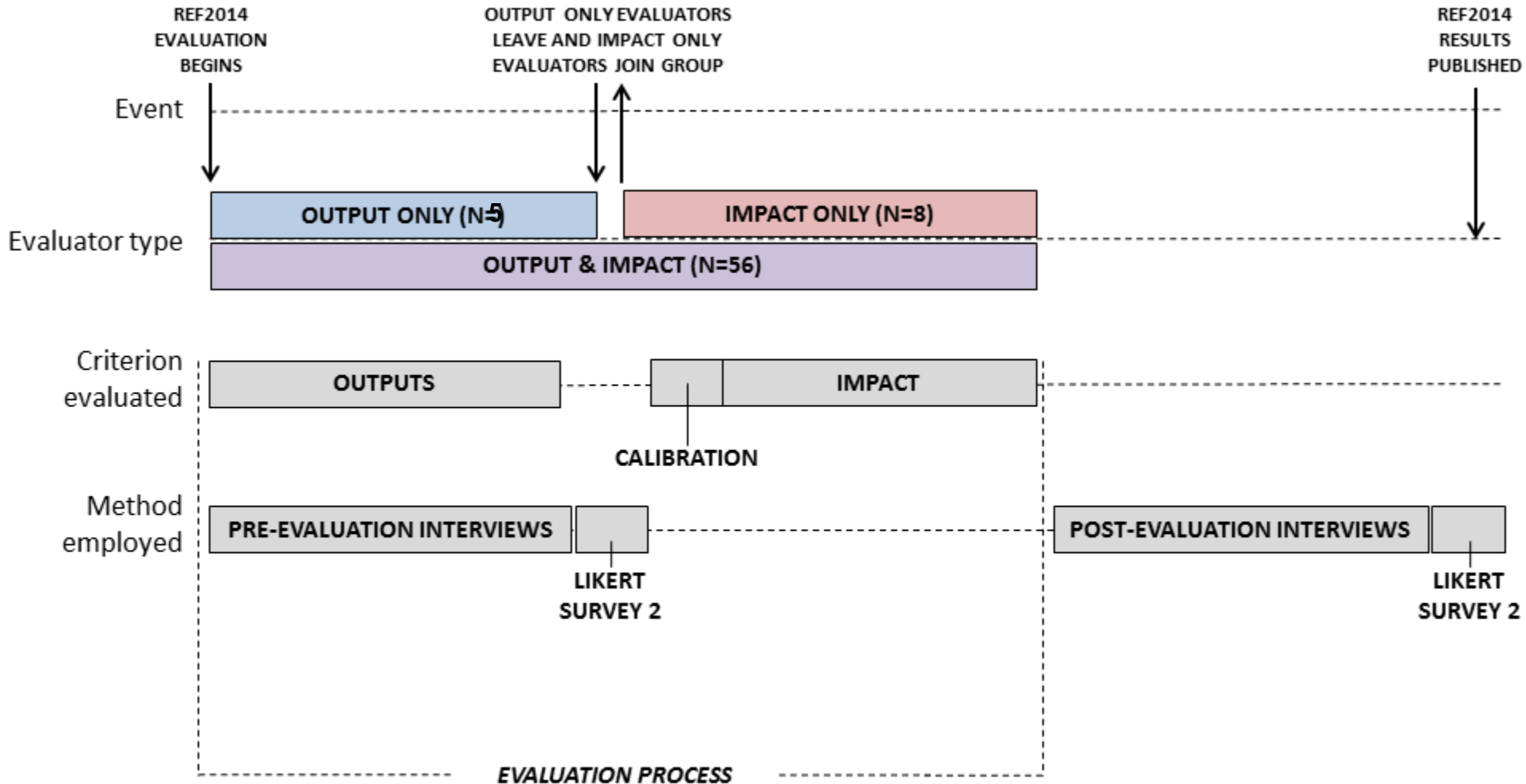
- **Type II** - Ideal type for peer review
  - Facilitative with intended outcomes – but mechanics should not overshadow the group-level deliberation of peers
  - Expert-driven peer review should not be mechanic-driven review
  - No mechanic artefacts present in evaluation outcomes
- **Type I** and **Type III** to be avoided
  - Mechanic artefacts will be present in evaluation outcomes



# Scale of types



# Research design



## Overall quality profile

Quality level	4*	3*	2*	1*	u/c
% of research	12	37	41	10	0

The overall quality profile comprises the aggregate of the weighted sub-profiles produced for outputs, impact and environment.

### Outputs

4*	3*	2*	1*	u/c
12.8	32.8	43	11.4	0

65%

### Impact

4*	3*	2*	1*	u/c
20	45	35	0	0

20%

### Environment

4*	3*	2*	1*	u/c
0	40	40	20	0

15%

# No precedent for Impact evaluation in practice

---

## Uncertainty of Impact

### EX-POST EVALUATION FOR REF2014 ONE “BIG EXPERIMENT”

*“I think it's all new territory for all of us, and none of us know – we are going to learn on the job I think.” P4OutImp6*

### LACK OF DEFINITION

*“I’m still not convinced everybody shares exactly the same definition of what constitutes impact or where they place the weight or if it’s impact or isn’t” P3Imp1*

### LACK OF EXPERIENCE

*“I’m very happy to describe the quality of the research [but] the valuing of impact is something I have no idea about” P0P2 OutImp1*

*“And I don’t believe that we know how to do it- you have to contrast this with the assessment of outputs which is really just reviewing, which is bread and butter stuff for an academic. That’s what we cut our teeth on, that’s what we do every day and so there may be an awful lot of it...but it is just what we do. Whereas this impact stuff we just don’t know. So I feel a little bit nervous about it.” P2 OutImp1*

# REF2014 Impact mechanics

Submissions	Evaluation	Interpretation
Impact definition	Impact definition	Impact definition
Star ratings (U-4*)	Star ratings (U-4*)	Star ratings (U-4*)
Criteria (Significance and Reach)	Criteria (Significance and Reach)	
Submission guidelines		Submission guidelines
Template – Impact and Impact Case studies		
	Main Panel A	
Panel membership		Panel membership
	Panel working methods	Panel working methods
	Peers and Experts	
		Report on panel working methods
	Calibration exercises	

# REF2014 “Impact” assessment

*“...an effect on, change or benefit to the economy, society, culture, public policy or services, health, the environment or quality of life, beyond academia”*

Impact template (REF3a)



Institution:
Unit of Assessment:
a. Context
b. Approach to impact
c. Strategy and plans
d. Relationship to case studies

Impact case study (REF3b)



Institution:
Unit of Assessment:
Title of case study:
1. Summary of the impact <i>(indicative maximum 100 words)</i>
2. Underpinning research <i>(indicative maximum 500 words)</i>
3. References to the research <i>(indicative maximum of six references)</i>
4. Details of the impact <i>(indicative maximum 750 words)</i>
5. Sources to corroborate the impact <i>(indicative maximum of 10 references)</i>

# Main panels and UoA (sub-panels)

## Main Panel A

### **UoA 1**

Clinical Medicine

### **UoA 2**

Public Health, Health Service  
and Primary care

### **UoA 3**

Allied Health Professions,  
Dentistry, Nursing and  
Pharmacy

### **UoA 4**

Psychology, Psychiatry and  
Neuroscience

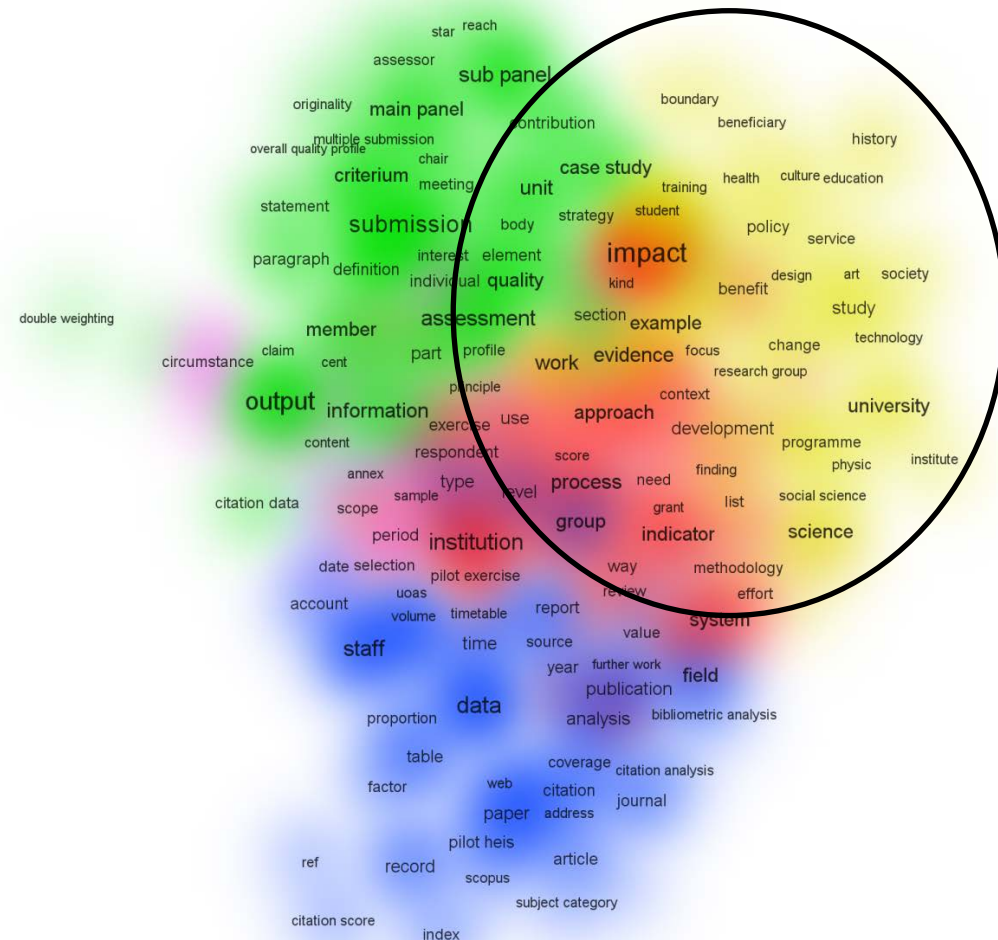
### **UoA 5**

Biological Sciences

### **UoA 6**

Agriculture, Veterinary and  
Food Science

*Derrick et al (2014) A co-word analysis of the UK REF2014 policy documents using VOSviewer. Research Evaluation.*





(Kings College London & Digital Science, *The Nature, Scale and Beneficiaries of Research Impact: An Initial Analysis of Research Excellence Framework (REF) 2014 Impact Case Studies*, HEFCE, 2015: 29)







# Research on “impact case studies”

- Acceptance of “impact” examples independent of consideration of the group-construction of its evaluation
  - Reinforces the role of the mechanics – does not evidence impact
- REF2014 “Impact” is not “Impact”
  - Are constructed narratives
  - 4 star no more “impactful” than 1 star
- Type I or Type III mechanics?
  - Depends on how intended were the consequences?
  - Further examined through evaluative practice (cross tabulation between pre- and post-evaluation interviews)

# Mechanics in practice

## - Evaluation guidelines

---

- POST-EVALUATION INTERVIEW DATA – “da bible”

*“...it helped to clarify what the direction of travel should be for impact.”*

*P0 OutImp2(POST)*

*“...it was also helpful in that over and over it was pulling out -- all the REF teams were able to refer back to the REF guidance which after all was a Bible for our assessment, whatever your view of impact or otherwise. The REF guidance was the thing that you should be referring back to.”*

*P3 Imp2(POST)*

*“-- we were so well trained that I think if somebody put me into a physics subpanel I probably could have judged their impact.”*

*P3OutImp9(POST)*

*“we stuck very much to the letter of what was in the instructions”*

*P4OutImp1(POST).*

*“it was a fairly mechanistic operationalization [of impact]”*

*P2OutImp6(POST).*

# Mechanics in Practice

## - Calibration exercises

---

### Pre-evaluation

- *“part of our calibration exercise will be to make sure that we have heard and digested, and hopefully adopted, each other’s points of view”*  
P0P2OutImp1(PRE)
- *“Research excellence, I think it’s more straightforward....for impact its new, and we haven’t gone through comparing different assessors and how consistent their assessment is and how we’re going to get agreement at the end, it might be very high, it might not be, I don’t know”*  
P1OutImp2(PRE)

### Post-evaluation

- *“get the exact ground rules sorted out”*  
P1OutImp4(POST)
- *“the REF people...went through it very, very clear[ly] about what we were to look for and the steps you were to go through in assessing impact”*  
P3OutImp4(POST)

# Infiltrative mechanics?

---

*“Despite the fact that we all went in saying, **how are we ever going to get this right, how are we going to judge what is a one, two, three or four with these rather vague written cases. But when it came to it, it wasn’t nearly as hard....and that’s partly due to the fact that we got very clear advice from HEFCE as to how to interpret what we saw and how to grade what we saw”.***

*P1OutImp4(POST)*

# Group consensus

---

## Pre-evaluation

- Concern that there would be a sway by some evaluators towards the narrative of case studies, and not the Impact

- Big concern that User-evaluators would be the worst offenders

*“I would expect a difference on the weight they place on the real quality of the underpinning research that it was offering something genuinely new and has strong, if you like, evidence of the narrative or persuasive the narrative if the causality might be”*

P2OutImp1(PRE)

## Post-evaluation

- Common agreement about the value of the narrative

*“Everyone agreed that the narrative aspect was an important part”*

P0OutImp5(POST)

# Group dynamics in peer review

---

Group psychology tendencies such as social loafing, happy talk, groupthink and identity setting all identified

*“The panel, I'd just like to emphasis, you know, we treated the whole exercise as an exercise in collective responsibility, and our panel decided to do things and I'm with the panel on that. I have a personal view but that is not my public view.”*

**Academic (Post-evaluation)**

*“And so we spent a reasonable amount of time discussing these because it was a way of establishing our principles when we looked at all the others.”*

**Academic (Post-evaluation)**

*“...we were like a team, suffering together”*

**Academic (Post-evaluation)**

*“....I don't know how other panels work, but my panel was quite – perhaps had quite a strong culture shall we say.”*

**User (Post-evaluation)**



# Conclusion

---

- Examining just the submissions, and evaluation outcomes to find evidence of impact is not sufficient.
- REF2014 Impact is not Impact
  - It is not separate from the evaluation process
- Role of mechanics – food for thought
  - A process that is “expert-driven” should not let evaluation mechanics trump expertise.
  - For a ambiguous evaluation object like impact was..... Not such a bad thing?
- Book “Inside Impact....” by *Palgrave Macmillan* in 2017.