Peer Review, Evaluation and Policy Learning

Erik Arnold
www.technopolis-group.com

9 November 2007
If you don’t understand history, you understand nothing … Peer review is one of the most deeply political aspects of R&D evaluation.

- **Theory**
  - 1: Technology Push
  - 2: Needs Pull
  - 3: Coupling, Complex Systems

- **Subsidy Focus**
  - Big Cos, National Champions
  - SMEs, Tax Incentives

- **Policy**
  - Foresight
  - New programme forms
  - Funding reforms
  - University reforms
  - Collaborative programmes
  - Economic, military competition
  - Commercialise RIs, RAs

- **Timeline**
  - 1950s: Build up Universities, RIs, RAs
  - 1960s
  - 1970s: Coupling, Complex Systems
  - 1980s
  - 1990s
The intellectual battle has culminated in a ‘national innovation systems’ perspective that undermines the special status of science and the scientific community.

Source: Arnold and Kuhlmann, 2001
Peer review

• Originally developed to decide about the suitability of articles proposed for publication in academic journals (17th century) - still probably the dominant mode essentially a judgement about scientific quality

• Much later (20th century), peer review was extended to play a ‘gate keeping’ role in the access to research resources via Research Councils

• In the late 20th century, the concept is extended further (‘extended peer review’) to tackle non-quality questions such as relevance (ex ante) and impact (ex post)
## John Rigby’s classification (Plattform FTE, 21, June 2004)

<table>
<thead>
<tr>
<th>Sub-Type of Expert Review</th>
<th>Science</th>
<th>Level of Specialisation</th>
<th>Level of Professionalization</th>
</tr>
</thead>
<tbody>
<tr>
<td>Traditional Peer-Review</td>
<td>Academic Science</td>
<td>Increasing</td>
<td>Increasing</td>
</tr>
<tr>
<td>(Canonical Academic Review)</td>
<td>Republican Science (Fuller, 2000)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Direct Peer Review</td>
<td>Post-Academic Science</td>
<td>Increasing</td>
<td>Increasing</td>
</tr>
<tr>
<td>Modified Direct Peer Review</td>
<td>(Ziman, 1995) &amp; Liberalized</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Science (Fuller, 2000)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pre-Emptive Peer Review</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Indirect Peer Review</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Merit Review (extended form</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>of Peer Review)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ancillary Peer-Review</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Expert Panels/Peer Review</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Panel Review</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Professional Evaluator</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Extended Peer Communities</td>
<td>Post Normal Science</td>
<td>Specialisation non-relevant</td>
<td>Wider communities - anti-professional</td>
</tr>
<tr>
<td></td>
<td>(Funtowicz, Ravetz)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
A key to understanding peer review is to see its role in the self-organisation of the scientific community

- Rooted in the Mertonian and Humboldtian ideas of science and researchers’ roles in society
- Reinforces Authority within the intensely hierarchical world of science (cp Feyerabend)
- Connects access to resources with conformity to the views of the Establishment (Lakatos)
- Presents the conventional wisdom of the scientific community as the highest form of Truth in the debate between science and society
  - Shorthand: In the UK debate, homeopathy is attached, not because it is wrong but because it isn’t peer reviewed
- Forces conformity: making science students “victims of a history rewritten by the powers that be” (Kuhn)
Despite these important political legitimation roles, the scientific community increasingly questions the adequacy of peer review:

- Slow
- Expensive
- Prone to bias (friendship, tit for tat, positive ignorance bias, intimidation)
- Open to abuse (especially in small systems)
- Sometimes incompetent
- Unable to detect fraud
- No audit trail
- Boundary problem (?): often judges the ‘real project’ rather than the evaluation object

And it’s

- Running out of capacity - eating up the research system it’s supposed to support
In evidence-base evaluation practice, peer review is not a last-resort method choice but a rich source of insight - as long as the evaluators stay in charge of the process by structuring it.
Experience suggests there’s intelligent life out there

Experts’ Scores Compared to Respondents’ Scores

![Graph showing comparison between experts' and respondents' scores across various categories such as General S&T relevance, Relevance to programme goals, Performance, Quality, General S&T impact, Impact on the organisation, Policy relevance, Dissemination, Impact on Users, Overall project score. The graph includes two lines: one for self-assessment (n=35) and another for expert scores (n=35).]
Where do we see peer review done?

<table>
<thead>
<tr>
<th></th>
<th>Before</th>
<th>During/After</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>System</strong></td>
<td></td>
<td>☺</td>
</tr>
<tr>
<td><strong>Institutions</strong></td>
<td></td>
<td>☻</td>
</tr>
<tr>
<td><strong>Programmes</strong></td>
<td></td>
<td>☻</td>
</tr>
<tr>
<td><strong>Projects</strong></td>
<td>☻ ☻ ☻</td>
<td>☻ ☻</td>
</tr>
</tbody>
</table>

- National reviews: SF, CREST, OECD
- High-level: Authorities attacking or defending institutions, eg FPs
- Lower-level: additional sources of evidence
- Scientific self-management
Leaving aside scientific self organisation, the reasons why peers are wanted seem to have less and less to do with their domain knowledge the higher the level of enquiry.
Leaving aside scientific self organisation, the reasons why peers are wanted seem to have less and less to do with their domain knowledge the higher the level of enquiry.
The political incentive is obvious. A practical corollary is that the higher up you go, the harder it is to make evidence based judgements - especially where planning is inexplicit, cp FP5

<table>
<thead>
<tr>
<th>Overall Objectives</th>
<th>Programme Purposes</th>
<th>Results</th>
<th>Activities</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strengthening the scientific and technological bases of European industry and contributing to the quality of life of its citizens</td>
<td>Goals of Strategic Programme 1</td>
<td>Expected results of Action 1.1</td>
<td>Project 1.1.1</td>
</tr>
<tr>
<td>Goals of Strategic Programme 2</td>
<td>Expected results of Action 1.2</td>
<td>Project 1.1.2</td>
<td></td>
</tr>
<tr>
<td>Goals of Strategic Programme 3, etc</td>
<td>Expected results of Action 1.3, etc</td>
<td>Project 1.1.3</td>
<td></td>
</tr>
</tbody>
</table>

Goals

Criteria

Activities
Recent attempts to use peer review at a high level recognise systems complexity - and are being driven to use background studies, increasingly using innovation system and evaluation specialists

Examples

- Finnish NIS Review
- EU 5-year assessments
- OECD ‘Innovation System’ reviews
- EU-CREST ‘Policy Mix’ reviews
Issues arising

- If the wonks do the work, who really makes the judgements?
- How do you embed learning mechanisms? The degree of learning seems to vary
  - Finland
  - EU
  - OECD
  - CREST-OMC
- At what point do you empower the wonks to speak truth to power?