The Use and Influence of Research Evaluation Studies

Science Europe Working Group ‚Research Policy and Programme Evaluation'

Maarja-Liisa Kärp, Katrin Milzow, Anke Reinhardt, Sten Söderberg, Klaus Zinöcker

Openevaluation2016, 23 November 2016

Motivation and objectives

Motivation

- Origin in Science Europe Working Group on ‚Research Policy and Programme Evaluation‘
- How were evaluation studies commissioned by our organisations actually used? What impact did they have?
- What could be done to make evaluations more worthwhile?

Objectives

On the basis of practical experience at the interface between evaluators, program managers, heads of institutions

- Illustrate influence of evaluation activities engaged in by Science Europe Member organisations
- Better understand the role evaluation plays in the research (funding) system
- Identify different patterns of evaluation use and conditions under which high quality, independent evaluations contribute to the knowledge base for policy implementation and definition
- Draw lessons learnt and explore whether recommendations can be provided on this basis to ensure evaluations are targeted and fruitfully employed

Methods

Case study selection

Case studies were selected to

- Be retrospective, completed evaluation studies
- Cover evaluation objects at the level of the institution, research field or instrument/program
- Cover various degrees of use, from extensive to lesser use (-> enable exploration of potential factors influencing evaluation use)

Biases cannot be excluded

Case studies – Research institutes and research fields

Case study	Evaluator	Date of publication
1. Evaluation of INRA-Research Units by HCERES	Haut Conseil de l'évaluation de la recherche et de l'enseignement supérieur (HCERES)	several evaluations
2. Swedish Research in Mechanical Engineering	International expert committee	2013
3. Swedish Research in Mathematics	International expert committee	2010
4. Research in Mathematics at Norwegian Universities	International expert committee	2012
5. Evaluation of Plant & Soil Sciences in Estonia	Foundation Archimedes, International expert panel	2012

Case studies – Instruments or programmes

Case study	Evaluator	Date of publication
6. DFG Collaborative Research Centres	Joanneum Research	2012
7. DFG Emmy Noether Career Programme	Institut für Qualitätssicherung (IfQ)	2008
8. Swedish Research in Biodiversity	Two expert committees for Scientific Quality and for Relevance	2010
9. Education Programmes in Norway	Panel of experts in educational science from the Nordic countries	2013
10. National Prevention Research Initiative MRC-UK	Independent Scientific Review Group	2015
11. SNSF Fellowship Schemes	Observatoire Science, Politique et Société, University of Lausanne	2010

Case study writing and analysis

Case studies are

- Informed by at least two different sources contributing different perspectives on the evaluation

  -> interviews with commissioning instances, program managers or evaluators
- Written in accordance with a template tested and refined in pilot study
- Analysed and interpreted by at least two persons

General questions		Information gathered	Source
1	How are evaluation activities integrated into the governance of the organization evaluated?	At the Swedish Research Council (VR), evaluations are normally commissions, either internally from the organization´s eight scientific councils, or directly from the government. The VR has a special division for Research Policy and a unit for Evaluation, which is responsible for the management and conducting of most of these commissions. The programme manager (for e.g. a certain research programme or a specific research field) is normally not the project manager for the evaluation study - but the programme manager is often/normally a reference person for (or an adjunct member of) the evaluation team. The evaluation itself is normally carried out by international expert panels (peer review), where self-assessments, bibliometrics (if relevant) and other types of data constitute the basis for he expert panels´ evaluations.	Own knowledge Evaluation Policy of the Swedish Research Council (dnr 113-2009-91, revised 2014-12-15 (in Swedish)
2	Who commissioned the evaluation? What were the motives for conducting an evaluation?	The evaluation of Swedish research in Mathematics was commissioned by the VR´s Scientific Council for Natural and engineering sciences in 2009. The VR decided to carry out the evaluation together with another funding body, the Swedish Foundation for Strategic Research Foundation (SSF), which - together with VR - has traditionally been an important funding source for research in Mathematics in Sweden. The observation has been made, by two informants, that the fact that several funding bodies initiate and commission an evaluation study together – as in the case of research in Mathematics – in itself adds to the substance of the evaluation and increases the probability that the results "land" and will be used. The evaluation can be regarded as a traditional evaluation of a research field. There has been a tradition, not least from the Scientific Council for Natural and engineering sciences, to commission evaluations of research fields on a regular basis. One motive for conducting the evaluation was a "knowledge gap", i.e. a long time – 14-15 years - had elapsed since research in Mathematics had last been evaluated (in 1995), and there was a perceived need to know what had happened. According to one informant, in research, a period of 10-15 or 15-20 years can be a long time in which new research-based knowledge or research-based applications have been developed.	Evaluation report (Evaluation of Swedish Research in Mathematics, Report 16:2010) Project manager for the evaluation at VR (Former) Secretary general for Natural and Engineering Sciences at VR Programme manager at SSF Coordinator for Natural and engineering sciences at VR

1. Which of the following types of evaluation use can be identified in this case study? How relevant are they [1]

	Not at all				Very much
Legitimation e.g. justifying the use of public funds	X Not relevant. The evaluation was *not* used to legitimize a decision that had already been taken - the use was rather the opposite	O	O	O	O
Information providing the public with information on how public funds are being used and to what effect	O	O	O	O	X Relevant. The evaluation provided knowledge about the scientific quality of the research field that was not known for the research

An example: Evaluation of research units at INRA

Commissioned by	French Parliament
Conducted by	Haut Conseil de l'évaluation de la recherche et de l'enseignement supérieur
Explicit goals of the study	All public research units in France are evaluated every 5 years by HCERES in a methodologically standardized fashion. The objectives are to: • Help the units evaluated • Help the 'tutelles' • Highlight strengths of a particular sight
Main findings	The evaluations have resulted in: • Recommendations to continue a unit or not, to amend its area of activity, and/or to merge units

Evaluation of research units at INRA

Types of use observed, Discrepancy observed between formal goals and actual uses, if any	• The main uses are steering and learning uses. • Since research units are evaluated every 5 years there is also a ritualistic element.
Observed impact of the evaluation study	• The evaluations have had impact on budgetary decisions and have led to structural changes in French research towards the general objective of reducing the number of entities and establishing more interdisciplinary groups. • The evaluations have also had learning effects for units concerned
Factors noted behind the observed impact	• Qualitative criteria are perceived to be more conducive to use than merely quantitative indicators or summary grades • Use is facilitated if the emphasis is on future projects of the research units evaluated rather than past achievements • The use of the evaluation is facilitated if the findings are communicated in a less prestigious context to encourage a more constructive discussion and avoid a situation where the evaluand takes a defensive attitude. • Insufficient coordination of evaluations at many levels (projects, individual researchers, units, institutions…) may reduce the impact of evaluations

Stakeholder validation

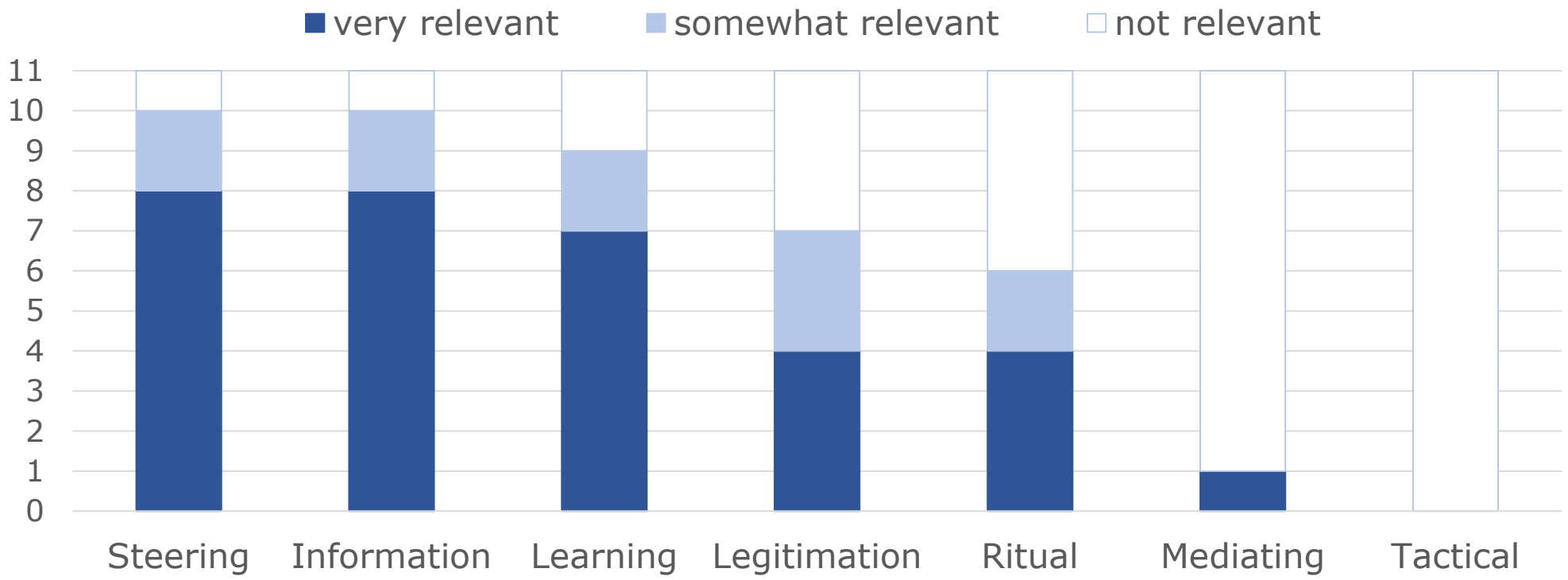
- **Dr Dorothea Sturn**, former Managing Director of the Austrian Science Fund (FWF)
- **Professor Pierre Glaudes**, Director of the Evaluation of Research Units Department, Haut Conseil de l'Évaluation de la Recherche et de l'Enseignement Supérieur (HCERES)
- **Professor Lars Kloo**, Secretary General, Natural and Engineering Sciences, Swedish Re-search Council (VR)

Findings

1) How were evaluation studies used?

All evaluations considered were used in various ways

Types of use identified in the 11 case studies

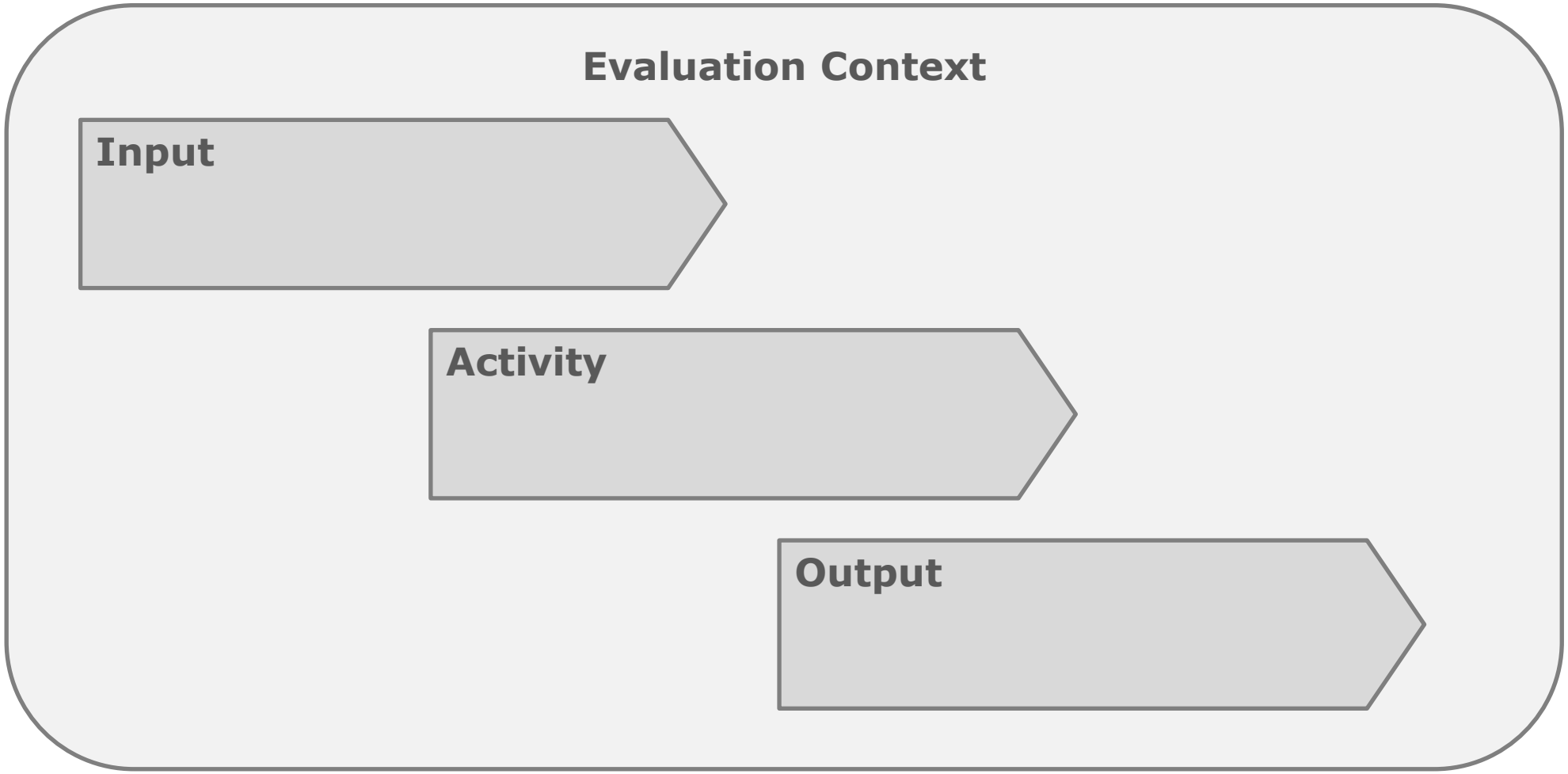


Types of use based on fteval Evaluation Standards and Vedung (1997)

Many uses of evaluation can coexist

- Contribution to knowledge base and cognitive resources should not be dismissed.
- Nevertheless, 10 of 11 case studies display steering use.
- May wide-spread impression that evaluation studies are frequently not used depend on a narrow definition of use?

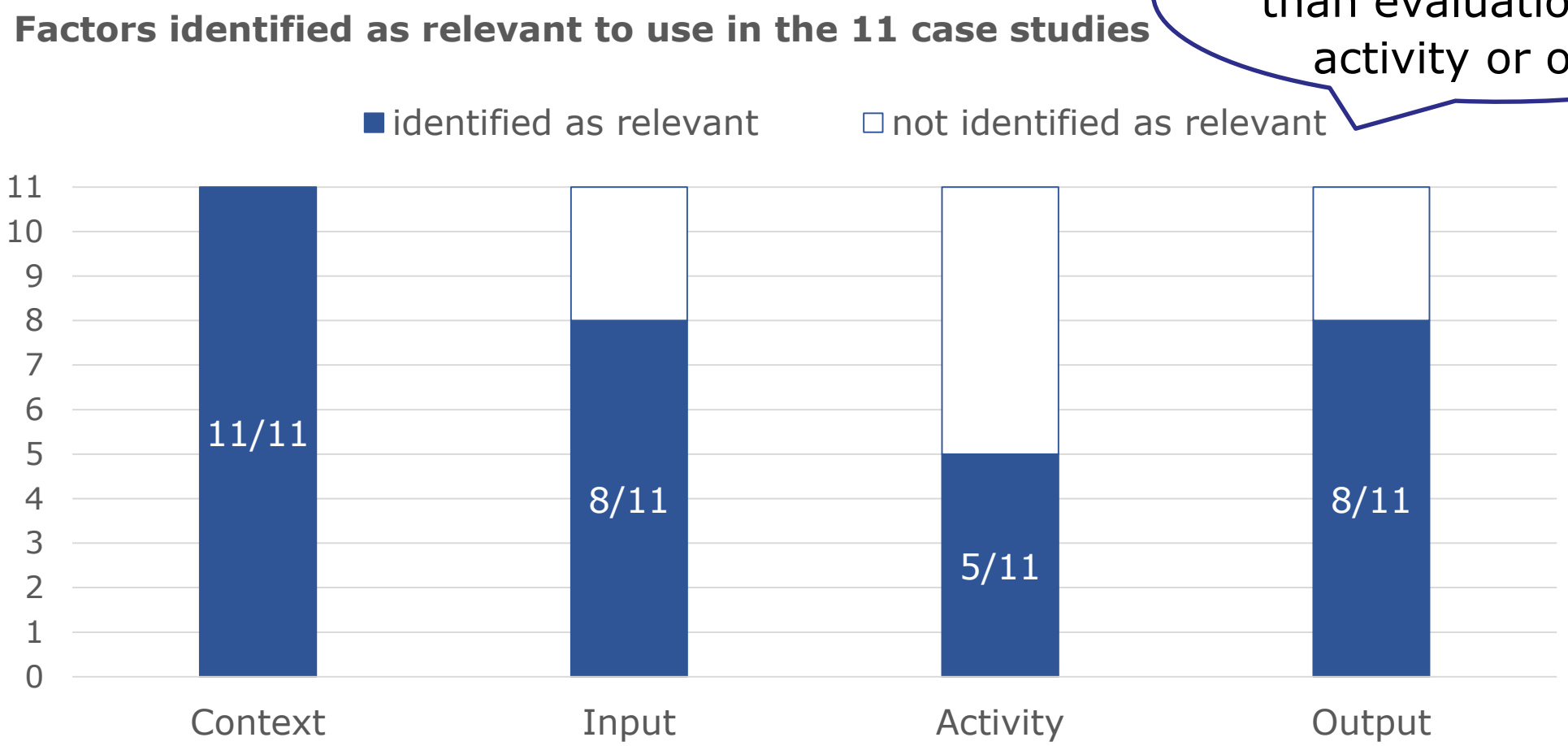
A logic model of evaluation use?



Representation inspired by Mark & Henry (2004)

What influenced the use of evaluation studies?

Context more relevant than evaluation input, activity or output



Evaluation use depends on a number of interrelated factors, without clear discernable patterns.

- High quality is a necessary but not sufficient condition for use
- Context and other factors external to the evaluation itself, but often within the realm of organisations' control, play an essential role
- Certain aspects of the institutional environment may support evaluation use (e.g. follow-up committees, funds for follow-up, generally capacity for reform)
- Above all, evaluation use seems to be determined by case-specific constellations of factors

Conclusions and recommendations

Underlying tensions in evaluation processes

- Power and possibility v. willingness and understanding
- Intended v. unintended uses
- Continuity v. timeliness and methodological innovation
- Time for evaluation v. time for research

Recommendations

Since evaluation use is multiple and depends on a number of interrelated factors, without clear discernable patterns

- Consider scope for evaluation use on a case by case basis before initiating an evaluation
- Apply logic models on the use of evaluation so as to consider underlying mechanisms and chains of influence
- Consider cancelling an evaluation if insufficient scope for use
- Consider interests and possibilities of commissioning and implementing instances. Make underlying tensions/goal conflicts explicit
- Evaluators can support organisations in this process

Possible discussion points

- Do you recognize yourself in the findings? Are there any surprises?
- Are there specificities to the use of *research* evaluation, as oppose to evaluation in other areas?
- Are the recommendations useful?