Practice Insight

Strengthening practice-research connections to improve evaluation practice: perspectives of science communication practitioners from different scientific institutions and countries

Appendix A. Characteristics and recommendations for "good" evaluations
As in Volk and Schäfer, 2024.

The four characteristics of "good" evaluation
1. Be holistic. Evaluations should encompass inputs, outputs, outcomes, and impacts of science communication projects (i.e. the stages of the logic model), and so consider the full spectrum of activities and their effects.

2. Use mixed methods. Evaluations should employ both quantitative and qualitative methods (e.g. surveys, interviews, content analysis, observations, web tracking), and their combination through triangulation to fully capture the results of a project.

3. Be carried out at multiple time points during a project, including during analysis and planning (formative evaluation) and implementation (process evaluation), and not just at its conclusion (summative evaluation).

4. Fit the target audience and format. Questionnaires are not suitable all audiences; for children other formats should be used. Exhibitions and citizen science projects have different levels of engagement and thus need different evaluation approaches.

Seven non-prescriptive recommendations to advance evaluation practice and research
1. Conduct more and whenever possible more robust and well-designed evaluations that go beyond measuring output and aim to capture both short- and long-term effects, using diverse methods.

2. Increase demand and support for evaluation from funders, scientific institutions and policy-makers.

3. Set-up shared evaluation standards that would improve comparability, learning from others and benchmarking.

4. Develop refined and realistic impact measures that capture the non-academic, long-term societal impacts. Since cause-effect relationships will be difficult to establish, these measures should be agreed with funders and scientific institutions.

5. Invest in capacity building, encompassing hands-on evaluation templates, guides, training and networking opportunities, including with science communication researchers.

6. Promote responsibility towards participants in science communication activities (i.e. privacy and data protection), and also the scientific organisations and funders that provide the resources for the activities.

7. Promote open access to evaluation data, for both practice and research purposes.

Appendix B. Methods

Sample selection
Communication leaders from three types of scientific organisation – universities, research institutes and Intergovernmental Research Organisations (IRO) – were chosen to maximise organisational diversity and explore possible commonalities or differences across institutions.

The research organisations are located or headquartered in four countries of which the authors have professional and research experience as well as contextual knowledge: Germany, Portugal, Switzerland, United Kingdom. All four countries have some sort of policy and/or funding programmes for public communication of science [Entradas et al, 2023; Volk, 2024], they have been included in studies investigating structures and communication practices in research centres and universities [Fürst et al, 2022; Entradas et al, 2023], and studies into science communication cultures in European countries [Mejlgaard et al, 2012].

For the semi-structured interviews, Heads of Communications/Research Engagement/Public Engagement were chosen, as they are well-placed to reflect on their own evaluation practice and that of their teams within the context of their organisations' strategic goals. They are also key players (and often the drivers) of the increasing professionalisation of the communications teams [Entradas et al, 2023].

Interview guide
The 10 interviews were conducted through video calls by the first author (AG), between November 2024 and January 2025. Each interview lasted between 60-75 minutes. A semi-structured interview guide was used, with questions distributed in six broad sections (Table B1)

Table B1. Example of an interview guide

#	Question	Sub-question
Section 1: Introduction and Context		
Q0	Could you please introduce yourself - name and position - and your organisation (in 2 sentences)	
Section 2: Concept of evaluation		
Q1	Thinking about science communication, what does "evaluation" mean to you?	
Section 3: Practice of evaluation		
Q2	Do you evaluate or monitor your science communication projects?	**If yes.** Ask for a measure of extent: how often? On which projects specifically? For which channels? With what tools?

		Could you give a few examples?
		If No. Why not?
Q3	Is there someone dedicated to evaluation in your team?	**If yes.** What was the rationale behind having a dedicated function?
		If no. Why not? Would it be desirable? Who is thus responsible for evaluation?
Q4	Who is the target audience for your evaluation reports? Management, own team, other teams, for publication, funders, researchers, …	**If managers.** Do they show interest in the reports? **If no.** Why not? **If the team.** Does the team use evaluation insights? **If yes.** How? **If no.** Why not?
Q5	What are the biggest obstacles to carrying out evaluation and compiling evaluation reports? Please name up to three.	What needs to change to overcome these obstacles?
Section 4: Awareness of evaluation research and resources		
Q6	Are you aware of resources/toolkits/guidelines in your evaluation? Have you used them? E.g. WiD resources, SFI Guidelines, EU projects, others	**If yes to having used them.** Which? How did you find them? And what is your assessment?
		If no to having used them. Why not?
Q7	Have you ever worked with science communication researchers, consultants or agencies in evaluating your projects?	**If yes.** With which? How did that collaboration come about? What was the outcome, both practical and behavioural.
		If no. Why not? What kind of collaboration could be useful?

Q8	Are you aware of research into evaluation of science communication?	**If yes.** How? Which? Do you find it useful?
		If no. Do you think it could be useful for you?

Section 5: Views of the characteristics and recommendations described in Volk and Schäfer, 2024

Q9	Your views on the four ideal characteristics of "good" evaluation (Volk and Schäfer, 2024): 1. Be holistic (logic model) 2. Use mixed methods 3. Cover different time points during a project (formative, summative) 4. Match evaluation to the target audience and format How familiar are these characteristics to you? How relevant? How feasible to put into practice?	
10	Your views on the recommendations from Volk and Schaffer, 2024: 1. More and better evaluations; 2. More demand and support from organisations; 3. Define shared standards to make data comparable; 4. Capacity building: training and networking, shared guides 5. Be responsible to participants, funders, organisations; 6. Better impact measures to capture long-term contributions 7. Make reports and data publicly available Are they relevant? Are they feasible?	**If not relevant or feasible.** Why? Which recommendation would be a priority?

Section 6: Proposals for researcher-practitioner interfaces

11	Name one way in which evaluation research(ers) could help you in your practice	

Data collecstion and analysis

All interviews were recorded and automatically transcribed using Amberscript, and transcriptions were manually checked for accuracy. Qualitative content analysis was conducted by the first author in MaxQDA, followed by discussions with the other authors. The category system in MaxQDA was informed by the structure of the interview guide and deductive categories were derived from the literature on the characteristics and

recommendations for evaluation (e.g. concept of evaluation). Following Rädiker and Kuckartz (2019), inductive categories were developed through iterative coding, beginning with an initial reading and open coding of the interview transcripts, followed by the formation and refinement of new categories until no new themes emerged. Coding was understood as a reflexive process (Schreier, 2012), and the first author engaged in ongoing reflexivity during analysis and interpretation, continuously reflecting on prior professional experiences, comparing the data with existing literature, and discussing interpretations with the co-authors. The dual role of the first author and her professional familiarity with several of the research organisations involved is openly acknowledged, as it could lead to interpretations being affected by biases. To mitigate any biases and raise the validity of the findings, the other authors unfamiliar with the research organisations (SV and MS) were involved in the development of the analysis and interpretations. All direct quotes were sent to interviewees for approval.