

ESSAY

Evaluations in science communication. Current state and future directions

Sophia C. Volk and Mike S. Schäfer

Abstract

Evaluations are becoming more important in science communication. But both science communication practitioners and researchers are not sufficiently utilizing the potential of evaluations yet. In this essay, we first define four requirements for rigorous evaluations of science communication activities and projects. To substantiate our argument, we take stock of the scientific literature, uncover deficiencies in current evaluation practices and identify potential causes. We conclude with laying out how different actors in the field — including science communication practitioners, professional associations, scientific institutions and funding bodies as well as researchers — can contribute to advancing evaluation practices in science communication as well as research on it.

Keywords

AI tools in science communication; Professionalism, professional development and training in science communication; Science and policy-making

Additional Keywords

Evaluation, Impact Measurement, Methods

Received: 22nd June 2024

Accepted: 22nd August 2024

Published: To be published

1 - Introduction

The importance of evaluating science communication activities and projects has been emphasized in recent years and is reflected in a growing number of publications about evaluation practices and methods [e.g., Jensen & Gerber, 2020; Niemann, van den Bogaert & Ziegler, 2019; Weitkamp, 2015]. This is due to several factors: on the one hand, the professional field of science communication has expanded, diversified and professionalized [e.g., Sörensen, Volk, Fürst, Vogler & Schäfer, 2024; Trench, 2017], partly driven by the transformation of digital media environments which led to an increase in channels, formats and target audiences. On the other hand, public and political pressures to demonstrate the societal impact of science and justify research spending have increased [Hill, 2016]. This has led to growing expectations toward scientific institutions and scientists to communicate publicly about their research [King, Steiner, Hobson, Robinson & Clipson, 2015; Palmer & Schibeci, 2014; Rose, Markowitz & Brossard, 2020; SFI, 2020] – and to ensure that such communication has an impact.

Along with the question of impact, the evaluation and measurement of science communication activities and projects is gaining importance. Systematic evaluation is crucial to determine whether the goals of science communication are being met or not, and to improve science communication practices. Typically, evaluations of science communication activities evolve around questions such as: how do (which) science communication formats change people's perceptions, attitudes, or behaviors, if at all? What are the long-term impacts of science communication projects? Which channels are most effective in reaching which target audiences?

This essay critically assesses how science communication activities and practices are typically evaluated. It is aimed at science communication practitioners who conduct evaluations in practice but also at researchers who study the effects of science communication. With both audiences in mind, we argue that the professional field and the research field on science communication do not fully utilize the potential of evaluation: practitioners evaluate too little and not robustly enough, and researchers do not take sufficient advantage of the opportunity to conduct accompanying evaluations of science communication practices.

2 - How should science communication be evaluated?

Evaluation of science communication is broadly understood as the systematic, data-based assessment of communication activities against predefined objectives [Raupp & Osterheider, 2019].

This is different from research on science communication. Both evaluation and research have in common that they use social scientific research methods and quantitative and qualitative measures to capture the effects of science communication. But evaluation differs from research in that it typically assesses science communication activities against organizational or project-specific goals [Raupp, 2017; Volk, 2023; Ziegler, Hedder & Fischer, 2021]. Hence, evaluation is goal-driven, aiming to assess specific programs within predefined target audiences, unlike most research, which is hypothesis-driven and seeks to contribute to generalizable knowledge. It follows that evaluation requires the definition of objectives. Typical objectives of science communication include intermediate objectives like

increasing awareness or interest in scientific topics, improving knowledge, or creating enthusiasm for a scientific discipline, and actual goals like maintaining or building trust in science or influencing behavior [e.g., Besley & Dudo, 2022; Weingart & Joubert, 2019]. For evaluations, formulations of objectives should be “SMART” [e.g., Spicer, 2017]: Specific (as precise and concrete as possible), Measurable (empirically verifiable), Achievable (realistically attainable), Relevant (meaningful and accepted) and Time-bound (time of goal achievement must be specified). Second, it necessitates the definition of target audiences – from children news media to political actors or disinterested publics [Ziegler et al., 2021].

Evaluation can serve different purposes for the practice of science communication [e.g. Jensen & Gerber, 2020; Volk, 2023]: it can demonstrate the “success” of science communication and thereby justify spending, contribute to a learning process and optimization of science communication, serve as a decision-making aid for resource allocation, or function as an early warning system to detect issues or monitor crises. Evaluations can, and should, be an important building block for evidence-based science communication, which is particularly important as “a substantial and thorough concern about the quality of science communication is still lacking in many contexts and institutions” [Pellegrini, 2021, p. 305].

Against this background, many colleagues have discussed what constitutes “good” evaluation, both in the field of science communication and the closely related fields of informal science learning and education. In our view, these discussions can be summarized in four core requirements:

1. **Evaluation should be holistic.** Effects of science communication should be measured and evaluated holistically [Friedman, 2008; Weitkamp, 2015]. Ideally, evaluations should cover entire science communication projects (and not just selected activities), comprise both short-, medium- and long-term effects (and not only immediate or single effects), and include different evaluation objects such as the media or audiences (and not only a single object). A holistic evaluation can be supported by the use of logic models that systematize and visualize how a project or activity leads to a desired result through a sequence of steps. The components of a logic model, hereafter called “stages” [following Raupp & Osterheider, 2019; see also Macnamara & Gregory, 2018], are typically divided into inputs, outputs, outcomes and impacts.¹ Figure 1 illustrates such a logic model, differentiating the stages, the evaluation focus and different evaluation objects at each stage. The *input stage* assesses what resources (e.g. human, financial or time resources) have been invested in a science communication project. The *output stage* asks for the “activities” developed with these resources, e.g. how many exhibitions were created, or social media posts were published, and what their reach was in terms of, for example, website visits, social media impressions, or media coverage. At the *outcome stage*, the question is whether the science communication activities had cognitive, affective or conative effects on the target audiences, for example whether they raised awareness, changed attitudes or behavioral intentions. Importantly, this stage should go beyond desired effects to also measure undesired

1. In the broader literature on evaluation of communication, the term “stages” has become established and is also used by the International Association for Measurement and Evaluation of Communication (<https://amecorg.com/barcelona-principles-3-0-translations/>). In the science communication literature, the components of a logic model are sometimes also labeled “phases” [cf. Pellegrini, 2021] or “elements” [cf. Friedman, 2008].

Stage	INPUTS	OUTPUTS	OUTCOMES	IMPACTS
Focus	What financial and personnel expenses are invested in science communication activities?	How many science communication activities are created and what is the reach of the contact offers?	To what extent do the activities influence the cognitions, attitudes, emotions or behaviors of the target audiences?	What is the long-term contribution or added value for the scientific institution and for society?
Example Methods	Budget analysis Time recording Process analysis	Counting of activities Media resonance analysis Social media tracking Website tracking ...	Surveys / interviews Informal feedback methods Observations Social media analysis ...	Impact stories / case studies Reputation analysis Counting of acquisition of funds / enrolment ...
Object	Institution / Project	Activity Media	Target audiences	Institution / Society

Figure 1. Stages, focus, example methods, and objects of evaluation (source: adapted from Volk [2024] inspired by Deutsche Public Relations Gesellschaft (DPRG) and Internationaler Controller Verein (ICV) [2011]).

and unexpected side effects and dysfunctional or negative consequences. The *impact stage*, then, assesses the long-term value of science communication, for example for a scientific institution (e.g., a university, or museum) or society at large. From an institutional perspective, impact indicators range from student enrolment numbers over the acquisition of new funds or donations to an improved reputation. From a societal perspective, impact can involve various contributions to society, for example, in the field of public health, environment, policy, the economy or practice [e.g., Bornmann, 2013; Jensen, Wong & Reed, 2022]. By examining science communication effects from inputs to outputs, outcomes and impacts, evaluators can relate the resources that were invested to what has been achieved [Raupp & Osterheider, 2019], i.e. making cost-benefit calculations – which is key given that resources are limited and should be invested in activities most suitable to achieve the desired goals. In practice, such logic models should already guide the planning phase and be thought about backwards [e.g., Taplin & Clark, 2012]: wHat is the actual impact we want to achieve? To achieve this, what do we need to change about the opinions and attitudes of the target audiences? Which channels and activities are suitable for this? What resources are required for this?

2. Evaluation should use mixed methods. Holistic evaluations of science communication projects as described above – from inputs to impacts – require the use of different methods and their combination through triangulation [e.g., Niemann et al., 2019; Raupp & Osterheider, 2019]. After all, the results of science communication projects, which often combine social media posts, media relations, and various informative or entertaining events, can typically not be captured with one method alone. The use of multiple or mixed methods to enable holistic evaluation along the input, output, outcome and impact stage is therefore desirable [e.g., Frechtling, 2015; White, 2009].² However, it is important to note that the use of mixed methods does not per se lead to a better evaluation; the selection of methods should always follow the evaluation question [e.g., Funnell & Rogers, 2011] and the quality of

2. This is also part of the recommendations – the so-called Barcelona Principles 3.0 – of the International Association for Measurement and Evaluation of Communication (<https://amecorg.com/barcelona-principles-3-0-translations/>).

the methodological design and implementation naturally remains decisive. In cases where evaluations focus only on selected stages or only parts of science communication activities and projects, individual methods may be sufficient, if they enable the evaluation question to be answered.

Figure 1 outlines typical methods that can be used on each stage. While it may not be feasible in practice to combine multiple methods at each stage, ideally at least one method per stage should be selected, resulting in an overall mixed methods evaluation. Importantly, the combination of methods should not be an end in itself, but should arise from the evaluation question and be oriented towards the utility of the evaluation for the evaluators or the scientific organization [Patton, 1997]. At the output and outcome stage, the entire spectrum of quantitative and qualitative social science research methods (e.g. surveys, interviews, content analysis, observations, web tracking, the use of trace data etc.) is generally suitable for evaluation [Pellegrini, 2021]. In the case of digital media, especially at the stages of outputs and outcomes, a variety of external tool providers (e.g., Meltwater) can be used to collect digital metrics such as clicks, likes, or comments [e.g., Volk & Buhmann, 2023]. In addition, informal feedback methods often used in the field of informal science learning and education like feedback cards or short (exit) interviews with visitors can be used [e.g., Davies & Heath, 2014; Grand & Sardo, 2017]. At the input stage, methods from the business sector such as budget analysis, time tracking, or process analysis are also appropriate [Volk, 2023]. At the impact stage, narrative impact stories or case studies can be used to reconstruct the long-term impact of science communication [e.g., Jensen et al., 2022]. At each stage, different indicators can be used to measure results, including both quantitative indicators (e.g., amount of media coverage, number of participants) and qualitative indicators (e.g., tonality of media coverage, qualitative feedback of participants).

3. **Evaluation should be conducted at multiple points in time.** Ideally, evaluation should not be limited to one-time, post-hoc measurements but occur at multiple points in time and throughout a project [e.g., Pellegrini, 2021]. This is particularly relevant for a robust measurement of changes in the cognitions, attitudes, emotions, or behavior of audiences, which ideally require using pre- and post-test-designs that compare results before and after a science communication activity [Jensen, 2019]. In scholarship on science communication, different types and time points of evaluation are distinguished. Most authors differentiate between formative and summative evaluation [e.g., Pellegrini, 2021], while others additionally speak of process evaluation³ (sometimes also referred to as “monitoring”) [e.g., Macnamara & Gregory, 2018; Valente & Kwan, 2013]. The three types can be related to the typical phases that a project or activity in the field of science communication goes through [Volk & Buhmann, 2023] — from analysis and planning to implementation to evaluation [e.g., Besley & Dudo, 2022], as depicted in Figure 2. Formative, process and summative evaluation complement each other and should ideally be combined in an evaluation design:

- *Formative* evaluation takes place early in a project during the analysis and planning phase, i.e. before implementation begins. It typically asks which

3. This understanding of “process evaluation” differs from other understandings of the term [see e.g. Friedman, 2008; Rossi, Lipsey & Freeman, 2004], which refer to the conduct of process analyses as part of a program evaluation. As a method, process analysis can be used at the input stage (see Figure 1) to examine how efficiently processes and collaborations are running.

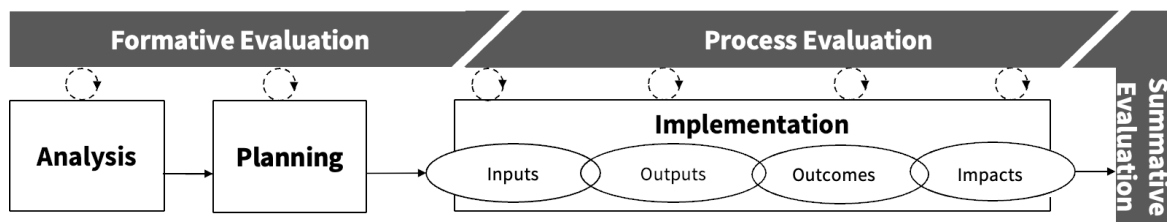


Figure 2. Formative, process, and summative evaluation (source: adapted from Volk and Buhmann [2023]).

messages and channels would be most suitable for the target audiences. In some cases, previous evaluation results (e.g., social media analysis) can be used or usability tests and pretests (e.g., of campaign slogans) can be carried out.

- *Process* evaluation takes place during implementation phase [Macnamara & Gregory, 2018]. This type has gained in importance and attention with the rise of digital communication, media and tools which enable the continuous monitoring and almost real-time optimization of communication [Volk & Buhmann, 2023]. For example, it asks whether the channels are really reaching the desire target audiences during a campaign (e.g., social media analysis). Ideally, this identifies any problems in the implementation process promptly and allows formats to be optimized on the fly.
- *Summative* evaluation takes place retrospectively, i.e. at the end of a project. It asks what effects the activities had on the target audiences (e.g., surveys, feedback forms), compares desired and achieved results and is therefore often understood as a measurement of success.

4. **Evaluation should be suitable to the target audience and format.** Finally, a good evaluation should fit the target audiences and formats to be evaluated [e.g. Campos, 2022; Jensen, 2014]. In science communication, there are numerous target audiences with different characteristics, for example in terms of age (e.g. children, senior citizens), education or scientific literacy, and attitudes towards science (e.g., science-skeptical target groups) [e.g., Humm, Schrögel & Leßmöllmann, 2020; Schäfer, Füchslin, Metag, Kristiansen & Rauchfleisch, 2018]. It is evident that an activity aimed at elementary school children or migrant populations, who cannot yet read or write well or may not own a smartphone, cannot be evaluated via online questionnaires (unless it is filled in by teachers, parents, or translators). Borrowing from the field of informal science learning and education, methods like drawing exercises, paper and pencil questionnaires in simple language, or observations can be used instead [e.g., Campos, 2022]. Moreover, evaluation designs need to suit the formats — from permanent exhibitions over one-off science slams to citizen science apps — which differ in their degree of interactivity and the context of participation. When evaluations build on reactive methods like self-reports or participant observations, they need to ensure not to interrupt or distort participants' experience or discourage their involvement due to expected time constraints of taking part in an evaluation [Grand & Sardo, 2017]. For instance, a highly interactive exhibit at a museum may use quick, unobtrusive and informal feedback methods like a feedback terminal or short in-person interviews [e.g., Davies & Heath, 2014], while a long-term citizen science project might ask involved

citizen scientists to make time for a longer in-depth interviews or an online survey through an app developed for the project For entertaining formats like science slam or performances, evaluations can also be playfully integrated into the format [e.g., Grand & Sardo, 2017], for example, by using an applause meter or live voting. As a result, not every audience and format can be examined with the same evaluation design or traditional methods, but evaluations must always be adapted to their specific context [Spicer, 2017].

3 - Taking stock of evaluation practices – what does the literature say?

But how often, and how well are evaluations in science communication actually done? What is known about how evaluations are carried out in practice and to what extent do they meet requirements of “good” evaluation?

Despite the growing relevance of evaluations in science communication, published empirical research in the field’s English-language journals on the topic is scarce. Although there is a bulk of empirical studies on the effects of science communication, these are of limited practical relevance, as they are often not linked to specific projects and tend to be conducted in laboratory settings rather than the field. Published evaluations of specific science communication and informal science learning and education projects as well as research on the evaluation practices of science communication practitioners are rare and scattered. Such publications mostly stem from Anglophone countries (mainly the US and the UK, but also Australia, Canada, New Zealand or South Africa) as well as German-speaking countries (mainly Germany and Switzerland).⁴ Studies have examined organizational communication from universities or science centers, specific formats such as science festivals, or specific channels such as YouTube. As a result, their findings are often only comparable to a limited extent.

We believe that the different studies found under the keyword of evaluation in science communication scholarship can be categorized into three types (Table 1), when considering the *form* of the evaluation, the study *object*, the role of *researchers*, and the *methodological* focus.

The first type are (1) *external evaluations* of specific science communication activities or projects. Here, scientists typically function as external partners of science communication practitioners and assess whether specific formats work and what effects they have. For example, they may use a survey to question participants of a science festival. In this type of study, the study objects are typically participants in a science communication project, the data are self-reports by those participants, and researchers take the role of providing scientific support or conducting evaluations. An example for this type of study is Rose, Korzekwa, Brossard, Scheufele and Heisler [2017], who conducted an evaluation (as noted in the footnote) of a Wisconsin Science Festival by surveying 183 attendees.

4. This is also due to the selection of international English-language journals; since the Latin American scientific community, for example, publishes in its own regional journals (e.g., Journal of Science Communication – América Latina), it is quite possible that corresponding studies have been overlooked for the purpose of this essay. Future research could conduct a systematic review of journals in different languages to address this limitation and further develop the categorization proposed in Table 1.

Table 1. Categorization of studies on evaluation of science communication.

<i>Type of study</i>	<i>Form of evaluation</i>	<i>Study object</i>	<i>Role of researcher</i>	<i>Methodological focus</i>	<i>Example studies</i>
(1) External evaluations	External evaluation of a science communication activity / project	Participants in a science communication activity / project	Conducting an evaluation and/or providing scientific support to science communication practitioners	Surveys, interviews, feedback forms, observations, drawings	Adhikari et al. [2019]; Boyette and Ramsey [2019]; Campos [2022]; Falk et al. [2016]; Grand and Sardo [2017]; Jensen, Jensen, Duca and Roche [2021]; Kennedy, Jensen and Verbeke [2018]; Niemann, Bittner, Schrögel and Hauser [2020]; Pennisi and Lackey [2018]; Rose, Korzekwa, Brossard, Scheufele and Heisler [2017]
(2) Scientific analyses of evaluation practices	Self-evaluation of a science communication activity / project / department	Science communication practitioners	Collecting and analyzing the narrated evaluation practices	Surveys, interviews	Bühler, Naderer, Koch and Schuster [2007]; Fu, Kannan, Shavelson, Peterson and Kurpius [2016]; Höhn [2011]; Impact Unit [2019]; Impact Unit [2023]*; Phillips, Porticella, Constas and Bonney [2018]; Jensen [2019]; Sörensen, Volk, Fürst, Vogler and Schäfer [2024]; Ziegler, Hedder and Fischer [2021]
(3) Scientific meta-analyses of evaluation reports	Self-evaluation of a science communication activity / project / department	Documents (public / non-public) written by science communication practitioners	Collecting and analyzing the written evaluation practices	Document / content analysis, meta-analysis	Fu, Kannan, Shavelson, Peterson and Kurpius [2016]; Jensen, Wong and Reed [2022]; Volk [2024]; Ziegler and Hedder [2020]*

Note: studies may appear twice if they represent multiple types. Studies marked with * are not peer-reviewed.

The second type are (2) *scientific analyses of evaluation practices*. Here, scientists ask science communication practitioners how they evaluate what works and what effects their activities or projects have. In this type of study, the study objects are science communication practitioners, the data is based on their self-reports about self-evaluation practices, and the researcher takes the role of collecting such statements by means of surveys and interviews and analyzing the narrated evaluation practices. An example for this type is the study by Phillips, Porticella, Conostas and Bonney [2018], who conducted a survey among 99 citizen science practitioners to analyze what learning outcomes they measure.

The third and rarest type are (3) *scientific meta-analyses of evaluation reports*. Here, scientists typically analyze documents written by science communication practitioners (and occasionally communicating researchers) that show what effects their activities or projects have. Typically, they use meta- or content analyses of documents, sometimes including unpublished reports exclusively made available for such an analysis, for instance by funders [e.g., Volk, 2024]. In this study type, the study objects are documents written by science communication practitioners, the data is based on written self-reports about self-evaluation practices, and the researcher takes the role of collecting such documents and analyzing these statements. For example, Fu, Kannan, Shavelson, Peterson and Kurpius [2016] analyzed 36 evaluation reports from the year 2012 publicly posted on the website informal.science.org.

All three types of studies indicate how often evaluations are carried out in practice, in what forms, and how current practices fulfill requirements of “good” evaluation:

1. **Evaluations are not done often and hardly ever holistically.** Although several studies show that evaluations are widely seen as important among science communicators [e.g., Impact Unit, 2019], they are *not* widely done. For example, two surveys on the evaluation practices of science communication practitioners in German-speaking countries in 2019 and 2023 show that 32 to 46 percent of practitioners never or rarely evaluate, while only 36 percent often or always evaluate [Impact Unit, 2019, 2023]. Similarly, a survey by Phillips et al. [2018] among citizen science practitioners in the US and Canada found that only 57% of respondents had ever conducted project evaluations.

But there are indications that — at least in universities — evaluations are done more often nowadays than 15 years ago. Bühler, Naderer, Koch and Schuster [2007] found that in 2007, only 28% of German universities evaluated their PR. A bit later this was still true for less than half of German universities [Höhn, 2011]. However, a more recent study by Sörensen et al. [2024] indicates only 10% of (in this case: Swiss) universities do not conduct any type of evaluation at all. A second finding from the published literature is that evaluations are almost never done holistically, i.e. along the stages of inputs, outputs, outcomes and impact. The qualitative study by Sörensen et al. [2024] indicates that evaluations at Swiss universities often focus on outputs like the number of media releases or created social media posts, or on media coverage or social media impressions achieved. Typically, evaluation focuses on short-term direct outcomes such as social media engagement (e.g., likes or shares), while more meaningful indirect outcomes (e.g., attitudinal changes) are rarely measured. Systematic input and impact measurement are only carried out by few universities [Sörensen et al., 2024].

Medium-term outcomes appear to be, however, often evaluated as part of external evaluations (see e.g., Fogg-Rogers, Bay, Burgess and Purdy [2015] and Rose et al. [2017]). For example, Falk et al. [2016] assessed understanding of and interest in

science and technology, using a quantitative survey of 6,089 adults across 13 countries. Interestingly, unexpected, undesired or dysfunctional effects are hardly ever recorded, so it is often not clear whether science communication activities had any side effects.

- 2. Evaluations mostly use simple methods and easily quantifiable metrics.** Based on the review of published literature, evaluations in science communication seem to focus on a few simple methods rather than on triangulating methods. Meta-analyses of evaluation reports, like the analysis of 36 evaluation reports in the US by Fu et al. [2016], the analysis of 128 Swiss project reports by Volk [2024], or the analysis of 55 evaluation reports in the German-speaking region by Ziegler and Hedder [2020] paint a similar picture: Most evaluations conducted by practitioners or communicating researchers employ relatively simple methods like feedback methods, surveys or interviews. These methods primarily depend on self-reported measures instead of direct observations [for a critique of this approach see Jensen & Lister, 2017]. More complex — and often more costly — methods such as (survey-based) experimental designs with control groups, focus group studies or eye-tracking are rarely used [Ziegler & Hedder, 2020; for an exception see e.g., Niemann, Bittner, Schrögel & Hauser, 2020]. These findings are further corroborated by science and university communicators' self-reports, which indicate that their evaluation practices predominantly focus on web analytics, social media and media monitoring, with less frequent use of comparatively more expensive surveys among target groups like employees or students [e.g., Impact Unit, 2023; Sørensen et al., 2024]. It has to be noted that the first study type — external and scientifically supported evaluations — are often solely based on surveys, albeit of high methodological quality. Some of these studies use complex measures and validated items and questionnaires for measuring interest, knowledge, understanding or motivation through surveys [see e.g., Falk et al., 2016; Phillips et al., 2018; Rose et al., 2017]; behavioral changes are comparatively rarely measured and realized [e.g., Phillips et al., 2018]. Fogg-Rogers et al. [2015] measured audience preferences for different science festival formats in New Zealand and the impact on knowledge acquisition and engagement with on-site surveys among 661 visitors over three years. However, the use of a single method does often not allow for a holistic evaluation of a project from inputs to impacts — for example, even well-designed and executed surveys can often only provide information about the outcome stage. A holistic evaluation usually necessitates a combination of methods in order to assess whether the resources used were proportionate to the effects achieved and whether accompanying communication activities (e.g., social media posts) were also effective. Yet, it seems that different methods are rarely combined — and if they are, it is often only two or three simple methods focused on the output stage, like social media analyses and media monitoring [Sørensen et al., 2024; Volk, 2024]. For example, Adhikari et al. [2019] evaluated the “Pint of Science” format in Thailand by combining fairly simple self-reported surveys among 125 participants with qualitative interviews and focus groups discussions, measuring the motivations for attendance, knowledge, interest, participation and engagement, thus focusing on the outcome stage. Other studies combine simple informal evaluation techniques, which may provide valuable immediate feedback but often come with methodological limitations like low response rates, self-selectivity and restricted insights into more meaningful outcomes. For example, Grand and Sardo [2017] integrated short online questionnaires and “snapshot” interviews with autonomous graffiti walls and feedback cards to evaluate

science festivals in the UK. In a study with children in Portugal, Campos [2022] combined photo-elicitation interviews with drawing exercises. While these examples illustrate the integration of different methods, they also demonstrate that mixed method designs are not of better quality per se and also come with limitations. Beyond the frequent use of simple methods, easily measurable, quantitative metrics seem to dominate, especially in scientific institutions — such as the number of participants or visitors, clicks, likes or subscriptions [Bühler et al., 2007; Impact Unit, 2023; Sörensen et al., 2024; Volk, 2024]. For example, Donhauser and Beck [2021] evaluated videos on the Max Planck Society's YouTube channel and assessed their success by analyzing the number of views and subscriptions, broken down by the age distribution of viewers. Qualitative metrics, in contrast, are rarely reported, for example participants' qualitative feedback [for an exception see Robinson et al., 2017].

3. **Summative evaluations dominate.** Regarding the timing of evaluations, published studies suggest that most evaluations use one-time measurements after a project or activity, making robust statements about effects difficult. Both self-reports of science communication practitioners [Impact Unit, 2023; Sörensen et al., 2024] as well as meta-analyses of evaluation reports [Volk, 2024; Ziegler & Hedder, 2020] show that summative evaluations clearly dominate, with the majority of practitioners collecting data only once at the end of a project. Fewer studies make use of process evaluations (often related to social media), and formative evaluations hardly ever take place. More demanding pre-post-test designs — with before and after measurements — are almost exclusively found in external and scientifically supported evaluations. At the example of a South African MOOC and a traveling “World Biotech Tour”, Jensen [2019], for example, reports on the use of pre-, mid-, and post-test surveys with Likert-scales to measure participants' understanding, experiences and attitudes through repeated measures over time. Rose et al. [2017] assessed the effects of attending a panel at the “World Science Festival” in the US on perceived knowledge, risk perceptions, benefit perceptions, and moral and ethical views using pre- and post-test surveys. In most published studies, however, evaluations take place only once and directly after participation, for example while exiting a museum. An exception is, for example, a study by Pennisi and Lackey [2018] in the US that conducted a multiyear evaluation for an annual science festival, including a follow-up survey 6 months after the festival to track knowledge and behavior change. In many evaluations, however, it remains unclear whether effects are stable or whether there may be time-delayed effects that could not be measured with the design.

4. **Evaluations rarely reveal who was reached and if the communication was suitable.** Strikingly, the review of published evaluation studies reveals that the reached audiences are often not really known in science communication projects, making it unclear who is reached. Often, only basic demographics are known about the reached audiences, i.e. age and gender [Donhauser & Beck, 2021; Fogg-Rogers et al., 2015], race [in the US, e.g., Boyette & Ramsey, 2019], and in some cases levels of education and income [Adhikari et al., 2019; Boyette & Ramsey, 2019]. If projects target the “broader population”, evaluations rarely compare the reached audience with the general population [for exceptions see, e.g., Jensen, Jensen, Duca & Roche, 2021; Kennedy, Jensen & Verbeke, 2018], so it is unclear how representative the reached audience is. Since many countries regularly publish official statistics on the demographic profile of specific regions, census data is often available as a comparative reference point and

there is no need to collect such data during an evaluation. In a few evaluations, for example the study by Rose et al. [2017], such data was collected — presumably in a costly and time-consuming way — by means of a state-representative population survey in the US in order to compare residents with attendees of a science festival. The comparison revealed that attendees were more educated, liberal, and had higher trust in scientists than residents, and also pointed to a strong self-selection bias of participants. Overall, science-related attitudes are surveyed rather rarely — and when they are, especially in external evaluations, it turns out that the target audiences addressed and reached often already have a high level of interest in science anyway [Volk, 2024], as demonstrated for instance for science festival attendees in the UK by Kennedy et al. [2018]. In a study comparing science centers across 13 countries, Falk et al. [2016] found that individuals who visited science centers self-reported significantly higher levels of understanding, interest, curiosity and participation in science-related activities compared to non-visitors, even after accounting for income and education level. This naturally raises the question of what types of audiences are not being reached, and whether some are systematically underserved [Humm et al., 2020]. Despite existing research on disengaged audiences and its relevance for practice, in the published studies, there is however little reflection on the question of who was not reached or whether the evaluation was appropriate for the target audiences, the format, and the context [e.g., Campos, 2022; Grand & Sardo, 2017].

Overall, published studies show significant shortcomings in evaluation practices around science communication. Scholarship suggests that these deficiencies arise both from the science communicators themselves, the organizations where evaluation takes place, and the broader professional field: first, practitioners often lack financial resources or time for evaluation, and partly also methodological skills and tools for measuring communication effects [Jensen, 2014; King et al., 2015; Sørensen et al., 2024; Weitkamp, 2015; Ziegler et al., 2021]. Second, the fact that evaluation is hardly conducted holistically, and sometimes not at all, may also be due to a lack of demand for such studies from organizational leaders and funding agencies [e.g., Banse, Panzer & Fischer, 2024; Sørensen et al., 2024]. Third, the professional field lacks agreed, standardized metrics and normative pressures for evaluation [e.g., Banse et al., 2024; Ziegler et al., 2021; see also Volk, 2024].

4 - Evaluating science communication — the way forward

The published literature shows that evaluations of science communication are not (yet) widespread, and when they are carried out, they are usually based on simple methods, selective indicators and one-off measurements. But evaluations will presumably be increasingly required in the future. We believe that several requirements need to be addressed to move both evaluation practices and evaluation research forward. In our view, this can be condensed in seven points:

- *More, and more importantly, better evaluations are needed.* While this requirement is certainly not new and has already been emphasized by previous essays [e.g., Jensen, 2014, 2019], it seems necessary to reiterate it in light of the identified gaps. Ideally, science communicators should conduct elaborate evaluations along inputs, outputs, outcomes, impacts, with a mix of social scientific research methods and pre-post test designs. However, with ongoing resource and time restrictions, we realize that this is

and will often not be feasible in every project. Therefore, science communicators may need to be selective and conduct systematic evaluations as described above only for specific, strategically relevant projects or in larger intervals. But importantly, if they evaluate, evaluations should be well designed and robust. Better evaluations can be achieved by devising evaluation plans at the outset of a project [Spicer, 2017] and relying on already validated instruments (e.g., for measuring knowledge, trust etc. in surveys) [Jensen, 2019]. Moreover, they involve not only positive, intended effects, but should also assess potential unexpected, dysfunctional effects, and audiences that may not have been reached [Jensen, 2015]. Better and more evaluation will require building up solid methodological expertise among practitioners to conduct robust self-evaluations. In addition, greater collaboration with researchers to conduct independent external evaluations would be useful and have the added benefit of researchers having the incentive to publish from the data and incorporate the results into the scientific evidence base.

- *More demand for and support of evaluations is needed by leaders in scientific organizations as well as by funding institutions and foundations* in the long run. Universities, science centers, museums, and funders alike have a duty to demand evaluations to learn how well money was spent on science communication. Hence, they should support better evaluations by reallocating or setting aside resources for evaluation in the future [Banse et al., 2024]. More valorization, i.e. symbolic appreciation for evaluation activities, is also needed — as are sanctions for lack of evaluations, for example in final project reports submitted to funders. Larger scientific organizations should consider establishing designated positions for coordinating evaluation activities and developing tools for evaluating science communication that are shared with colleagues [e.g., Sørensen et al., 2024]. Science communicators in turn should request and plan a separate budget and personnel resources for evaluation and should compile evaluation reports for organizational leadership that demonstrate how science communication adds to strategic goals [Spicer, 2017].
- *Shared standards for evaluations* are needed. With few common indicators being used in the field [Banse et al., 2024], evaluations of different projects, formats or organizations are often not comparable, presenting an obstacle to learning from others and broader benchmarking [Volk, 2024; Davies & Heath, 2014]. Developing and agreeing on a set of common standards is needed to address growing expectations of policymakers and stakeholders to demonstrate impact on the one hand, but also to avoid selective, less robust and non-representative evaluations on the other hand. Both national as well as international professional associations should initiate a collaborative process of harmonizing and negotiating standards for evaluation of science communication, drawing inspiration from similar endeavors initiated, for example, by the International Association for Measurement and Evaluation of Communication (AMEC).⁵ Such standards should neither stifle nor suffocate flexibility and creativity but aim at increasing comparability of evaluations and offering suitable indicators for demonstrating outputs, outcomes, and impacts. Time is of the essence here: We think an initiative to define them *within* the field is necessary before standards are imposed from *outside* the field, in a process science communicators and researchers have no say in.

5. AMEC [2016]. AMEC Integrated Evaluation Framework. <https://amecorg.com/amecframework/>.

- *Refined impact measures* are needed that capture non-academic, long-term contributions or values of science communication. Since the term impact is understood and defined differently [Bornmann, 2013; Watermeyer & Chubb, 2019], a common understanding of such impact must first be developed together with professional associations and funding bodies. Suitable methods, such as narrative impact statements or impact case studies, potentially involving key stakeholders or external evaluators [Hill, 2016], and suitable indicators considering different types of impact, for example, on the environment or politics, should be developed [Jensen et al., 2022]. Evaluation periods need to be adapted, as impacts often occur with considerable time lags. It must be clear to everyone involved that more rigorous impact measurement will be demanding, resource- and time-consuming [Jensen, 2019]. Since practitioners may not be able to empirically trace cause-effect relationships back to science communication, they will likely need to agree on logically plausible pathways to impact together with funders and scientific institutions. Reflection on impact indicators is urgently needed – both in science communication and in academia more generally – to counteract inflationary impact statements and unrealistic impact expectations [King et al., 2015].
- *Capacity building* is needed. Given the limited resources and different levels of methodological expertise among many science communicators [Jensen, 2015], developing and sharing evaluation guides and survey or interview templates with instructions is desirable so that practitioners can flexibly put these together themselves and use them without much effort rather than reinventing the wheel. Professional associations in science or university communication as well as funding bodies should (continue to) invest in capacity building and offer platforms for sharing hands-on instructions and best practices. They can follow the example of initiatives like the “Impact Unit”⁶ in Germany and the World Initiative for Science Evaluation (SciWise)⁷ in the US, or funders like the Commonwealth Scientific and Industrial Research Organization (CSIRO)⁸ in Australia, the UK Research and Innovation (UKRI)⁹ in England, or the Science Foundation Ireland (SFI)¹⁰ in Ireland. They should also foster an understanding of evaluation as an opportunity to improve and optimize science communication by learning from mistakes [Jensen, 2019]. More methods training and networking opportunities for practitioners interested in evaluation are needed. Researchers could be involved, for example, in continuing education courses that impart methodological expertise.
- *Responsible evaluation must be key*. This has always been true and includes, for example, being responsible to participants, from protecting their privacy and confidentiality, over following ethical standards during evaluation to being aware of and accountable for effects of the evaluation on participants. It also includes being responsible to the scientific organization, funder, or project, for example, by using evaluation resources wisely and efficiently, by ensuring that evaluations provide useful

6. Impact Unit [2023]. How-To-Reihe Wisskomm evaluieren. Wissenschaft im Dialog. <https://impactunit.de/tools/>.

7. <https://www.sciwise.org/en/mission>.

8. Commonwealth Scientific and Industrial Research Organisation. [2020]. Impact evaluation guide. <https://www.csiro.au/en/about/Corporate-governance/Ensuring-our-impact/Evaluating-our-impact>.

9. UK Research and Innovation. [2020]. Evaluation: practical guidelines. <https://www.ukri.org/publications/evaluation-practical-guidelines/>.

10. Science Foundation Ireland. [2015]. Evaluation toolkit. <https://www.sfi.ie/engagement/guidance/>.

information, and by making evaluation processes and (unwanted or unmet) results transparent. Responsible evaluation will be increasingly important given the expanding availability of digital trace data and new technologies like generative artificial intelligence, which will likely have a massive impact on science communication [Schäfer, 2023] and its evaluation — especially regarding the collection, analysis, and use of evaluation data [Volk & Buhmann, 2023]. AI-powered tools can be used, for example, for the automated collection of digital data (e.g., through scraping) or automated transcriptions (e.g., of audio data), for real-time data analysis (e.g., machine learning) or visualization of data (e.g., dashboards), as well as for real-time optimization of communication (e.g., through message distribution) or prognostic evaluation (e.g., predictive analytics) [Volk & Buhmann, 2023]. Professional associations and scientific institutions will need to develop codes and guidelines and raise awareness for ethical and responsible evaluation.

- *Open evaluation data is needed* whenever possible. Since there are few incentives for science communication practitioners or evaluators to publish results from evaluations in peer-reviewed journals [Fu et al., 2016], evaluation reports are often not publicly accessible. Moreover, only 19% of surveyed science communication practitioners report that they make evaluation data available for research purposes [Impact Unit, 2023]. In principle, it would be desirable to make the results of evaluations, including the instruments, sample descriptions, and descriptive data publicly accessible both for practical and research purposes. Open evaluation data would enable science communicators to draw comparisons between projects, institutions, etc. and learn about realistically achievable science communication effects [e.g., Pellegrini, 2021]. Of course, with open data, new problems of confidentiality and anonymization arise, but more publicly available evaluation reports will allow to build up a better evidence base for science communication. For instance, researchers could conduct meta-analyses of evaluation data and publish results [see Volk, 2024; Jensen et al., 2022]. Moreover, researchers could use evaluation reports to engage in a meta-critical reflection on a particular evaluation design in order to stimulate and promote improvements [e.g., Jensen, 2015; Jensen & Lister, 2017]. Both scientific institutions as well as funding institutions and foundations should make such data available, where possible [Davies & Heath, 2014].

In general, we emphasize that systematic, rigorous evaluation of science communication practices and activities needs to be taken more seriously. We call for the relevant actors in science communication — from practitioners to professional associations over scientific institutions and funding bodies all the way to researchers — to reflect jointly on ways to improve evaluation practices. We hope that the seven requirements outlined above provide a basis for such a discussion and reflection.

Acknowledgments

This essay summarizes a keynote by the authors at the “Evaluation in Science Communication” conference at LMU Munich, which was organized by the Munich Science Communication Lab (MSCL) and the Impact Unit in March 2024.

References

- Adhikari, B., Hlaing, P. H., Robinson, M. T., Ruecker, A., Tan, N. H., Jatupornpimol, N., ... Cheah, P. Y. (2019). Evaluation of the Pint of Science festival in Thailand. *PLoS ONE* 14 (7), e0219983. doi:10.1371/journal.pone.0219983
- Banse, L., Panzer, J. & Fischer, L. (2024). *Hürden und Herausforderungen effektiver Evaluationen in der Wissenschaftskommunikation. Erkenntnisse einer qualitativen Untersuchung mit Praktiker*innen der Wissenschaftskommunikation*. Wissenschaft im Dialog. Berlin, Germany. Retrieved from https://impactunit.de/wp-content/uploads/2024/05/Analyse_ImpactUnit_Huerden-und-Herausforderungen.pdf
- Besley, J. C. & Dudo, A. (2022). *Strategic science communication: a guide to setting the right objectives for more effective public engagement*. Baltimore, MD, U.S.A.: Johns Hopkins University Press. doi:10.56021/9781421444215
- Bornmann, L. (2013). What is societal impact of research and how can it be assessed? A literature survey. *Journal of the American Society for Information Science and Technology* 64 (2), 217–233. doi:10.1002/asi.22803
- Boyette, T. & Ramsey, J. (2019). Does the messenger matter? Studying the impacts of scientists and engineers interacting with public audiences at science festival events. *JCOM* 18 (02), A02. doi:10.22323/2.18020202
- Bühler, H., Naderer, G., Koch, R. & Schuster, C. (2007). *Hochschul-PR in Deutschland: Ziele, Strategien und Perspektiven* [Public relations in German higher education institutions: goals, strategies, and perspectives]. Deutscher Universitätsverlag Wiesbaden. doi:10.1007/978-3-8350-9148-1
- Campos, R. (2022). Including younger children in science-related issues using participatory and collaborative strategies: a pilot project on urban biodiversity. *JCOM* 21 (02), N07. doi:10.22323/2.21020807
- Davies, M. & Heath, C. (2014). “Good” organisational reasons for “ineffectual” research: evaluating summative evaluation of museums and galleries. *Cultural Trends* 23 (1), 57–69. doi:10.1080/09548963.2014.862002
- Deutsche Public Relations Gesellschaft (DPRG) and Internationaler Controller Verein (ICV) (2011). *Communication controlling: how to maximize and demonstrate the value creation through communication*. Retrieved from http://www.communicationcontrolling.de/fileadmin/communicationcontrolling/sonst_files/Position_paper_DPRG_ICV_2011_english.pdf
- Donhauser, D. & Beck, C. (2021). Pushing the Max Planck YouTube channel with the help of influencers. *Frontiers in Communication* 5, 601168. doi:10.3389/fcomm.2020.601168
- Falk, J. H., Dierking, L. D., Swanger, L. P., Staus, N., Back, M., Barriault, C., ... Verheyden, P. (2016). Correlating science center use with adult science literacy: an international, cross-institutional study. *Science Education* 100 (5), 849–876. doi:10.1002/sce.21225
- Fogg-Rogers, L., Bay, J. L., Burgess, H. & Purdy, S. C. (2015). “Knowledge is power”: a mixed-methods study exploring adult audience preferences for engagement and learning formats over 3 years of a health science festival. *Science Communication* 37 (4), 419–451. doi:10.1177/1075547015585006
- Frechtling, J. A. (2015). *Logic modeling methods in program evaluation*. New York, NY, U.S.A.: Wiley.
- Friedman, A. (2008). *Framework for evaluating impacts of informal science education projects. Report from a National Science Foundation workshop*. The National Science Foundation. Retrieved from https://informalscience.org/wp-content/uploads/2022/05/Eval_Framework.pdf
- Fu, A. C., Kannan, A., Shavelson, R. J., Peterson, L. & Kurpius, A. (2016). Room for rigor: designs and methods in informal science education evaluation. *Visitor Studies* 19 (1), 12–38. doi:10.1080/10645578.2016.1144025

- Funnell, S. C. & Rogers, P. J. (2011). *Purposeful program theory: effective use of theories of change and logic models*. San Francisco, CA, U.S.A.: Jossey Bass.
- Grand, A. & Sardo, A. M. (2017). What works in the field? Evaluating informal science events. *Frontiers in Communication* 2, 22. doi:[10.3389/fcomm.2017.00022](https://doi.org/10.3389/fcomm.2017.00022)
- Hill, S. (2016). Assessing (for) impact: future assessment of the societal impact of research. *Palgrave Communications* 2 (1), 16073. doi:[10.1057/palcomms.2016.73](https://doi.org/10.1057/palcomms.2016.73)
- Höhn, T. D. (2011). *Wissenschafts-PR: eine Studie zur Öffentlichkeitsarbeit von Hochschulen und außeruniversitären Forschungseinrichtungen* [Science PR: a study on public relations at higher education institutions and non-university research organizations]. Konstanz, Germany: UVK.
- Humm, C., Schrögel, P. & Leßmöllmann, A. (2020). Feeling left out: underserved audiences in science communication. *Media and Communication* 8 (1), 164–176. doi:[10.17645/mac.v8i1.2480](https://doi.org/10.17645/mac.v8i1.2480)
- Impact Unit (2019). *Evaluation and impact in science communication. Results of a community survey November/December 2019*. Wissenschaft im Dialog. Berlin, Germany. Retrieved from https://impactunit.de/wp-content/uploads/2021/08/Summary_Community_Survey.pdf
- Impact Unit (2023). *Evaluation and impact in science communication. Results of a community survey November/December 2023*. Wissenschaft im Dialog. Berlin, Germany. Retrieved from https://impactunit.de/wp-content/uploads/2024/04/WiD_ImpactUnit_CommunitySurvey2023.pdf
- Jensen, A. M., Jensen, E. A., Duca, E. & Roche, J. (2021). Investigating diversity in European audiences for public engagement with research: who attends European Researchers' Night in Ireland, the UK and Malta? *PLoS ONE* 16 (7), e0252854. doi:[10.1371/journal.pone.0252854](https://doi.org/10.1371/journal.pone.0252854)
- Jensen, E. A. (2014). The problems with science communication evaluation. *JCOM* 13 (01), C04. doi:[10.22323/2.13010304](https://doi.org/10.22323/2.13010304)
- Jensen, E. A. (2015). Highlighting the value of impact evaluation: enhancing informal science learning and public engagement theory and practice. *JCOM* 14 (03), Y05. doi:[10.22323/2.14030405](https://doi.org/10.22323/2.14030405)
- Jensen, E. A. (2019). Why impact evaluation matters in science communication: or, advancing the science of science communication. In P. Weingart, M. Joubert & B. Falade (Eds.), *Science communication in South Africa: reflections on current issues* (pp. 213–228). doi:[10.5281/zenodo.3557213](https://doi.org/10.5281/zenodo.3557213)
- Jensen, E. A. & Gerber, A. (2020). Evidence-based science communication. *Frontiers in Communication* 4, 78. doi:[10.3389/fcomm.2019.00078](https://doi.org/10.3389/fcomm.2019.00078)
- Jensen, E. A. & Lister, T. (2017). The challenges of 'measuring long-term impacts of a science center on its community': a methodological review. In P. G. Patrick (Ed.), *Preparing informal science educators: perspectives from science communication and education* (pp. 243–259). doi:[10.1007/978-3-319-50398-1_13](https://doi.org/10.1007/978-3-319-50398-1_13)
- Jensen, E. A., Wong, P. & Reed, M. S. (2022). How research data deliver non-academic impacts: a secondary analysis of UK Research Excellence Framework impact case studies. *PLoS ONE* 17 (3), e0264914. doi:[10.1371/journal.pone.0264914](https://doi.org/10.1371/journal.pone.0264914)
- Kennedy, E. B., Jensen, E. A. & Verbeke, M. (2018). Preaching to the scientifically converted: evaluating inclusivity in science festival audiences. *International Journal of Science Education, Part B* 8 (1), 14–21. doi:[10.1080/21548455.2017.1371356](https://doi.org/10.1080/21548455.2017.1371356)
- King, H., Steiner, K., Hobson, M., Robinson, A. & Clipson, H. (2015). Highlighting the value of evidence-based evaluation: pushing back on demands for 'impact'. *JCOM* 14 (02), A02. doi:[10.22323/2.14020202](https://doi.org/10.22323/2.14020202)
- Macnamara, J. & Gregory, A. (2018). Expanding evaluation to progress strategic communication: beyond message tracking to open listening. *International Journal of Strategic Communication* 12 (4), 469–486. doi:[10.1080/1553118x.2018.1450255](https://doi.org/10.1080/1553118x.2018.1450255)

- Niemann, P., van den Bogaert, V. & Ziegler, R. (2019). *Evaluationsmethoden der Wissenschaftskommunikation*. Springer Fachmedien Wiesbaden.
doi:[10.1007/978-3-658-39582-7](https://doi.org/10.1007/978-3-658-39582-7)
- Niemann, P., Bittner, L., Schrögel, P. & Hauser, C. (2020). Science slams as edutainment: a reception study. *Media and Communication* 8 (1), 177–190. doi:[10.17645/mac.v8i1.2459](https://doi.org/10.17645/mac.v8i1.2459)
- Palmer, S. E. & Schibeci, R. A. (2014). What conceptions of science communication are espoused by science research funding bodies? *Public Understanding of Science* 23 (5), 511–527.
doi:[10.1177/0963662512455295](https://doi.org/10.1177/0963662512455295)
- Patton, M. Q. (1997). *Utilization-focused evaluation: the new century text* (3rd ed.). Thousand Oaks, CA, U.S.A.: SAGE Publications.
- Pellegrini, G. (2021). Evaluating science communication: concepts and tools for realistic assessment. In M. Bucchi & B. Trench (Eds.), *Routledge handbook of public communication of science and technology* (3rd ed.). doi:[10.4324/9781003039242](https://doi.org/10.4324/9781003039242)
- Pennisi, L. & Lackey, N. Q. (2018). A multiyear evaluation of the NaturePalooza Science Festival. *Journal of Extension* 56 (7), 8. doi:[10.34068/joe.56.07.08](https://doi.org/10.34068/joe.56.07.08)
- Phillips, T., Porticella, N., Conostas, M. & Bonney, R. (2018). A framework for articulating and measuring individual learning outcomes from participation in citizen science. *Citizen Science: Theory and Practice* 3 (2), 3. doi:[10.5334/cstp.126](https://doi.org/10.5334/cstp.126)
- Raupp, J. (2017). Strategische Wissenschaftskommunikation [Strategic science communication]. In H. Bonfadelli, B. Fähnrich, C. Lühje, J. Milde, M. Rhomberg & M. S. Schäfer (Eds.), *Forschungsfeld Wissenschaftskommunikation* (pp. 143–163).
doi:[10.1007/978-3-658-12898-2_8](https://doi.org/10.1007/978-3-658-12898-2_8)
- Raupp, J. & Osterheider, A. (2019). Evaluation von Hochschulkommunikation [Evaluation of higher education communication]. In B. Fähnrich, J. Metag, S. Post & M. S. Schäfer (Eds.), *Forschungsfeld Hochschulkommunikation* (pp. 181–205). doi:[10.1007/978-3-658-22409-7_9](https://doi.org/10.1007/978-3-658-22409-7_9)
- Robinson, M. T., Jatupornpimol, N., Sachaphimukh, S., Lönnkvist, M., Ruecker, A. & Cheah, P. Y. (2017). The first Pint of Science Festival in Asia. *Science Communication* 39 (6), 810–820.
doi:[10.1177/1075547017739907](https://doi.org/10.1177/1075547017739907)
- Rose, K. M., Korzekwa, K., Brossard, D., Scheufele, D. A. & Heisler, L. (2017). Engaging the public at a science festival: findings from a panel on human gene editing. *Science Communication* 39 (2), 250–277. doi:[10.1177/1075547017697981](https://doi.org/10.1177/1075547017697981)
- Rose, K. M., Markowitz, E. M. & Brossard, D. (2020). Scientists' incentives and attitudes toward public communication. *Proceedings of the National Academy of Sciences* 117 (3), 1274–1276.
doi:[10.1073/pnas.1916740117](https://doi.org/10.1073/pnas.1916740117)
- Rossi, P. H., Lipsey, M. W. & Freeman, H. E. (2004). *Evaluation: a systematic approach* (7th ed.). Thousand Oaks, CA, U.S.A.: SAGE Publications.
- Schäfer, M. S. (2023). The Notorious GPT: science communication in the age of artificial intelligence. *JCOM* 22 (02), Y02. doi:[10.22323/2.22020402](https://doi.org/10.22323/2.22020402)
- Schäfer, M. S., Fuchslin, T., Metag, J., Kristiansen, S. & Rauchfleisch, A. (2018). The different audiences of science communication: a segmentation analysis of the Swiss population's perceptions of science and their information and media use patterns. *Public Understanding of Science* 27 (7), 836–856. doi:[10.1177/0963662517752886](https://doi.org/10.1177/0963662517752886)
- SFI (2020). *Science in Ireland Barometer 2020. Research report*. Science Foundation Ireland. Dublin, Ireland. Retrieved from <https://www.sfi.ie/engagement/barometer/>
- Sörensen, I., Volk, S., Fürst, S., Vogler, D. & Schäfer, M. (2024). "It's not so easy to measure impact": A qualitative analysis of how universities measure and evaluate their communication. *International Journal of Strategic Communication* 18 (2), 93–114.
doi:[10.1080/1553118X.2024.2317771](https://doi.org/10.1080/1553118X.2024.2317771)

- Spicer, S. (2017). The nuts and bolts of evaluating science communication activities. *Seminars in Cell & Developmental Biology* 70, 17–25. doi:[10.1016/j.semcdb.2017.08.026](https://doi.org/10.1016/j.semcdb.2017.08.026)
- Taplin, D. H. & Clark, H. (2012). *Theory of change basics: a primer on theory of change*. ActKnowledge. New York, NY, U.S.A. Retrieved from https://www.theoryofchange.org/wp-content/uploads/toco_library/pdf/ToCBasics.pdf
- Trench, B. (2017). Universities, science communication and professionalism. *JCOM* 16 (05), C02. doi:[10.22323/2.16050302](https://doi.org/10.22323/2.16050302)
- Valente, T. W. & Kwan, P. P. (2013). Evaluating communication campaigns. In R. E. Rice & C. K. Atkin (Eds.), *Public communication campaigns* (4th ed., pp. 83–97). Thousand Oaks, CA, U.S.A.: SAGE Publications.
- Volk, S. (2023). Evaluation der Wissenschaftskommunikation: Modelle, Stufen, Methoden. In P. Niemann, V. van den Bogaert & R. Ziegler (Eds.), *Evaluationsmethoden der Wissenschaftskommunikation* (pp. 33–49). doi:[10.1007/978-3-658-39582-7_3](https://doi.org/10.1007/978-3-658-39582-7_3)
- Volk, S. (2024). Assessing the outputs, outcomes and impacts of science communication: A quantitative content analysis of 128 science communication projects. *Science Communication*. doi:[10.1177/10755470241253858](https://doi.org/10.1177/10755470241253858)
- Volk, S. & Buhmann, A. (2023). Digital corporate communication and measurement and evaluation. In V. Luoma-aho & M. Badham (Eds.), *Handbook on Digital Corporate Communication* (pp. 118–133). doi:[10.4337/9781802201963.00018](https://doi.org/10.4337/9781802201963.00018)
- Watermeyer, R. & Chubb, J. (2019). Evaluating ‘impact’ in the UK’s Research Excellence Framework (REF): liminality, looseness and new modalities of scholarly distinction. *Studies in Higher Education* 44 (9), 1554–1566. doi:[10.1080/03075079.2018.1455082](https://doi.org/10.1080/03075079.2018.1455082)
- Weingart, P. & Joubert, M. (2019). The conflation of motives of science communication — causes, consequences, remedies. *JCOM* 18 (03), Y01. doi:[10.22323/2.18030401](https://doi.org/10.22323/2.18030401)
- Weitkamp, E. (2015). Between ambition and evidence. *JCOM* 14 (02), E. doi:[10.22323/2.14020501](https://doi.org/10.22323/2.14020501)
- White, H. (2009). Theory-based impact evaluation: principles and practice. *Journal of Development Effectiveness* 1 (3), 271–284. doi:[10.1080/19439340903114628](https://doi.org/10.1080/19439340903114628)
- Ziegler, R. & Hedder, I. R. (2020). *Evaluationspraktiken in der Wissenschaftskommunikation — eine Betrachtung veröffentlichter Evaluationsberichte im deutschsprachigen Raum*. Wissenschaft im Dialog. Berlin, Germany. Retrieved from https://impactunit.de/wp-content/uploads/2021/08/Erg_ebnisbericht_Evaluationspraktiken_der_Wisskomm.pdf
- Ziegler, R., Hedder, I. R. & Fischer, L. (2021). Evaluation of science communication: current practices, challenges, and future implications. *Frontiers in Communication* 6, 669744. doi:[10.3389/fcomm.2021.669744](https://doi.org/10.3389/fcomm.2021.669744)

About the authors

Dr. Sophia Charlotte Volk is a Senior Research and Teaching Associate at the Department of Communication and Media Research (IKMZ) at the University of Zurich (Switzerland). Previously, she was a Research Associate at the Chair of Strategic Communication at Leipzig University (Germany). Her research interests include science and university communication, evaluation and impact measurement, strategic communication, digital media environments and technologies like artificial intelligence, and international comparative research.

✉ s.volk@ikmz.uzh.ch

✉ [@sophia_c_volk](https://twitter.com/sophia_c_volk)

Dr. Mike S. Schäfer is a Full Professor of Science Communication and Head of Department at the Department of Communication and Media Research (IKMZ) of the University of Zurich. He is also director of the university's Center for Higher Education and Science Studies (CHESS). His research focuses on public communication and public perceptions and attitudes towards science and science-related topics, currently with a focus on climate change and artificial intelligence.

✉ m.schaefer@ikmz.uzh.ch

✕ [@mss7676](https://twitter.com/mss7676)

How to cite

Volk, S. C. and Schäfer, M. S. (2024). 'Evaluations in science communication. Current state and future directions'. *JCOM* 23(06), Y01. <https://doi.org/10.22323/2.23060401>.



© The Author(s). This article is licensed under the terms of the Creative Commons Attribution – NonCommercial – NoDerivativeWorks 4.0 License. All rights for Text and Data Mining, AI training, and similar technologies for commercial purposes, are reserved. ISSN 1824-2049. Published by SISSA Medialab. jcom.sissa.it