



Evaluationen in der Wissenschaftskommunikation

Ein Zwischenbericht zu theoretischen und konzeptionellen Grundlagen

Teil der aktuellen Förderung von Wissenschaftskommunikation.de ist auch die Entwicklung von Evaluations-Toolkits zur aussagekräftigen und wirkungsorientierten Evaluation von verschiedenen Formaten der externen, nicht medienvermittelten Wissenschaftskommunikation. Grundlage dieser Entwicklung ist unter anderem eine Desktop-Recherche zu den theoretischen und konzeptionellen Grundlagen und dem Stand der Forschung zu Evaluationen in der Wissenschaftskommunikation. Der folgende Bericht beinhaltet die Ergebnisse dieser Recherche, welche im Jahr 2019 durch Projektmitarbeitende bei Wissenschaft im Dialog durchgeführt wurde.

Erscheinungsdatum: 30.09.2020



Inhalt

Einleitung

Relevanz des Themas „Wirkungen und Evaluation in der Wissenschaftskommunikation“	2
Entwicklungen und Zukunftsperspektiven der Evaluationspraxis in der Wissenschaftskommunikation	3
Die Rolle von Wissenschaft im Dialog	4

Methodik der Recherche	5
------------------------------	---

Theorie zur Evaluation der Wissenschaftskommunikation

Zielebenen von Projekten in der Wissenschaftskommunikation.....	6
Wirkungen von Wissenschaftskommunikation	10
Ziele, Motivationen und Arten von Evaluationen.....	12
Objekte von Evaluationen	14
Evaluationsdesigns und Erhebungsmethoden.....	15

Fazit und Ausblick: Erkenntnisse und Perspektiven für die Evaluationsplanung ..	19
---	----

Literaturverzeichnis	21
----------------------------	----



Einleitung

Relevanz des Themas „Wirkungen und Evaluation in der Wissenschaftskommunikation“

Der öffentliche Austausch über Forschung und die Rolle von Wissenschaft in einer demokratischen Gesellschaft werden derzeit sowohl im Wissenschaftssystem als auch im politischen Raum außergewöhnlich intensiv diskutiert und der (externen) Wissenschaftskommunikation eine zunehmende Bedeutung zugesprochen. Hintergrund sind öffentliche Diskussionen um einen möglichen Vertrauensverlust der Öffentlichkeit in Forschungsergebnisse und wissenschaftliche Expertise sowie die Wahrnehmung, dass auch in Deutschland die öffentliche Finanzierung und Förderung sowie die Freiheit von Wissenschaft und Forschung nicht mehr als sakrosankt gelten. Damit einher geht eine verstärkte Auseinandersetzung um die richtigen Ziele von und Strategien in der Wissenschaftskommunikation. Dies findet vor dem Hintergrund statt, dass sich die Wissenschaftskommunikation – ihre Formate wie auch Rollen und Akteure – seit der PUSH-Bewegung Anfang der 2000er Jahre enorm diversifiziert, pluralisiert und bezüglich der Umsetzung von Maßnahmen und Aktivitäten sowie Durchführung von Projekten professionalisiert hat.

Parallel hierzu lässt sich im internationalen Kontext feststellen, dass sich besonders im englischsprachigen Raum, aber auch in Skandinavien staatliche wie gemeinnützige Akteure der Wissenschaftskommunikation seit einigen Jahren verstärkt Fragen der Wirkung, Evaluation und Qualität von Wissenschaftskommunikation widmen. Die entsprechenden Organisationen blicken dabei qualitäts- und wirkungsorientiert auf ihre eigenen Tätigkeiten, wie zum Beispiel die *British Science Association* mit ihrer jährlichen Veröffentlichung von Impact-Reports (British Science Association, 2020) oder auch die *Science Foundation Ireland*, die sich 2014 die strategische Aufstellung ihrer Aktivitäten zur öffentlichen Partizipation und Vermittlung naturwissenschaftlicher Themen zur Aufgabe machte und im folgenden Jahr ihr Evaluation Toolkit veröffentlichte (Science Foundation Ireland, 2017). Des Weiteren wirken übergreifende Akteure durch politische Vorgaben oder eine Verknüpfung an Förderungen im Bereich Wissenschaftskommunikation auf eine wirkungsorientierte Weiterentwicklung der Wissenschaftskommunikation in ihren Ländern und Regionen ein. Beispielsweise initiierte die australische Regierung im Jahr 2010 ein Programm zur Entwicklung einer nationalen Wissenschaftskommunikationsstrategie, mit dem Ziel, die Qualität solcher Maßnahmen zu leiten (Inspiring Australia, 2010). Ihre Evaluation galten demnach als Hauptmerkmal qualitativ hochwertiger Initiativen in der Wissenschaftskommunikation. Diesen Beispielen folgend bekannte auch das Bundesministerium für Forschung und Bildung in Deutschland 2019 mit der Veröffentlichung seines Grundsatzpapiers zur Wissenschaftskommunikation (BMBF, 2019) sein verstärktes Interesse an der Evaluation und Wirkungsorientierung in diesem Bereich. Einher geht dies mit einer übergreifenden Entwicklung im Wissenschaftssystem, welches sich verstärkt Fragen des ‚impacts‘ und insbesondere des ‚societal impacts of research‘ widmet. Es adressiert damit auch die Frage, wie Kommunikationsaktivitäten von Forschenden und wissenschaftlichen Einrichtungen in die Gesellschaft bewertet und verglichen werden können und bewertet Wissenschaft und Forschung auch anhand ihrer Nützlichkeit im gesellschaftlichen Kontext – womit (externer) Wissenschaftskommunikation eine umfassende Rolle zukommt.



Gleichzeitig können Fragen nach genau dieser Rolle der Wissenschaftskommunikation und der Erreichung mit ihr verbundener Zielsetzungen nur dann sinnvoll diskutiert und beantwortet werden, wenn dies auf Basis gesicherter Erkenntnisse über die Reichweiten und Wirkungen, aber auch Grenzen von Wissenschaftskommunikation in verschiedenen Formaten erfolgt. Aussagekräftige Evaluationen in der Wissenschaftskommunikation liefern hierzu wichtige Erkenntnisse. Diese ermöglichen nicht nur eine kritische Überprüfung der Praxis der Wissenschaftskommunikation, sondern liefern auch wichtige Hinweise für die Formulierung von Strategien für Akteure und Institutionen, die Wissenschaftskommunikation fördern. Diese können anhand möglichst standardisierter Evaluationspraktiken ihr Förderportfolio anhand der Fragen, ob geförderte Aktivitäten tatsächlich auf formulierte strategische Ziele einzahlen und dies in effektiver (und effizienter) Weise tun, auswählen und überprüfen. Dies ermöglicht eine Weiterentwicklung der Wissenschaftskommunikation auf strategischer wie praktischer Ebene.

Wissenschaft im Dialog (WiD) möchte sich dieser Thematik in Zukunft stärker widmen und in Zusammenarbeit mit wissenschaftlichen Partnern sowie Praktikerinnen und Praktikern der Branche die Etablierung einer stärkeren Evaluationskultur und Wirkungsorientierung in der deutschsprachigen Wissenschaftskommunikation unterstützen – eine Evaluationskultur, die sich quantitativer wie auch qualitativer Elemente bedient, auf gesicherten Erkenntnisse der ‚science of science communication‘ und sozialwissenschaftlichen Forschungsmethoden beruht und gleichzeitig Kontexte und Realitäten der Praxis der Wissenschaftskommunikation und damit auch die Grenzen von Evaluationen anerkennt. Erste Maßnahmen – und damit auch die Erläuterungen in folgendem Bericht – konzentrieren sich auf die nicht medienvermittelte Wissenschaftskommunikation, die auf externe Bezugsgruppen (nicht Forschende) ausgerichtet ist. Evaluationen des Wissenschaftsjournalismus, der Presse- und Medienarbeit oder etwa von Angeboten zur Nachwuchsrekrutierung im MINT-Bereich sind damit zunächst ausgeklammert.

Entwicklungen und Zukunftsperspektiven der Evaluationspraxis in der Wissenschaftskommunikation

Einhergehend mit dem Ausbau und einer steigenden Professionalisierung in der Wissenschaftskommunikation ist auch die stärkere Hinwendung zu Evaluation von Maßnahmen in diesem Bereich (Southwell & Marshall, 2010, S. 288). Dies mag auch an der erhöhten Nachfrage von Förderern liegen, die immer öfter eine Einordnung der geplanten bzw. erbrachten Leistungen sowie der gesellschaftlichen und zielgruppenspezifischen Wirkung voraussetzen.¹

Zeitgleich wurden in den letzten Jahren Stimmen lauter, die Verbesserungen in der bisherigen Evaluationspraxis verlangen: Ein Kritikpunkt betrifft die Qualität der Evaluationen unter sozialwissenschaftlichen Gesichtspunkten. Wie Jensen etwa kommentierte, weisen klassische Evaluationsbefragungen häufig Mängel hinsichtlich der Validität ihrer Aussagen auf, beispielsweise wenn Eltern Erlebnisse stellvertretend für ihre Kinder einordnen oder Fragen durch implizite Formulierungen

(„Haben Sie durch den Museumsbesuch heute irgendetwas Neues gelernt?“) zu unkritischen

¹ Beispielhaft zu nennen sind Projektrückblicke, wie im Fall der *Robert Bosch Stiftung* (2020) oder Spezifika in der Antragsstellung von Projekten, wie im Fall von EU-geförderten Maßnahmen (Europäische Kommission, 2018). Im gleichen Ton ermutigt das *Center for Advancement of Informal Science Education* (CAISE, 2019) dazu, bei der Antragsstellung geplante Auswertungsstrategien zur Ermittlung der Effekte eines Projekts einzubringen.



Reflexionen und Bewertungen verleiten (Jensen, 2014); Probleme, die Jensen auf fehlende Ressourcen, aber auch fehlende sozialwissenschaftliche Expertise und unzureichend durchdachte Evaluationsplanung zurückführt (2015, S. 2 f.). Diese Bedenken finden ein Echo in den Kritiken von Joubert und Weingart (2019) wie auch Weitkamp (2015), denen Evaluationen oftmals zu sehr auf deskriptiven Daten wie Reichweite basieren, wodurch sie die Chance verspielten, ein tieferes Verständnis der Erlebnisse im Rahmen solcher Maßnahmen zu erlangen. Dass die Planung von Evaluationen im Allgemeinen Verbesserungspotenzial aufweist, veranschaulicht eine Studie von Phillips und Kollegen (2018), die Evaluationsberichte von Citizen Science-Projekten mit den angegebenen Zielen des Projektes abgleicht. Hierbei wurde deutlich, dass zahlreiche Evaluationen nicht das messen, was sie als Ziel ihres Projektes auf ihrer Website angeben. Dies mag ein weiterer Hinweis sein, dass die Ziele eines Wissenschaftskommunikationsprojekts generell und besonders während der Planung ihrer Evaluation bis heute in zahlreichen Fällen unzureichend reflektiert werden, oder wie Stilgoe, Lock und Wilsdon es ausdrücken: „The *how* trumps the *why*“ (2014, S. 5).

Es lässt sich festhalten, dass Evaluationen zwar in weiten Kreisen als wichtiger Schritt in der Maßnahmenplanung bei Wissenschaftskommunikationsprojekten gelten, doch mit der Verbreitung und stetigen Etablierung wird auch die Frage nach ihrer Funktion, ihrer Optimierung und ihren Konsequenzen lauter. Gleichzeitig entsteht eine Art Gegenbewegung, welche die Erwartungen, die an Evaluationen gestellt werden, für zu hoch empfinden. King und Kollegen kommen etwa zum Schluss, dass keine sicheren Aussagen über gesellschaftliche Effekte auf Basis eines Evaluationsberichts getroffen werden könnten und stattdessen weniger ambitionierte Ziele anvisiert werden sollten (2015, S. 3). Eine Position, die nicht die Bedeutung von Evaluationen abschwächen soll, sondern vielmehr eine Besinnung auf realistische Zielsetzungen und entsprechend bedachte Evaluationsdesigns nahelegt.

Die Rolle von Wissenschaft im Dialog

Seit der Gründung im Jahr 2000 hat sich *WiD* als gemeinsame Initiative der Wissenschaftsorganisationen zu einer nationalen Organisation und einem Kompetenzzentrum für Wissenschaftskommunikation in Deutschland entwickelt. Ziel war es zunächst, das Verständnis der Bevölkerung von und für Wissenschaft zu stärken. Inzwischen hat *WiD* weitreichende Erfahrungen aufgebaut, was auch die Entwicklung und Durchführung innovativer Ansätze und Formate der Wissenschaftskommunikation angeht, um den Dialog zwischen Gesellschaft und Wissenschaft zu stärken, neue Zielgruppen zu erschließen und kontroverse Themen zu behandeln.

Im 2017 verabschiedeten *WiD*-Strategiepapier haben die Gesellschafterorganisationen beschlossen, dass sich *WiD* in Zukunft verstärkt der „Entwicklung von Strategien, [des] Monitoring[s] und [der] Bewertung von Kommunikationsaktivitäten“ (Wissenschaft im Dialog, 2017, S. 4) in der Wissenschaftskommunikation annehmen wird. Mit der Entwicklung einer „Evaluations-Toolbox“ im Rahmen eines erweiterten Angebots von Wissenschaftskommunikation.de ist nun die Möglichkeit geschaffen, einführende Informationen, Praxisbeispiele sowie nutzbare Materialien und Tools für Evaluationen in der Wissenschaftskommunikation zu entwerfen und bereitzustellen. Praktikerinnen und Praktikern der Wissenschaftskommunikation sollen damit zukünftig Hilfsmittel zur Verfügung gestellt werden, die auf die Praxis der Wissenschaftskommunikation mit ihren Kontexten und Realitäten eingehen, aber dennoch möglichst auf dem Forschungsstand zu Wissenschaftskommunikation und sozialwissenschaftlichen Gütekriterien basieren. Mit dieser Toolbox möchte *WiD* einen Beitrag dazu



leisten, das Thema der Wirkungsorientierung in der Wissenschaftskommunikation weiter ins Zentrum zu rücken und die Etablierung einer methodisch sinnvollen sowie wirkungsorientierten und damit aussagekräftigen Evaluationspraxis in der deutschen Wissenschaftskommunikation voranzutreiben.

Methodik der Recherche

Eine Grundlage für die Entwicklung von nutzbaren Materialien und Tools ist eine systematische Beantwortung folgender Fragen: Was sind mögliche Ziele von Wissenschaftskommunikation und nach welchen Dimensionen lassen sich diese differenzieren? Welche angestrebten Wirkungen durch Formate und Aktivitäten der Wissenschaftskommunikation sind damit verbunden? Wer sind Akteure und Zielgruppen unterschiedlicher Arten von Evaluationen und welchem Zweck dienen diese jeweiligen Arten? Welche verschiedenen Gründe, Motive und Anlässe lassen sich für Evaluationen identifizieren? Welche Formen von Evaluationsdaten können durch welche Methoden erhoben und ausgewertet werden? Eine Desktoprecherche, die den Forschungsstand hierzu erarbeitete und deren Ergebnisse im Folgenden geschildert werden, soll Antworten liefern und verschiedene Modelle und Sichtweisen vorstellen, die bei der Projekt- und Evaluationsplanung als Entscheidungshilfe dienen können.

Durch die Nutzung peer-reviewter wissenschaftlicher Quellen sollte die akademische Perspektive auf die Wirkungen und Ziele von Wissenschaftskommunikationsformaten, Evaluationskonzepte und -modelle erarbeitet und analysiert werden. Zu den herangezogenen Journals gehörten das *Journal of Science Communication*, *Public Understanding of Science* und *Science Communication* als internationale Quellen, darüber hinaus die deutschsprachigen Zeitschriften *Psychologische Rundschau* und die *Zeitschrift für Pädagogische Psychologie*. Letztere sollten vor allem den kommunikationswissenschaftlichen Fokus um eine psychologische Sichtweise ergänzen. Fokussiert wurden dabei die Jahrgänge von 2012 bis 2019; ein Zeitraum, der im Anbetracht der begrenzten Zeit und personellen Ressourcen im Projekt einen aktuellen Einblick ermöglichen sollte. Relevante Beiträge wurden anhand einer schlagwortgeleiteten² Recherche identifiziert. Darüber hinaus wurden Tagungsbeiträge der PCST-Konferenz 2018 (*Public Communication of Science and Technology*) und Projekte des abgeschlossenen DFG-Schwerpunktprogramms „Wissenschaft und Öffentlichkeit“ (2009-2015) gesichtet. Ergänzt wurde die Recherche durch eine Google-Scholar-Suche nach zentralen Schlagworten. Selbstverständlich kann eine Recherche, die unter zeitlichen Begrenzungen und auf Basis einer begrenzten Anzahl an Schlagworten und Quellen durchgeführt wird, die vielen eröffneten Fragen nur beginnen zu beantworten. In ihrer explorativen Form hat sie dennoch das Potenzial, in das komplexe Thema einzuführen und zentrale Elemente aufzuzeigen.

An dieser Stelle können beispielsweise bereits erste Erkenntnisse aus der Schlagwortsuche für künftige Recherchen gezogen werden: Die deutschen Schlagworte *Wissenschaftskommunikation*, *Wissenschaftspädagogik* und *Wissenschaftsvermittlung* werden selten in der Branche genutzt, da

² Zu den Kernbegriffen der Schlagwortsuche zählten: *Science Education*, *Science Communication*, *Science Engagement*, *Public Engagement with Science and Technology* sowie *Public Engagement*. Diese Schlagwörter wurden in separaten Rechenschritten mit sogenannten Kombinationsbegriffen gemeinsam abgefragt: *Impact*, *Evaluation*, *Tools*, *Meta analysis* und *Outreach*. Im Fall der deutschen Quellen wurden darüber hinaus deutsche Schlagwörter hinzugezogen. Konkret wurden hierbei die Kernbegriffe *Wissenschaftskommunikation*, *Wissenschaftspädagogik* und *Wissenschaftsvermittlung* mit *Evaluation*, *Wirkung*, *Einfluss* und *Metaanalyse* kombiniert.



deutschsprachige Autorinnen und Autoren des Feldes ebenfalls internationale Standards gebrauchen. Zudem publizieren sie eher in internationalen Journals – nur wenige für die Recherche relevante Beiträge konnten in den deutschen Wissenschaftszeitschriften ausfindig gemacht werden. Gleiches galt für die Googlerecherche, bei der relevante deutschsprachige Quellen nicht direkt aufzufinden waren. Auch wenn in Geschäftsberichten über die Existenz von Evaluationen und deren Dokumentation gesprochen wird, sind diese oftmals nicht öffentlich zugänglich. Zu den Schlagworten, die sich bei der Fachliteratur als ergiebig erwiesen, gehören *Public Engagement with Science and Technology*, *Science Education* und *Science Communication* in Kombination mit den Begriffen *Evaluation*, *Impact* oder *Outreach* im Text.

Die für die Recherche genutzte Systematik ermöglicht einen Überblick über relevante Informationen zu den erläuterten Themen und Schlagworten. Dennoch ist zu betonen, dass kein Anspruch auf Vollständigkeit oder Repräsentativität erhoben werden kann oder soll. Grund ist, dass neben den klaren Parametern, die für die Recherche gesetzt wurden (z. B. Zeitspanne und Schlagworte), die subjektive Einschätzung der Relevanz eines Artikels als zentrales Kriterium gilt. Das bedeutet, dass über die Schlagwortsuche identifizierte Beiträge nicht einbezogen wurden, wenn die Inhalte keinen Mehrwert für die Recherche darstellten. Umgekehrt wurden Beiträge einbezogen, wenn sie als besonders erkenntnisreich für die Beantwortung Recherchefragen galten, obwohl sie beispielsweise nicht über das geschilderte Verfahren identifiziert wurden. Die Recherchestrategie folgt eher dem Prinzip des Schneeballverfahrens, da auf Basis des Quellenbestands weitere Artikel identifiziert wurden, die durch den ersten Kriterienkatalog nicht auffindbar waren. Als weitere Quelle galten zudem Expertinnen und Experten oder Partnerinnen und Partner des Projekts, die Literatur und Projekte empfahlen. Außerdem wird vermutet, dass der (Nicht-)Zugang zu Literatur den Quellenbestand verzerrt. Open-Access-Beiträge hatten eine höhere Chance, in der Recherche einbezogen zu werden, als Closed-Access-Beiträge. Trotz dieser Limitationen wird die Recherche als für ihren Zweck ausreichend befunden: ein erster Abriss der Erkenntnisse zur Wirkungsmessung und theoretischer Evaluationsstandards. Es wurden 80 Beiträge aus der Fachliteratur zur näheren Betrachtung ausgewählt und 33 – hauptsächlich aus dem Ausland stammende – Toolkits und Ratgeber gesammelt. Nur 19 der 113 Quellen wurden von deutschen Instituten herausgegeben. Hier bestätigt sich der Eindruck, dass gerade in Nordamerika, Großbritannien und Australien das Forschungsinteresse an der Evaluation von Wissenschaftskommunikationsformaten bisher ausgeprägter ist als im deutschsprachigen Raum.

Theorie zur Evaluation der Wissenschaftskommunikation

Zielebenen von Projekten in der Wissenschaftskommunikation

Eine Beschäftigung mit Wirkungen und Evaluation von Wissenschaftskommunikation setzt zunächst eine Auseinandersetzung mit den Intentionen von Wissenschaftskommunikationsprojekten voraus. Schließlich ist die Durchführung einer Evaluation dadurch motiviert, den Erfolg eines Projekts zu messen – idealerweise mit Blick auf vorab definierte, konkrete Ziele. Die andauernde Konfrontation mit der Frage, was Wissenschaftskommunikation im Allgemeinen und im Fall konkreter Projekte erreichen soll, ist von zentraler Bedeutung, um zu Erkenntnissen zu gelangen und richtungsweisende Entscheidungen zu treffen. Trotzdem verlieren sich Projekte zu oft noch in unklaren Zielsetzungen, was sich beispielsweise dann zeigt, wenn zu viele Vorhaben vermischt werden – etwa die Absicht, Begeisterung für Forschung oder gar wissenschaftliche Karrieren zu erzeugen und gleichzeitig Dialog und Beteiligung anzuregen – oder sehr vage Zielgruppen formuliert



werden. Die Problematik überträgt sich – wie bei der Schilderung der Relevanz des Themas bereits aufgegriffen wurde – auch auf systemischer Ebene, mit Blick auf Projektförderungen und politische Strategieentwicklungen. Umso bedeutsamer für die Beschäftigung mit Wirkungen, wie auch die Entwicklung von Evaluationsmaterialien, ist es, sich an dieser Stelle nochmals mit den bisherigen Zielformulierungen und -ebenen auseinanderzusetzen, die bei der Planung und Umsetzung von Wissenschaftskommunikation zum Tragen kommen.

In einer vergleichenden Analyse von deutschen Wissenschaftskommunikationsformaten aus dem Jahr 2008 wurden vier generelle Ziele solcher Projekte identifiziert: Wissensvermittlung, Nachwuchsförderung, Partizipation und die Nutzung von Wissenschaft als unterhaltsame, kulturelle Aktivität (Weingart et al., 2008, S. 4). Auch wenn diese vier Aspekte bis heute so wichtig wie erstrebenswert sind, haben sich seitdem die verfolgten Zielrichtungen ebenso ausdifferenziert wie die Strukturen und Programme der Wissenschaftskommunikation selbst. Mit Wachstum und Professionalisierung des Feldes geht auch eine steigende Vielfalt einher, die sich nicht nur in der Entwicklung neuer Formate und Ansprache bisher unbeachteter Zielgruppen bemerkbar macht, sondern inzwischen zahlreiche Wissenschaftsfelder wie auch Initiatorinnen und Initiatoren betrifft. All diese Aspekte erweitern dementsprechend auch die Bandbreite an verfolgten Zielen.

Um die Ergebnisse einer Aktivität in ihrem Ausmaß zu unterscheiden und strukturiert betrachten zu können, wird in der Evaluationsliteratur häufig auf die Trennung von Output, Outcome und Impact verwiesen (Kurz & Kubek, 2018; Phillips et al., 2014; Phillips et al., 2018). Dieser Logik folgend werden zur Umsetzung eines Projekts stets Ressourcen investiert, die als Input gelten. Hierzu können Materialien und Ausstattung zählen, aber auch personelle Ressourcen. Daraus folgende Entwicklungen können in Outputs, Outcomes und Impacts eingeordnet werden.

- **Output:** erbrachte Leistungen, Nutzung der Leistungen durch Zielgruppen, Zufriedenheit der Teilnehmenden
- **Outcome:** neues Wissen, Fertigkeiten, Fähigkeiten, Bildung eigener Meinungen, Veränderungen im Handeln oder in den Lebenslagen der Teilnehmenden
- **Impact:** Veränderungen innerhalb der Gesamtgesellschaft

Outputs gelten dieser Logik zufolge noch nicht als Wirkung einer Maßnahme, sondern sie sind die Leistung, die durch das Angebot und die Beteiligten erbracht wird: die Vorlesung oder der Workshop vonseiten der Verantwortlichen und, im Fall eines interaktiven Programms, die erbrachten Ergebnisse vonseiten der Teilnehmenden. Was Einzelne aus diesem Erlebnis mitnehmen oder was auf lange Sicht auf gesellschaftlicher Ebene zum Tragen kommt, gilt dabei als Wirkung (für eine ausführliche Ausführung, siehe Kurz & Kubek, 2018, S. 33 ff.).

Welche Zielvorstellungen in der Wissenschaftskommunikation besonders präsent sind, zeigt eine Sammlung der gemeinnützigen Treuhand *Wellcome Trust*, die Forschung und Forschende insbesondere im Gesundheitssektor finanziert und seit 2012 verstärkt verschiedenartige Reports zu ihren Public Engagement-Aktivitäten und -Projekten veröffentlicht (Wellcome Trust, 2020). In ihrem jährlichen „UK Centres Public Engagement Workshop“ wurden 2015 die Ziele der Praktikerinnen und Praktiker aus der Wissenschaftskommunikation (Forschende verschiedener Fachbereiche, Künstlerinnen und Künstler, Führungskräfte aus Research Centres und Projektmanagerinnen und -manager) gesammelt. Hierbei wurde eine Trennung in betroffene



Systeme oder Personengruppen vorgenommen: Auswirkungen, welche die Öffentlichkeit oder Teilöffentlichkeiten betreffen sollen; Auswirkungen, die man sich bei Forschenden und Institutionen erhofft und solche, die sich auf Beziehungen zwischen den Systemen – Öffentlichkeit, Wissenschaft, ggf. auch Politik oder die öffentliche Ordnung – beziehen (Wellcome Trust, 2015, S. 6).

Die erwünschten Veränderungen in der Öffentlichkeit bezogen sich vor allem auf **die Steigerung von wissenschaftlichem Interesse, Inspiration, Verständnis, Verbindung und Vertrauen**. Für Forschende erhoffte man sich die **Verbesserung ihrer kommunikativen Fertigkeiten** wie auch ihrer **Inspiration und neue Perspektiven für die eigene Forschung**, doch dieses Feld regte noch diversere Antworten an. So wurden auch pragmatische Funktionen genannt: **Team building** zu betreiben oder zentrales **Wissen an Stakeholder** zu leiten. Der **Wissens- und Erfahrungsaustausch** zwischen Gesellschaft und Wissenschaft, das Knüpfen neuer Kontakte, die **Debatte von Forschungsthemen** sowie das **Erfassen bestehender Meinungen** und Relevanzempfindungen in der Öffentlichkeit zählte beispielhaft zu den Zielen hinsichtlich der System-Beziehungen.

Diese Zielformulierungen wurden auch in der weiteren Recherche wiedererkannt. Auffällig ist jedoch, dass im Fall des *Wellcome Trust* die langfristige Impact-Orientierung deutlicher zu Tage tritt als Bestrebungen, Outcomes bei den Individuen innerhalb der Öffentlichkeit zu erreichen. Andere Projekte setzen den Fokus auf die Ebene des Einzelnen, wie der folgenden Abbildung (Abb. 1) zu entnehmen ist. Sie eröffnet außerdem eine weitere Perspektive: Die Ziele variieren nicht nur in den Zielgruppen und Wirkungsebenen, sondern auch die Projektbeteiligten weisen unterschiedliche Priorisierungen und Beweggründe auf. Die Aufschlüsselung nach Forschenden, die sich in Wissenschaftskommunikationsformaten beteiligen, und nach den nicht-wissenschaftlichen Projektbeteiligten veranschaulicht, dass die Zielrichtung ähnlich ausfällt, aber von unterschiedlichen Motiven getrieben sein mag: Im Fall der Forschenden etwa von der Motivation, steuerzahlenden Bürgerinnen und Bürgern über die durch sie finanzierte Projekte zu berichten oder die eigene akademische Karriere zu fördern; im Falle der institutionellen Wissenschaftskommunikatorinnen und -kommunikatoren etwa von Bestrebungen, das allgemeine Wissenschaftsverständnis der Teilnehmenden zu unterstützen oder die Arbeit bestimmter Institutionen zu legitimieren.

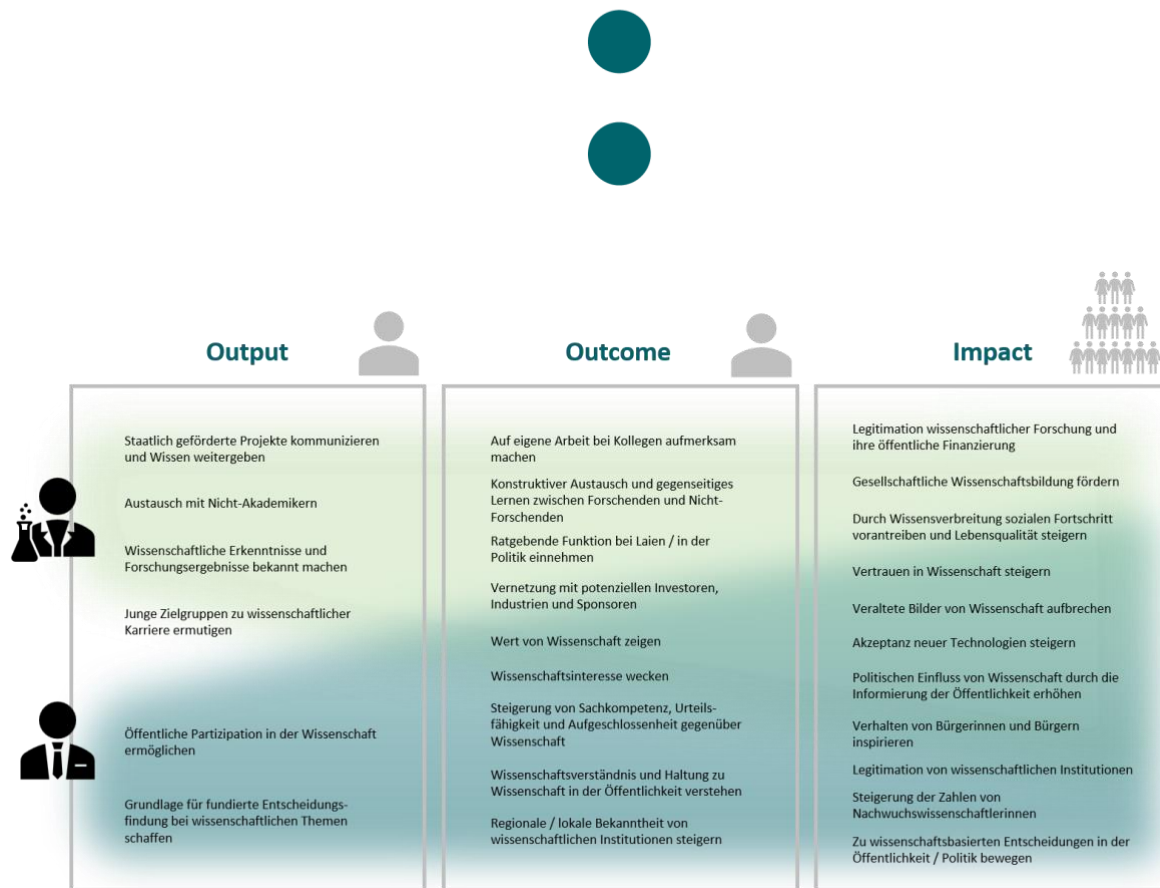


Abbildung 1: Beispielhafte Output-, Outcome- und Impact-Ziele aus Sicht der Forschenden und aus Sicht institutioneller Wissenschaftskommunikatorinnen und -kommunikatoren einer Forschungseinrichtung. (Eigener, nicht umfassender, vorläufiger Entwurf)

Darüber hinaus wird bei näherer Betrachtung deutlich, dass erhoffte Outcomes und Impacts (unabhängig von den Projektbeteiligten, die sie verfolgen) auf Motiven beruhen können, die sich primär über das gesellschaftliche Wohl oder primär über das institutionelle Wohl definieren – auch wenn im Fall gemeinnütziger Organisationen vom institutionellen Wohl gleichzeitig auch gesellschaftliche Verbesserungen erhofft werden. Die institutionelle Zielorientierung soll an dieser Stelle nicht als weniger wertvoll beurteilt werden. Beide Perspektiven, die in Abbildung 2 als ein Kontinuum aufgezeigt sind, sind bei der Betrachtung von Zielen der Wissenschaftskommunikation relevant und in der Realität in Vielzahl vorhanden.



Abbildung 2: Zielkategorien der Wissenschaftskommunikation auf Wirkungsebene, in Hinblick auf externe Zielgruppen. (W&F: „Wissenschaft und Forschung“; eigener, vorläufiger Entwurf)



Für die strategische Planung und eine erkenntnisreiche Evaluation ist die Auseinandersetzung mit den eigenen Zielen ein notwendiges Kriterium. Die vorgestellten Modelle können dabei behilflich sein, diese Ziele vor Augen zu führen: Verfolgt das Projekt Ziele in Form von konkreten Leistungen, Wirkungen auf Einzelne oder gesellschaftliche Langzeiteffekte (wenn nicht sogar alle drei)? Was konkret soll bei den angesprochenen Zielgruppen erreicht werden? Verfolgen Beteiligte und Mitarbeitende mit dem Projekt verschiedene Ziele und wenn ja, welche? Beruhen die Ziele auf unterschiedlichen Motivationen und hat dies Auswirkungen auf die Aktivitäten?

Mit Einigung auf konkrete Zielformulierungen mag sich bereits ergeben, wer die Zielgruppe für die Evaluation darstellt, für wen evaluiert wird und welche Methode die Erfassung der fokussierten Zielebenen zulässt. Bei der Planung eines Projekts kann nicht davon ausgegangen werden, dass alle Beteiligten automatisch einer Meinung sind, welches Ziel verfolgt werden soll. Die Planung von Evaluationen bietet hierfür eine gute Gelegenheit, da sie von allen an der Umsetzung oder Förderung beteiligten Akteuren verlangt, sich von Beginn an darauf zu verständigen, welche konkreten qualitativen wie quantitativen Zielsetzungen auf welcher Wirkungsebene in welcher Langfristigkeit angestrebt werden sollen.

Wirkungen von Wissenschaftskommunikation

Für eine wirkungsorientierte Formulierung von Zielen in der Wissenschaftskommunikation, stellt die (wissenschaftliche) Literatur zu Wirkungen von Wissenschaftskommunikation eine wichtige Informationsquelle dar. Im folgenden Abschnitt soll daher angerissen werden, wie sich Wirkungsformen unterscheiden lassen, welche empirischen Befunde bei Zielgruppen verzeichnet wurden und welche Schlüsse daraus für die Evaluationspraxis gezogen werden können.

Wirkungsmodelle von Wissenschaftskommunikationsmaßnahmen auf die Öffentlichkeit

Oftmals werden „die Öffentlichkeit“ oder bestimmte Teilöffentlichkeiten als zentrale Zielgruppen genannt. Die Auswirkungen, die Wissenschaftskommunikation auf diese haben kann, wurden bereits in ersten Modellen gruppiert. Beispiele hierfür bieten das *Framework for Evaluating Impacts of Informal Science Education* der *National Science Foundation (NSF)* in den Vereinigten Staaten (Friedmann et al., 2008, S. 19), die *LSIE*-Kategorien (Learning Science in Informal Environments) des *National Research Councils* (2009, S. 57 ff.) in den USA oder das *GLO*-Modell (Generic Learning Outcomes) des *Arts Council Englands* (2020). Alle Modelle unterscheiden fünf bis sechs Wirkungsdimensionen, die bei der Zielgruppe eines Angebots erreicht werden können. Zur Veranschaulichung soll das Framework der *NSF* herangezogen werden (Friedmann et al., 2008, S. 19) – zum einen, weil es als Vorbild vieler Modelle im Rahmen informellen Lernens gilt (Phillips et al., 2018, S. 3), zum anderen, da es trotz seiner Erstanwendung im MINT-Bereich in seiner Form leichter auf verschiedene Formate der Wissenschaftskommunikation übertragbar erscheint als die auf Citizen Science bezogene *LSIE*-Vorlage oder das im kulturellen Bereich zu verortende *GLO*-Modell.

Das NSF-Evaluation-Framework

Teilnehmende demonstrieren nach, während bzw. durch das Angebot eine messbare Bewertung, Veränderung oder Ausübung (von) ...



- **Wissen, Bewusstsein und Verständnis** für ein wissenschaftliches Thema, Phänomen, Konzept
- **gewecktem Engagement oder Interesse** für ein wissenschaftliches Thema, Phänomen, Konzept
- **(ggf. neu erlernten) Fähigkeiten**
- einer **Einstellung** gegenüber einem wissenschaftlichen Thema, Phänomen, Konzept. Damit gemeint sind relativ stabile Konstrukte, wie die Wertschätzung von Wissenschaft oder, um ein konkretes Beispiel zu nennen, die Befürwortung oder Ablehnung von Stammzellforschung
- **Verhaltensweisen** im Zusammenhang mit oder in Bezug auf ein wissenschaftliches Thema, Phänomen oder Konzept

Eine Verstärkung von bisherigen Einstellungen, Fähigkeiten, Verhaltensweisen usw. kann dabei ebenso als Wirkung gelten wie deren Veränderungen.

Wirkungsbefunde von Wissenschaftskommunikationsmaßnahmen auf die Öffentlichkeit

Das theoretische Framework deckt sich in einigen Bereichen mit der empirischen Befundlage. Gerade die Verbesserung des allgemeinen Wissensstands, Wissenschaftsverständnisses und der sensibilisierten Wahrnehmung wissenschaftlicher Themen (Bonney et al., 2016; Boyette & Ramsey, 2019) sowie das Erlernen und der Einsatz von Fähigkeiten (Bonney et al., 2016; Rodgers et al., 2018) konnte vonseiten der Forschung für verschiedene Wissenschaftskommunikationsprojekte bestätigt werden. Ergebnisse zur Einstellungsveränderungen fallen divers aus. In vielen Fällen werden diesbezügliche Wirkungen nicht bestätigt oder entsprechende Effekte sehr gering eingeschätzt (Bonney et al., 2016; Crall et al., 2013; Lin et al., 2015). Dies lässt sich zum Teil auch durch Deckeneffekte erklären. Da typische Teilnehmende bestimmter Formate bereits eine positive Haltung gegenüber Wissenschaft mitbringen, können jene Effekte nicht oder kaum nachgewiesen werden. Selbst wenn die evaluative Einstellung zu bestimmten wissenschaftlichen Themen im Fokus lag, wurden nicht immer nennenswerten Effekte gemessen (Morin, 2018). Es gibt allerdings auch Beispiele, die positive Veränderungen in der Einstellung verzeichneten (Nadkarni & Morris, 2018; Vitone et al., 2016). Befunde zu Verhaltensänderungen sind aufgrund der Messproblematik schwieriger einzuordnen. Die Abfrage der Verhaltensintention lässt nicht mit Sicherheit darauf schließen, ob solche Vorsätze auch in Taten enden. Hierfür wäre es nötig, die Teilnehmenden eines Formats über einen langen Zeitraum hinweg zu beobachten oder sie zumindest in regelmäßigen Abständen wiederholt zu befragen, ob sich Verhaltensintentionen auch dauerhaft durchsetzen. Forschungsökonomische Gründe oder Unsicherheiten bezüglich der Validität (wie die wahrheitsgemäße Beantwortung zugunsten der sozialen Erwünschtheit) stehen konkreten Aussagen zur Verhaltensänderung daher oftmals im Weg.

Wirkungen von Wissenschaftskommunikationsmaßnahmen auf Forschende

Wissenschaftskommunikation-Projekte können ebenso eine Wirkung auf die involvierten Wissenschaftlerinnen und Wissenschaftler ausüben. So ergab eine Befragung von Watermeyer (2012, S. 762), dass der Austausch mit Personen, die nicht fachkundig sind, die wissenschaftliche Arbeit bereichern kann, da ihre Sichtweisen und Ideen neue Denkanstöße geben und die Distanz zum Thema es ihnen erlaubt, Konzepte auf neue Weise kritisch zu hinterfragen. Zudem sollen Programme zur Kommunikation der eigenen Arbeit dazu beitragen, dass Forschende wissenschaftliche Konzepte besser erklären können, ihre Arbeit reflektieren und das Interesse an



den Implikationen der Forschung außerhalb des akademischen Raums steigern (Carpenter, 2015, S. 123 f.; Gamse, 2010) . Einige Studien lassen zudem vermuten, dass die aktive Teilnahme an Aktivitäten der Wissenschaftskommunikation positive Auswirkungen auf die Karriere der Forschenden haben kann: So kann die Partizipation in Wissenschaftskommunikation die Bekanntheit ihrer Forschungsprojekte steigern und Anfragen für Konferenzen und Förderungen nach sich ziehen (Alonso-Flores & Moreno-Castro, 2018, S. 39). Negative Wirkungen wurden allerdings auch erkannt. Forschende bemerken etwa, dass wissenschaftliche Erkenntnisse in den Medien verzerrt dargestellt (ebd.) und damit auch vom Publikum unter- oder überschätzt werden können.

Wirkungsorientierung in der Evaluationspraxis

Mit Blick auf die vorangegangenen Wirkungsstudien fällt auf, dass Nachweise auf Output- und Outcome-Ebene wesentlich leichter zu finden sind, als Belege für gesellschaftliche Veränderungen. Auch wenn es um die Verhaltensänderung von Individuen geht, erschwert sich die Beweislage. Trotzdem sind die Ziele für Wissenschaftskommunikationsprojekte nicht selten auf der Impact-Ebene verortet. Für die Evaluation ihrer Wirksamkeit stellt das eine Herausforderung dar, denn die Messung von gesellschaftlichen und langfristigen Effekten erweist sich als schwierig. Sie erfordert umfassende Evaluationsdesigns, den Einbezug von Kontrollgruppen und es ist nahezu unmöglich, äußere Einflussfaktoren wie das Vorwissen der Beteiligten auszuklammern (King et al., 2015, S. 2 f.). Aus diesem Grund konzentrieren sich viele Praktikerinnen und Praktiker auf leichter erhebbare Daten (Fogg-Rogers et al., 2015, S. 1 f.), die auf der Output-Ebene liegen. Ein Beispiel wäre die Evaluation der Junior Uni Wuppertal (Makles & Schneider, 2016), die zur Einordnung ihres Erfolgs vornehmlich Teilnehmerzahlen und demographische Daten der partizipierenden Kinder wie Geschlecht und sozioökonomischer Hintergrund zu Rate zieht, oder „Die Lange Nacht der Wissenschaften“ in Berlin, die zur Erfolgsbewertung ähnliche Daten wie Besuchszahlen und wiederholten Besuche heranzieht. Eine Metastudie von Phillips und Kollegen zur Evaluation von Citizen Science Projekten (2018) verfestigt den Eindruck, dass die Erhebung von Outputs wie die Motivation zur Teilnahme an und Zufriedenheit mit dem Projekt, neben Klickzahlen und leicht erkennbaren Erfolgsmetriken wie die Zahl an aus den Projekten resultierenden Publikationen eine große Rolle spielen.

Ziele, Motivationen und Arten von Evaluationen

Aussagekräftige Evaluationen nach möglichst objektiven Standards ermöglichen nicht nur die Bewertung des Bestehenden, sondern ihre Ergebnisse eröffnen auch Möglichkeiten zur Weiterentwicklung der Projekte. Welche Bereiche eines Formats oder einer Aktivität hierbei in den Blick genommen werden, hängt davon ab, ob die Evaluation prozessorientiert oder wirkungsorientiert angelegt wird. Während der Planung eines Projekts ist es daher sinnvoll, festzulegen, was die Evaluation messen soll und welche Funktion sie verfolgt, denn diese Entscheidungen bestimmen auch den Zeitraum und die Art der Evaluation. Vor diesem Hintergrund soll kurz auf vier verschiedene Arten von Evaluationen (Science Foundation Ireland, 2015) eingegangen werden. Sie begleiten unterschiedliche Projektphasen, können aber unabhängig voneinander durchgeführt werden.

1. *Die formative Evaluation:* Sie findet während der Projektentwicklung statt und dient der Prüfung vorangegangener Planungsprozesse, der Konzeptionalisierung und Strategie eines Projekts: Sind die Ziele, die Ausgestaltung und die Evaluationsmethode des



geplanten Formats sinnvoll aufeinander abgestimmt?

2. *Die Prozess-Evaluation:* Sie wird zu Beginn eines Projekts eingesetzt, um gesetzte Annahmen zu überwachen und ggf. frühzeitige Anpassungen vorzunehmen. Auf diese Weise kann Problemen noch rechtzeitig entgegengewirkt werden: Funktioniert das Projekt in der Praxis so wie geplant?
3. *Die summative Evaluation:* Bei dieser Form von Evaluation wird eine Zwischenbilanz während der Projektlaufzeit vorgenommen: Werden die gesteckten Ziele ansatzweise erreicht?
4. *Die Impact-Evaluation:* Sie findet während der Abschlussphase oder nach Vollendung eines Projekts statt.³ Hier werden die gesetzten Ziele geprüft und die Erfolge oder mögliche Probleme und Fehlentwicklungen ergründet: Wurde mit dem Projekt das erreicht, was vorgesehen war?

An anderen Stellen in der Literatur wird dabei lediglich zwischen *formativen Evaluationen als generell prozessorientierten Evaluationen* und *summativen Evaluationen als generell abschließenden Evaluationen* unterschieden. Projekten, die verschiedene Evaluationsarten in ihre Aktivitäten integrieren, begegnet man in der Wissenschaftskommunikation selten. Stattdessen sind es häufig summative Evaluationen, die durchgeführt werden und sich im allgemeinen Verständnis des Begriffs „Evaluation“ durchsetzen. Diese als bereits Impact-Evaluationen zu verstehen, wäre allerdings ein Irrtum. Denn während es sich bei der summativen Evaluation vielmehr um Monitoring, das Sammeln von deskriptiven Daten, handelt, setzt die Impact-Evaluation den Fokus auf das Verstehen und Nachvollziehen von Zusammenhängen. Was beide Arten gemein haben, ist eine wirkungsorientierte Ausrichtung (z. B. zur Betrachtung von Zielerreichungen), während formative Evaluationen, welche die Optimierung der Abläufe ins Zentrum rücken, prozessorientiert ausgerichtet sind.

Motive für Evaluationen

Fehlt es an Kapazitäten, um das Projekt durch ständige Evaluation zu begleiten, stellt sich die Frage, welche Evaluationsart fokussiert werden sollte. Diese Entscheidung hängt in erster Linie von der Motivation der Evaluation ab, die sich auch in den Funktionsbeschreibungen der verschiedenen Evaluationsarten widerspiegelt. Und auch diese Motivationen erweisen sich in der Literatur als divers. Evaluationen können einerseits der Erkenntnis dienen, welche Aspekte eines Erlebnisses funktionieren, in welcher Art und Weise, für welche Zielgruppen und aus welchem Grund (Jensen, 2015, S. 7). Hinter dieser Aufklärung der Wirkweisen kann auch der Anreiz stehen, anhand klarer Erfolgsfaktoren das Programm gegenüber Förderinnen und Förderern zu legitimieren (Friedmann et al., 2008, S. 8; Joubert, 2007; Kurz & Kubek, 2018, S. 8; Sanchez-Mora, 2014, S. 1). In anderen Zusammenhängen werden sie als Mittel zur Prüfung und Steigerung der Effizienz eingesetzt (Boyette & Ramsey, 2019; Grand & Sardo, 2017, 1 f.; Joubert, 2007; Phillips et al., 2018, S. 5), nicht zuletzt um Budget- und Ressourcen zu schonen und relevante Akteure und ihre Zusammenarbeit

³ Obwohl diese Form der Evaluation erst nach der Maßnahmendurchführung ansteht, sollte ihre Planung bereits während der Konzeption eines Projekts mitgedacht werden. Evaluationen, die im Nachgang geplant werden, laufen in Gefahr, die anfängliche Zielausrichtung zu verfehlen und nicht das zu messen, was zu Beginn angedacht war.



zu planen (Fischhoff, 2019). Außerdem kann sie Mitarbeiterinnen und Mitarbeiter motivieren (Kurz & Kubek, 2018, S. 8), denen anhand der Evaluation die Ergebnisse ihrer erbrachten Leistung verdeutlicht wird. Zusätzlich kann es ihnen die Möglichkeit bieten, eigene Lernerfahrungen aus dem Projekt einzubringen und gemeinsam mit der Leitungsebene Verbesserungen für kommende Aktivitäten zu erarbeiten, seien sie inhaltlicher oder formaler Natur.

Wer Evaluationen lediglich als einen ergänzenden Projektschritt wahrnimmt, verkennt ihr wahres Potenzial, wegweisende Erkenntnisse für das Projekt zu erschließen. Gleichzeitig müssen die eigenen Erwartungen den Ressourcen gegenübergestellt werden, die zur Verfügung stehen. Eine einmalige Online-Befragung wird kaum in der Lage sein, all den genannten Zielen gerecht zu werden und sämtliche Beteiligte zufrieden zu stellen. Ist die Motivation hinter den eigenen Evaluationsmaßnahmen (oder zumindest eine klare Priorisierung von Zielen und Motiven) geklärt und eine Entscheidung gefällt, wer als Adressat gilt und wie sie anschließend genutzt wird, so fällt die Wahl und Planung der adäquaten Evaluationsarten wesentlich leichter.

Objekte von Evaluationen

Mit der Definition wirkungsorientierter Ziele und der Wahl einer Evaluationsart ergeben sich auch erste Hinweise darauf, was konkret erhoben werden sollte. Generell kommen eine Reihe von Objekten in Frage, die das Format und die erreichte Zielgruppe beschreiben, aber auch solche, die auf Outcomes und Impacts, also kausale Effekte und Wirkmechanismen, abzielen (Fu et al., 2016, S. 13 f.).

Steht die Beschreibung der Aktivität bzw. die Identifikation von Outputs im Vordergrund, sind folgende Objekte häufig von Interesse:

- die Messung der Reichweite eines Formats (Besuchszahlen, Klickzahlen etc.)
- Soziodemografie und stabile Personenmerkmale der Teilnehmenden zur Zielgruppendefinition
- Motivation zum Besuch oder zur Teilnahme
- die Zufriedenheit mit dem Format
- Kritik und Verbesserungsvorschläge
- Annahme von Angeboten (z. B. die Nutzung von Informationsangeboten)
- Wahrscheinlichkeit, in Zukunft an ähnlichen Formaten teilzunehmen

Unabhängig davon, ob diese Objekte an beteiligte Forschende oder die teilnehmende Zielgruppe gerichtet sind: Die Erhebung derartiger Daten zielt darauf ab, diese Gruppen besser kennenzulernen und zu erkennen, wie eine Maßnahme bewertet und angenommen wird.

Wird das Ziel verfolgt, Wirkungen einer Aktivität auf die Bezugsgruppen zu prüfen, bietet die Erhebung folgender Daten an.

In Bezug auf Teilnehmende des Angebots:

- Interesse an, Bewertung oder Einordnung von konkreten wissenschaftlichen Themen
- Urteilsfähigkeit und Selbstwirksamkeit in wissenschaftlichen Debatten
- Wissen in Hinblick auf wissenschaftliche Inhalte oder wissenschaftliche Prozesse und Vorgehensweisen
- Bekanntheit von Institutionen



- Einstellung oder Vertrauen zu Wissenschaft, der Forschungsgemeinde oder konkreten Institutionen
- Verhalten(-sintentionen)

In Bezug auf beteiligte Forschende oder Beteiligte aus dem akademischen Feld:

- Einstellung zu Wissenschaftskommunikation, Partizipationsformaten, Open Science etc.
- Fähigkeiten (z. B. bei Forschenden: zur praktischen Vermittlung und Veranschaulichung ihrer Forschung)
- Einstellung zur Verantwortung von Wissenschaft

Zur Prüfung kausaler Zusammenhänge zwischen der Teilnahme an einem Format und der Veränderung vonseiten der Zielgruppe gilt es allerdings nicht nur die korrekte Erhebung zu beachten. Im Gegensatz zu deskriptiven Fragestellungen, deren Daten sich isoliert betrachten lassen, ist es bei der Beantwortung effektorientierter Fragestellungen zentral, die Beziehung verschiedener Daten zueinander und damit auch mögliche weitere Einflussfaktoren zu bedenken. Wurden die Vorkenntnisse, Einstellungen oder Verhaltensweisen der Zielgruppe vorab geprüft, um sichergehen zu können, dass eindeutige Veränderungen vorliegen? Kann ein anderer Umstand als die inhaltlichen Merkmale des eigenen Formats Veränderungen in der Zielgruppe verantworten? Ein Beispiel: Selbst der Vergleich aktueller und vergangener Daten (beispielsweise die Evaluationsergebnisse eines Science Slams im Vergleich mit jenen der letztjährigen Ausgabe), der etwa einen Zuwachs von Teilnehmenden andeutet, gibt keine verlässlichen Hinweise darauf, *wieso* ein Format besser oder schlechter ankommt als zuvor. Wird eine Veranstaltung besser besucht als im Jahr zuvor, kann dies schließlich an zahlreichen Faktoren liegen, die nichts mit dem Programm oder seiner Vorbereitung zu tun haben: Vielleicht fand eine weitere, attraktive Veranstaltung statt, öffentliche Verkehrsmittel fielen aus oder das Wetter spielte nicht mit. Zwar lassen sich auf Basis solcher Datenvergleiche Vermutungen anstellen, doch sie liefern keine Beweise und sind in ihrer Aussagekraft beschränkt.

Aus diesem Grund ist bei der Untersuchung von Outcomes und Impacts nicht nur relevant, was erhoben wird, sondern auch die Frage danach, wie diese Objekte erhoben werden – kurzum: ob die gewählte Analysemethode kausale Schlussfolgerungen zwischen den Merkmalen zulässt. Nur, wenn das Messvorgehen und die Auswertung bestimmte Bedingungen erfüllt, können effektorientierte Fragestellungen mit einer gewissen Sicherheit beantwortet und der Einfluss äußerer Faktoren am ehesten ausgeschlossen werden (mehr dazu im nächsten Abschnitt).

Neben deskriptiven und effektorientierten Fragestellungen können auch Fragen nach dem „Wie“ einer Effekt-Kette näher untersucht werden: Durch welche Mechanismen führt ein Merkmal des Formats zu Veränderungen in der Zielgruppe? In diesem Fall wird von kausal-prozessorientierten Fragestellungen gesprochen (Fu et al., 2016, S. 14), für deren Erörterung neben der kausal-zulässigen Methodik auch sozialwissenschaftliche Modelle herangezogen werden, um Annahmen zu Wirkungsabläufen aufzustellen. In der Evaluationspraxis sind dies Forschungsfragen, die gerade im Rahmen von Begleitforschung von Interesse sind.

Bevor eine Einordnung dieser Erkenntnisse für die Evaluationsplanung erfolgt, soll auf die möglichen Evaluationsdesigns und -methoden eingegangen werden.

Evaluationsdesigns und Erhebungsmethoden



Verbunden mit der Frage, welche Daten es im Rahmen einer Evaluation zu erheben gilt, um zu aussagekräftigen und zielorientierten Ergebnissen zu kommen, ist die Auswahl eines konkreten Evaluationsdesigns und der Methode zur Datenerhebung. Ob die Evaluation zu kausalen Zusammenhängen bezüglich des Outcomes oder Impacts eines Projekts oder zu deskriptiven Aussagen bezüglich seines Outputs führen soll, hat Konsequenzen für die Gestaltung der Datenerhebung. Der Zeitpunkt der Evaluation und die zeitlichen, personellen und finanziellen Ressourcen werden ebenfalls Auswirkungen auf das Evaluationsdesign haben und die Frage eröffnen, wie komplex das Design aussehen kann bzw. ob eine leicht umsetz- und planbare Methodik ausreicht, um aussagekräftige Ergebnisse zu generieren. Auch die Entscheidung, welche am Projekt beteiligten Akteure und Zielgruppen im Zentrum der Evaluation stehen, kann die Auswahl an Erhebungsmethoden eingrenzen. In jedem Fall erfordert die Evaluationsplanung einen Abgleich von Idealvorstellungen und den bestehenden Rahmenbedingungen. Möglicherweise müssen Korrekturen vorgenommen werden, um realistische Erwartungen zu formulieren, die den verfügbaren Ressourcen gerecht werden. Der folgende Abschnitt thematisiert die möglichen Varianten, die bei dieser Entscheidung in Betracht gezogen werden können, und ihren sinnvollen Einsatz.

Evaluationsdesign

Um die Spannweite von Evaluationsdesigns veranschaulichen zu können, wird in diesem Abschnitt eine grobe Einteilung in experimentelle, quasi-experimentelle und nicht-experimentelle Evaluationsdesigns vorgenommen.

Das eine „Extrem“ stellt das „echte Experiment“ (Schnell et al., 2018, S. 200) nach sozialwissenschaftlichen Standards dar. Es erfordert zwei Messungen, eine vor und eine nach dem Erleben der Wissenschaftskommunikationsmaßnahme, sowie den Abgleich mit einer untersuchten Kontrollgruppe, die dieser Maßnahme nicht ausgesetzt war. Zentral ist dabei, dass die Untersuchungen in allen Fällen gleich ablaufen, sodass Unterschiede zwischen den Gruppen tatsächlich auf das einzig variierende Merkmal – die Auseinandersetzung mit einer Wissenschaftskommunikationsmaßnahme – zurückzuführen sind. Zudem werden die Teilnehmenden zufällig einer der beiden Gruppen zugeordnet. Auf diese Weise soll der Einfluss des Projekts auf die Gruppen bzw. ihr Erlebnis geprüft werden (ebd.). Für Impact-Evaluationen mit einer kausalen Fragestellung ist so ein Design passend, wenn beispielsweise Effekte eines Angebots auf die Einstellung oder Wissensstände der Teilnehmenden gemessen werden sollen. Die Untersuchung von Kontrollgruppen und die Sicherstellung identischer Erhebungsbedingungen ist mit hohem Aufwand verbunden. Daher verwundert es nicht, dass diese strenge Form des Experiments in der Evaluationspraxis nicht so üblich ist wie ihre Alternativen (Fu et al., 2016, S. 20).

Um den Aufwand zu reduzieren, wird auf quasi-experimentelle Designs (Schnell, Hill & Esser, 2018, S. 203) zurückgegriffen. Sie weisen ebenfalls unterschiedliche Vergleichsgruppen auf, doch oftmals wird die Zielgruppe des eigenen Projekts mit Versuchspersonen aus anderen Untersuchungen verglichen. Im Gegensatz zum echten Experiment sind die Gruppen allerdings nicht zufällig in Kontroll- und Versuchsgruppe eingeteilt. Wird ein Quasi-Experiment durchgeführt, sollten mögliche Verzerrungen im Hinterkopf behalten werden: Erstens kann es vorkommen, dass die Vergleichsgruppen strukturelle Unterschiede aufweisen, wenn sie nicht zufällig zugeteilt werden. Zweitens ist es ratsam, beim Heranziehen fremder Stichproben zu prüfen, ob die gleichen Erhebungsbedingungen wie bei der eigenen Erhebung bestehen. Ist dies nicht gegeben, so können unterschiedliche Werte bei der Messung auch beispielsweise damit zusammenhängen, wie eine



Frage formuliert, mit welcher Methode die Daten generiert oder wie das Angebot in die Messung eingebunden wurde. Drittens sollte im Fall von Vorher-Nachher-Befragungen ohne Kontrollgruppe bedacht werden, dass das Stattfinden einer Vorab-Befragung die Teilnahme an der Aktivität oder auch die spätere Erhebung prägen kann.

Diese Vorwarnungen sollen nicht die Stärken des Quasi-Experiments überdecken, denn diese Variante kann durchaus interessante Hinweise geben, welche Wirkung ein Format entfalten kann und birgt mehr Informationsgehalt als Evaluationen ohne Vergleichswerte. Sollte es zur Erklärung von Wirkungszusammenhängen herangezogen werden, ist es dennoch sinnvoll, sich bei der Interpretation der Ergebnisse möglichen Einflussfaktoren bewusst zu machen.

Zuletzt soll auf das nicht-experimentelle Evaluationsdesign eingegangen werden. Hierbei handelt es sich um eine einmalige Erhebung ohne Vergleichsgruppen. Dieses Design ist wesentlich ressourcenschonender und fokussiert sich inhaltlich auf die Beschreibung von Zuständen. Damit ist es geeignet, wenn deskriptive Fragestellungen verfolgt werden, beispielsweise zur Dokumentation der Annahme eines Projekts im Rahmen einer summativen Evaluation. Diese Formen der Evaluation mögen nicht über Wirkmechanismen aufklären, können aber dennoch spannende Erkenntnisse generieren, die künftige Projekte verbessern.

Erhebungsmethoden

Ist die Entscheidung über ein Evaluationsdesign gefallen, kann sich der Erhebungsmethode gewidmet werden. Dabei gilt es zu klären, ob Variablen direkt erhoben werden oder die Teilnehmenden selbst Auskunft geben (müssen). Die Antwort hierauf ergibt sich aus den Evaluationsobjekten: Wer verfügt über die benötigten Informationen, woher lassen sie sich beziehen oder ablesen?

Eine Möglichkeit ist es, Daten direkt zu messen. In diesem Fall sind es Projektmitarbeiterinnen und -mitarbeiter oder Externe, die die Daten eigenständig erheben und oftmals auch im Folgeschritt analysieren. Zu direkten Messinstrumenten gehören unter anderem:

- Die Webanalyse: Besonders relevant zur Evaluation von Online-Formaten, beispielsweise die Analyse von Klickzahlen, Verweildauer, eindeutige Seitenansichten etc.
- Die Inhaltsanalyse: Zur Auswertung der medialen Berichterstattung oder beispielsweise der Reaktionen und Kommentare in sozialen Netzwerken
- Die Beobachtung: Zur Erhebung leicht beobachtbarer Umstände, wie der Austausch zwischen Teilnehmenden, der Annahme oder der Umgang mit Angeboten, beobachtbarer Ausdruck von Spaß, Neugierde etc.
- Das „Tracking“: Hiermit gemeint ist in diesem Zusammenhang die Form der Beobachtung, bei der Routen innerhalb einer Ausstellung nachverfolgt werden oder die Dauer gemessen wird, mit der sich Teilnehmende mit den Angeboten auseinandersetzen

Die andere Möglichkeit wäre, die gesuchten Informationen von anderen Personen zu erfragen, die über sie verfügen. In dem Fall werden Selbstreport-Instrumente eingesetzt. Manchmal werden die Daten über den Austausch mit der Zielgruppe erhoben, in anderen Fällen übernimmt die Zielgruppe die Erhebung selbstständig durch den Einsatz von Hilfsmitteln wie Online-Surveys. Zu Selbstreport-Instrumenten zählen:



- Der Fragebogen: Zur Erhebung von inneren Prozessen wie Gedanken, Gefühle oder Einstellungen der Zielgruppe, zur Erhebung von Lernfortschritten und Wissensständen auch als Test eingesetzt; entweder auf Papier, online oder mithilfe von Apps umsetzbar
- Das Interview: In Form eines Snapshot-Interviews⁴ für eine schnelle Abfrage erster Eindrücke oder in Form eines Leitfadenterviews für eine ausführliche Auseinandersetzung mit inneren Prozessen der Zielgruppe
- Die Fokusgruppe: Zur gemeinsamen Evaluierung etwa von Maßnahmenplanungen oder Erlebnissen einer größeren Gruppe
- „Automatisierte“ On-Site Feedback-Verfahren: Evaluationen, bei denen keine weitere Unterstützung notwendig ist, sondern Teilnehmende vollkommen selbstständig agieren. Beispiele hierfür wären Graffiti-Walls, offenliegende Feedbackkarten oder sogenannte Bean-Polls, bei denen Teilnehmende mithilfe von Tokens Bewertungen abgeben⁵
- Narrativer Selbstreport: In Form von Tagebucheinträgen, Storyboardings, Foto-Essays oder narrativen Interviews (mehr dazu, siehe Constant & Roberts, 2017), die zum Ziel haben, mehr über die individuelle Auseinandersetzung mit behandelten Themen, Motivationen und Einstellungen zu erfahren

Eine Sichtung von Toolkits und Evaluationsstudien deutet darauf hin, dass Selbstreport-Instrumente wie die Befragung zu den üblichsten Methoden gehören. Wenn die Abfrage innerer, nicht beobachtbarer Prozesse bei den Zielgruppen im Vordergrund steht, ist diese Wahl angebracht und sinnvoll (Möhring & Schlütz, 2010, S. 16). Gerade bei der Verarbeitung großer Datenmengen kann eine Online-Befragung etwa finanzielle, personelle und zeitliche Ressourcen einsparen (ebd., S. 133). Trotzdem sollte bedacht werden, dass Selbsteinschätzungen in Fragebögen und Interviews verzerrt sein und Störfaktoren die Ergebnisse beeinflussen können – durch ungeschulte Interviewer (Jensen, 2014), die Befragungssituation (Jensen & Laurie, 2016, 157 ff.) oder die oftmals freiwillige Selbstrekrutierung der Teilnehmenden (Schlütz & Möhring, 2010, S. 135). Außerdem stellt sich die Frage zu stellen, ob die Untersuchungspersonen überhaupt in der Lage sind, wahrheitsgemäße Aussagen zu treffen, selbst wenn sie es beabsichtigen. So ergab eine Studie von Bromme, Thomm und Ratermann (2016), dass unter Umständen nicht jeder in der Lage ist, das eigene Sachverständnis korrekt einzuordnen.

Besteht Interesse an Informationen, die von Außenstehenden erkennbar sind, hat dies den Vorteil, dass es keiner Involvierung der Teilnehmenden bedarf – die in bestimmten Fällen auch als störend empfunden werden kann. In solchen Fällen kann der Selbstreport durch eine direkte Messung ergänzt oder ersetzt werden. Sollte beispielsweise die Sorge bestehen, dass Teilnehmende keine wahrheitsgemäßen Antworten geben, wenn sie persönlich nach ihrer Meinung zu einem wissenschaftlichen Thema befragt werden, kann vielleicht die Analyse von Reaktionen auf Online-Kampagnen ein realistisches Bild der Lage verschaffen. Direkte Messmethoden können Praktikerinnen und Praktiker vor neue Herausforderungen stellen, gerade was die Vorbereitung der Datenerhebenden angeht. Damit die Beobachtenden nach den gleichen Regeln vorgehen oder Online-Kommentare systematisch analysiert werden, müssen Regeln formuliert und Mitarbeitende geschult werden. Im Allgemeinen empfiehlt sich nach Möglichkeit eine Kombination

⁴ Zwei-minütige Interviews mit klar formulierten Fragen, mehr dazu, siehe Grand & Sardo (2017).

⁵ Mehr zu diesem Verfahren und seine Anwendung zur Evaluation, siehe Elliott & Longnecker (2013, S. 4).



verschiedenartiger Quellen und Methoden, um die Befunde durch diverse Perspektiven abzusichern und Schwächen der einen oder anderen Methode auszugleichen (Fu et al., 2016, S. 17).

Fazit und Ausblick: Erkenntnisse und Perspektiven für die Evaluationsplanung

Wie im Verlauf deutlich wurde, beginnt die Reflexion eines Projekts nicht erst mit dem Blick auf die Evaluationsdaten, sondern im Idealfall bereits bei der Anlage der Evaluation als Teil des Projektdesigns und der Projektplanung. Evaluationen stellen dementsprechend auch keinen abschließenden Projektschritt dar, sondern sind vielmehr als konstanter Lernprozess zu verstehen, der ein Projekt von Anfang an und durchgehend begleitet. Ein Prozess, der zwar einen gewissen Aufwand erfordert, allerdings eine effektive und nachhaltige Zielfokussierung verspricht. Im vorherigen Kapitel wurde eine Vielfalt von Evaluationsvarianten präsentiert, was Praktikerinnen und Praktiker vor viele Entscheidungen stellen mag. Die Auswahl der passenden Verfahren lässt sich allerdings eingrenzen, indem einige zentrale Fragen bezüglich der Ziele, Motivationen und Adressaten einer Evaluation geklärt werden (Abb. 3, S. 20).

Nicht zuletzt ist die Ausgestaltung der Evaluation auch eine Frage der Ressourcen. Fehlen die Mittel für ein Design, das kausale Schlussfolgerungen erlaubt, so liegt es nahe, die bisherigen Erwartungen zu überdenken und weniger ambitionierte Evaluationsinteressen zu verfolgen. Denn die Alternative führt zu genau dem, was einige Forschende an der bisherigen Evaluationspraxis scharf kritisieren: In solchen Fällen lassen die Ergebnisse die postulierten Schlussfolgerungen aus methodischen Gründen nicht zu. Ihre Verbreitung verzerrt nicht nur die Wahrnehmung von Wissenschaftskommunikation, sondern auch die Erwartungen gegenüber ihrer Rolle, Wirkungen, Erfolge sowie ihrer Grenzen. Abschließend stellt sich damit die Frage: Wie lässt sich die Qualität und Aussagekraft von Evaluationen in der Wissenschaftskommunikation im deutschsprachigen Raum sichern? Zunächst bedarf es dafür eine Analyse des Status Quo der üblicher Evaluationspraxis in der deutschsprachigen Wissenschaftskommunikation, um auf dieser Basis Maßnahmen zur Weiterentwicklung bedarfsgerecht zu gestalten.

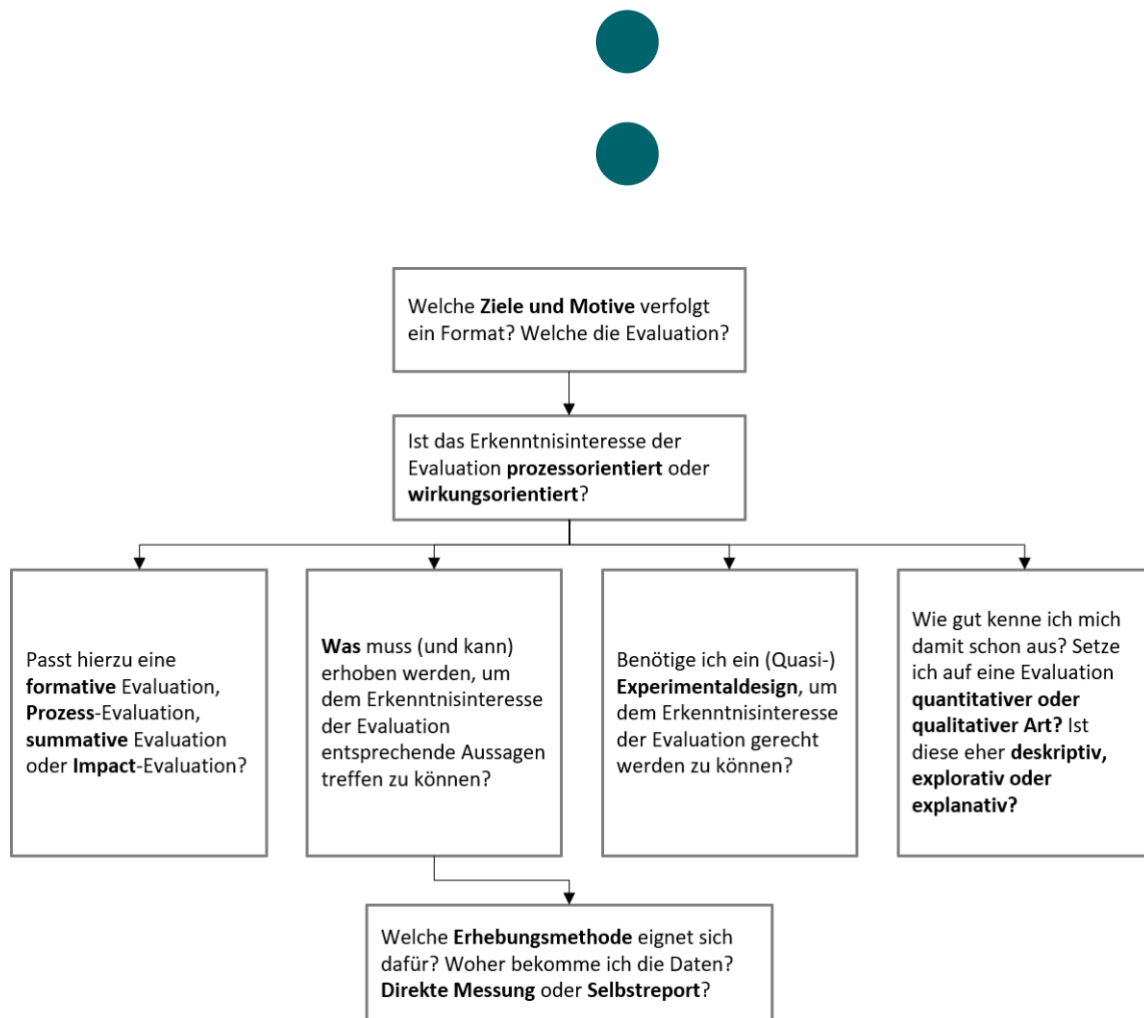


Abbildung 3: Zentrale Fragen zur Entwicklung einer Evaluation. (Eigener, vorläufiger Entwurf)

Internationale Quellen lassen zumindest erahnen, welche aktuellen Bedürfnisse bestehen: In der Praxis fehle es oftmals neben den (zeitlichen und personellen) Ressourcen besonders auch an sozialwissenschaftlicher Expertise für elaborierte und aussagekräftige Evaluationen. Auch wenn dies in einigen Fällen an der Priorisierung der Projektverantwortlichen liegen mag, die weniger Geld und Personal in die Evaluation ihrer Angebote investieren (Fu et al., 2016, S. 24), lässt sich dieser Umstand nur bedingt durch einen Kulturwandel bewältigen. Denn selbst, wenn die Bedeutung von Evaluationen hoch eingeschätzt wird, bedarf es Zeit, sozialwissenschaftliche Kenntnisse aufzubauen (Jensen, 2015, S. 2) und nicht jedes Projekt verfügt über die Mittel, ihre Evaluation von Externen übernehmen zu lassen. Folglich wird der Bedarf an standardisierten Instrumenten und Hilfsmitteln deutlich (Fu et al., 2016, S. 29; Weingart & Joubert, 2019, S. 6; Phillips et al., 2018, S. 12) – gerade in solchen Fällen, bei denen das Interesse an Evaluationen groß ist, doch gezwungenermaßen auf eigene, provisorische Instrumente zurückgegriffen wird (Phillips et al., 2018, S. 12). In diesem Zusammenhang wird auch der Wunsch nach Plattformen geäußert (The National Academies of Sciences, Engineering and Medicine, 2017, S. 9), die dem Austausch von Erfahrungen und der weiteren Professionalität des Feldes dienlich sein können. Nicht zuletzt wird auch Verbesserungspotenzial bei bestehenden Informationsmaterialien und Toolkits im englischsprachigen Raum erkannt. Zwar lassen sich Leitfäden zur Evaluation verschiedener Veranstaltungsformate und Projekte finden. Doch Hilfestellung zur Messung von Impact liefern diese den Kritikern zufolge selten (Weitkamp, 2015, S. 4). Inwieweit sich dieses Bild in gleicher Form in Deutschland abzeichnet oder ob andere Bedürfnisse an erster Stelle stehen, gilt es künftig zu betrachten.



Literaturverzeichnis

- Alonso-Flores, J. & Moreno-Castro, C. (2018). Does Science Communication Enhance Researcher Impact? A Survey among Scientists at Spanish Universities. *Journal of Education & Social Policy*, 5 (2), 34–44.
- Arts Council England (Hg.). (2020). *Generic Learning Outcomes*. <https://www.artscouncil.org.uk/measuring-outcomes/generic-learning-outcomes#section-1>
- BMBF. (2019). *Grundsatzpapier des Bundesministeriums für Bildung und Forschung zur Wissenschaftskommunikation*. https://www.bmbf.de/upload_filestore/pub/Grundsatzpapier_zur_Wissenschaftskommunikation.pdf
- Bonney, R., Phillips, T. B., Ballard, H. L. & Enck, J. W. (2016). Can citizen science enhance public understanding of science? *Public understanding of science (Bristol, England)*, 25 (1), 2–16. <https://doi.org/10.1177/0963662515607406>
- Boyette, T. & Ramsey, J. R. (2019). Does the messenger matter? Studying the impacts of scientists and engineers interacting with public audiences at science festival events. *Journal of Science Communication*, 18 (02). <https://doi.org/10.22323/2.18020202>
- British Science Association. (2020). *Impact Report and annual accounts*. <https://www.britishtscienceassociation.org/impactreport>
- Bromme, R., Thomm, E. & Ratermann, K. (2016). Wer könnte es wissen? Erklären beeinflusst die Einschätzung unseres eigenen Wissens sowie unsere Einschätzung des Wissens von Experten. *Zeitschrift für Pädagogische Psychologie*, 30 (2-3), 97–108.
- CAISE (Hg.). (2019). *Finding Funding*. <https://www.informalscience.org/projects/funding>
- Carpenter, S. L. (2015). Undergraduates' Perceived Gains and Ideas About Teaching and Learning Science From Participating in Science Education Outreach Programs. *Journal of Higher Education Outreach and Engagement*, 19 (3), 113–146.
- Constant, N. & Roberts, L. (2017). Narratives as a mode of research evaluation in citizen science: understanding broader science communication impacts. *Journal of Science Communication*, 16 (04). <https://doi.org/10.22323/2.16040203>
- Crall, A. W., Jordan, R., Holfelder, K., Newman, G. J., Graham, J. & Waller, D. M. (2013). The impacts of an invasive species citizen science training program on participant attitudes, behavior, and science literacy. *Public understanding of science (Bristol, England)*, 22 (6), 745–764. <https://doi.org/10.1177/0963662511434894>
- Elliott, J. & Longnecker, N. (2013). *Inspiring Australia Evaluation Resources*. https://www.industry.gov.au/sites/default/files/2018-10/inspiring_australia_evaluation_resources_2013.pdf?acsf_files_redirect
- Europäische Kommission (Hg.). (2018). *H2020 Programme. The EU Framework Programme for Research and Innovation*. https://ec.europa.eu/research/participants/data/ref/h2020/call_ptef/pt/2018-2020/h2020-call-pt-ria-ia-2018-20_en.pdf
- Fischhoff, B. (2019). Evaluating science communication. *Proceedings of the National Academy of Sciences of the United States of America*, 116 (16), 7670–7675. <https://doi.org/10.1073/pnas.1805863115>
- Fogg-Rogers, L., Sardo, A. M. & Grand, A. (2015). Beyond dissemination — science communication as impact. *Journal of Science Communication*, 14 (03). <https://doi.org/10.22323/2.14030301>
- Friedmann, A. J., Allen, S., Campbell, Patricia, B., Dierking, L. D., Flagg, B. N., Garibay, C., Korn, R., Silverstein, G. & Ucko, D. A. (2008). *Framework for Evaluating Impacts of Informal Science Education Projects*:

- Report from a National Science Foundation Workshop.*
https://www.informalscience.org/sites/default/files/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.
<https://doi.org/10.1080/10645578.2016.1144025>
- Gamse, B. (2010). *GK-12 Program Evaluation: Summary of Findings*. Abt Associates.
<https://www.abtassociates.com/insights/publications/presentation/gk-12-program-evaluation-summary-of-findings>
- Grand, A. & Sardo, A. M. (2017). What Works in the Field? Evaluating Informal Science Events. *Frontiers in Communication*, 2, Artikel 22. <https://doi.org/10.3389/fcomm.2017.00022>
- Inspiring Australia (Hg.). (2010). *Inspiring Australia Framework of Principles for Science Communication Initiatives*. <https://www.industry.gov.au/data-and-publications/inspiring-australia-framework-of-principles-for-science-communication-initiatives>
- Jensen, E. (2014). The problems with science communication evaluation. *Journal of Science Communication*, 13 (01). <https://doi.org/10.22323/2.13010304>
- Jensen, E. (2015). Evaluating impact and quality of experience in the 21st century: using technology to narrow the gap between science communication research and practice. *Journal of Science Communication*, 14 (03). <https://doi.org/10.22323/2.14030305>
- Jensen, E. & Laurie, C. (2016). *Doing real research: A practical guide to social research*. SAGE.
- Joubert, M. (2007). *Evaluating science communication projects*.
<https://www.scidev.net/global/communication/practical-guide/evaluating-science-communication-projects-1.html>
- King, H., Svalastog, A. L., Hobson, M., Robinson, A. & Clipson, H. (2015). Highlighting the value of evidence-based evaluation: pushing back on demands for ‘impact’. *Journal of Science Communication*, 14 (02). <https://doi.org/10.22323/2.14020202>
- Kurz, B. & Kubek, D. (2018). *Kursbuch Wirkung: Das Praxishandbuch für alle, die Gutes noch besser tun wollen*. <https://www.phineo.org/kursbuch-wirkung>. <https://www.phineo.org/kursbuch-wirkung>
- Lin, S.-F., Lin, H.-s., Lee, L. & Yore, L. D. (2015). Are Science Comics a Good Medium for Science Communication? The Case for Public Learning of Nanotechnology. *International Journal of Science Education, Part B*, 5 (3), 276–294. <https://doi.org/10.1080/21548455.2014.941040>
- Makles, A. & Schneider, K. (2016). *Evaluation der Junior Uni: Ein Update nach acht Jahren Begleitforschung*. https://www.researchgate.net/publication/311230194_Evaluation_der_Junior_Uni_Ein_Update_nach_acht_Jahren_Begleitforschung
- Möhring, W. & Schlütz, D. (2010). *Die Befragung in der Medien- und Kommunikationswissenschaft: Eine praxisorientierte Einführung* (2. Aufl.). *Lehrbuch*. VS Verlag für Sozialwissenschaften / GWV Fachverlage GmbH Wiesbaden. <https://doi.org/10.1007/978-3-531-92421-2>
- Morin, D. (2018). To debate or not debate? Examining the effects of scientists engaging in debates addressing contentious issues. *Journal of Science Communication*, 17 (04). <https://doi.org/10.22323/2.17040202>
- Nadkarni, N. M. & Morris, J. S. (2018). Baseline Attitudes and Impacts of Informal Science Education Lectures on Content Knowledge and Value of Science Among Incarcerated Populations. *Science Communication*, 40 (6), 718–748. <https://doi.org/10.1177/1075547018806909>
- The National Academies of Sciences, Engineering and Medicine. (2017). *Communicating Science Effectively: A Research Agenda*. Washington (DC). National Academies Press (US). <https://doi.org/10.17226/23674>
- National Research Council. (2009). *Learning Science in Informal Environments: People, Places, and Pursuits*. National Academies Press. <https://www.nap.edu/read/12190/> <https://doi.org/10.17226/12190>

- Phillips, T., Ferguson, M., Minarchek, M., Porticella, N. & Bonney, R. (2014). *User's Guide for Evaluating Learning Outcomes from Citizen Science*.
- 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). <https://doi.org/10.5334/cstp.126>
- Robert Bosch Stiftung (Hg.). (2020). *Hilfestellung und Dokumente für Projekte mit Beginn vor 2018*. <https://www.bosch-stiftung.de/de/hilfestellung-und-dokumente-fuer-projekte-mit-beginn-vor-2018>
- Rodgers, S., Wang, Z., Maras, M. A., Burgoyne, S., Balakrishnan, B., Stemmler, J. & Schultz, J. C. (2018). Decoding Science: Development and Evaluation of a Science Communication Training Program Using a Triangulated Framework. *Science Communication*, 40 (1), 3–32. <https://doi.org/10.1177/1075547017747285>
- Sanchez-Mora, C. (2014). Evolution of evaluation in science museums and centres. *Journal of Science Communication*, 13 (1).
- Schnell, R., Hill, P. B. & Esser, E. (2018). *Methoden der empirischen Sozialforschung* (11., überarbeitete Auflage). De Gruyter Oldenbourg.
- Science Foundation Ireland. (2015). *Evaluation Toolkit*. <https://www.sfi.ie/engagement/sfi-discover/guidance-and-best-practice/>
- Science Foundation Ireland (Hg.). (2017). *Guidance and best practice*. Science Foundation Ireland
- Southwell, B. G. & Marshall, J. (2010). Evaluation of Science Communication. In S. H. Priest (Hg.), *Encyclopedia of science and technology communication*. SAGE. <https://doi.org/10.4135/9781412959216.n101>
- Stilgoe, J., Lock, S. J. & Wilsson, J. (2014). Why should we promote public engagement with science? *Public understanding of science (Bristol, England)*, 23 (1), 4–15. <https://doi.org/10.1177/0963662513518154>
- Vitone, T., Stofer, K., Steininger, M. S., Hulcr, J., Dunn, R. & Lucky, A. (2016). School of Ants goes to college: integrating citizen science into the general education classroom increases engagement with science. *Journal of Science Communication*, 15 (01). <https://doi.org/10.22323/2.15010203>
- Watermeyer, R. (2012). Measuring the Impact Values of Public Engagement in Medical Contexts. *Science Communication*, 34 (6), 752–775. <https://doi.org/10.1177/1075547011432804>
- Weingart, P. & Joubert, M. (2019). The conflation of motives of science communication — causes, consequences, remedies. *Journal of Science Communication*, 18 (03). <https://doi.org/10.22323/2.18030401>
- Weingart, P., Pansegrau, P., Rödder, S. & Voß, M. (2008). *Bericht zum Projekt "Vergleichende Analyse Wissenschaftskommunikation"*. https://www.academia.edu/7062351/Bericht_zum_Projekt_Vergleichende_Analyse_Wissenschaftskommunikation_
- Weitkamp, E. (2015). Between ambition and evidence. *Journal of Science Communication*, 14 (02). <https://doi.org/10.22323/2.14020501>
- Wellcome Trust. (2015). *Engaging Science. Evaluating public engagement in the Wellcome Trust's UK Centres*. https://wellcome.ac.uk/sites/default/files/wtp059889_0.pdf
- Wellcome Trust (Hg.). (2020). *Reports*. https://wellcome.ac.uk/what-we-do/reports?&field_topic%5b17%5d=17&field_topic%5b16%5d=16
- Wissenschaft im Dialog. (2017). *Debatten anstoßen, Erkenntnisprozesse kommunizieren und neue Zielgruppen ansprechen: Strategiepapier 2017 der Wissenschaft im Dialog gGmbH*. https://www.wissenschaft-im-dialog.de/fileadmin/user_upload/Ueber_uns/WiD_dokumente/Strategiepapier_WiD_2017.pdf