



Evaluationspraktiken der Wissenschaftskommunikation

Eine Betrachtung veröffentlichter Evaluationsberichte im deutschsprachigen Raum

*Vorgelegt von der Impact Unit – Wirkung und Evaluation in der
Wissenschaftskommunikation*

Oktober 2020



Evaluationspraktiken der Wissenschaftskommunikation

Eine Betrachtung veröffentlichter Evaluationsberichte im
deutschsprachigen Raum

Inhaltsverzeichnis

Kurzfassung	3
Einleitung	4
Methodisches Vorgehen	5
<i>Datensammlung.....</i>	<i>5</i>
<i>Untersuchungseinheiten</i>	<i>7</i>
<i>Auswertungsverfahren</i>	<i>8</i>
<i>Kategorien der Kodierung</i>	<i>8</i>
<i>Generalisierbarkeit und Umgang mit Quellen.....</i>	<i>13</i>
Ergebnisse	14
<i>Stichprobenbeschreibung.....</i>	<i>14</i>
Welche Formate werden in der Stichprobe evaluiert?.....	14
<i>Ziele der Wissenschaftskommunikationsprojekte</i>	<i>16</i>
Welche Ziele werden in Evaluationsberichten genannt?	16
Wie werden die (definierten und erreichten) Zielgruppen beschrieben?.....	19
<i>Beweggründe der Evaluationen</i>	<i>20</i>
<i>Ausgestaltung der Evaluationen.....</i>	<i>20</i>
Welches Evaluationsdesign wird gewählt?	20
Welche Informationsträger*innen werden herangezogen?	22
Welches Erhebungsdesign wird angewendet?.....	23
Welche Erhebungsmethoden werden eingesetzt?.....	25
<i>Projektentwicklung auf Basis der Evaluationen</i>	<i>27</i>
<i>Genannte Bedarfe für bessere Evaluationen.....</i>	<i>28</i>
Zusammenfassung und abschließende Gedanken	29



Anhang.....	32
A. Literaturverzeichnis.....	32
B. Stichprobe.....	33
C. Liste der durchsuchten Stichworte.....	36
D. Ergebnisse der Community-Befragung der Impact Unit (2019)	39

Kurzfassung

Um die Qualität und Effektivität von Wissenschaftskommunikationsprojekten verbessern zu können, ist deren Evaluation von zentraler Bedeutung. Die *Impact Unit* von *Wissenschaft im Dialog* möchte daher die Entwicklung hin zu einer wirkungsorientierten, fundierten Evaluationspraxis vorantreiben. Um diesen Prozess zu unterstützen, braucht es Informationen über den Status Quo in der Evaluationspraxis. Mit der vorliegenden Analyse soll eine Bestandsaufnahme der Evaluationspraxis in der deutschsprachigen Wissenschaftskommunikation erfolgen, um deren Tendenzen, Stärken und Schwächen auszumachen.

Eine systematische Googlerecherche nach öffentlich zugänglichen Evaluationsberichten ergab eine Stichprobe von 51 Projektberichten, die 55 verschiedene Projektevaluationen behandeln. Diese 55 Projektevaluationen wurden inhaltsanalytisch ausgewertet. Von zentralem Interesse waren dabei die vorgestellten Zielsetzungen der Projekte, die Beweggründe für die Evaluation und die Ausgestaltung der Evaluation hinsichtlich der Erhebungsmethoden und Datenquellen, Erhebungs- und Evaluationsdesigns. Des Weiteren wurden Eindrücke zur Zielgruppenbeschreibung, der Projektentwicklung durch die Evaluation und der genannten Bedarfe für eine bessere Evaluationspraxis gesammelt.

Die Ergebnisse zeigen, dass die Nennung von Projektzielen sehr etabliert zu sein scheint, diese beziehen sich besonders häufig auf die Vermittlung von Informationen und die Einwirkung auf die Teilnehmenden auf kognitiver Ebene. Zielgruppen werden vor allem über Alter, Geschlecht und ihren Bildungsgrad definiert. Zu den Beweggründen für eine Evaluation gehören inhaltliche Neugierde, die Dokumentation von Erfolgen und die Effizienzprüfung des Projekts gleichermaßen. Bezuglich der genutzten Evaluationsdesign überwiegen in der Stichprobe eindeutig wirkungsorientierte Designs, bei denen es insbesondere darum geht, eine „Bilanz“ der Aktivität zu ziehen und ihren Erfolg einzuordnen. Prozessorientierte Designs treten deutlich seltener auf. Zu den üblichen Erhebungsmethoden gehören Fragebögen und Interviews, etwas seltener kommen Beobachtungen, Inhaltsanalysen oder Fokusgruppen vor. Passend dazu sind es meistens die Teilnehmenden der Aktivitäten selbst, die zur Datenerhebung herangezogen werden. Hinsichtlich der Erhebungsdesigns fällt auf, dass in den meisten Fällen einmalige Erhebungen durchgeführt werden, wobei auch Vorher-Nachher-Vergleiche stattfinden. Fälle, die zeigen, wie sich Projekte oder Aktivitäten durch Evaluationsergebnisse verändern oder verbessert werden, waren in der Stichprobe selten. An verschiedenen Stellen wurden allerdings Handlungsempfehlungen formuliert, die dafür sprechen, dass Evaluationsergebnisse zu sehr verschiedenen Verbesserungen anregen können, welche die Inhalte, Gestaltung und strategische Ausrichtung der Projekte betreffen.

Eine zentrale Erkenntnis ist, dass zwar die Zielbeschreibung, die Fokussierung auf die Teilnehmenden bei der Erhebung und die Wahl der Selbstreport-Methoden als Interesse an Wirkungen deutet werden kann. Die oftmals eingesetzten Erhebungsdesigns eignen sich jedoch nicht für die Wirkungsmessung. Zudem scheinen Wirkungsziele auf gesellschaftlicher Ebene eher legitimatorischen Zwecken zu dienen und nicht wirklich handlungsleitend zu sein, da sie häufig zwar eingangs genannt werden, aber ihre Erreichung im späteren Verlauf der Evaluation nicht resümiert wird. Des Weiteren ergeben sich Verbesserungspotenziale, die Diskrepanzen zwischen der Zielformulierung und der tatsächlichen Gestaltung der Projekte und Aktivitäten betreffen sowie den Einsatz prozessorientierter Evaluationen. Genannte Bedarfe für bessere Evaluationen beziehen sich in den Berichten auf fehlende personelle, zeitliche oder finanzielle Ressourcen.

Auch wenn diese Stichprobe sicherlich kein repräsentatives Abbild der Wissenschaftskommunikation und ihrer Evaluationspraxis darstellt, lassen sich einige Impulse für künftige Forschung, Diskurse sowie Hilfsmittel und Weiterbildungen ableiten.

Einleitung

In der Diskussion der gesellschaftlichen Bedeutung von Wissenschaft und Forschung wird der Wissenschaftskommunikation eine wichtige Funktion zugeschrieben. So lud der Forschungsausschuss des Deutschen Bundestags dieses Jahr zum öffentlichen Fachgespräch ein, um Wege zur Stärkung der Wissenschaftskommunikation zu besprechen (Deutscher Bundestag, 2020) und das Bundesministerium für Bildung und Forschung zeigte mit dem Grundsatzpapier zur Wissenschaftskommunikation (BMBF, 2019) sein verstärktes Interesse an der Professionalisierung des Feldes hierzulande. Eine solche Professionalisierung bedarf nicht zuletzt einer stärkeren Auseinandersetzung mit der Qualitätsmessung dieser Kommunikationsmaßnahmen. Zu diesem Schluss kamen auch die Expert*innen des Siggener Kreis:

„Je mehr sich die Wissenschaftskommunikation professionalisiert, je mehr Formate entstehen, sich verstetigen und je mehr Projekte gefördert werden, desto stärker wächst der Bedarf an ihrer Evaluation und an der Analyse ihrer Wirkungen.“ (Siggener Impulsschrift 2018, S. 5).

Für eine effektive Wissenschaftskommunikation braucht es eine vertrauenswürdige Beweislage, welche Praktiken bei welchen Zielgruppen eine bestimmte Wirkung entfalten. Die Schlüsselrolle kommt dabei einer gewissenhaften, auf gemeinsamen Standards beruhenden Evaluationspraxis zu.

Die *Impact Unit*, ein vom BMBF gefördertes Projekt von *Wissenschaft im Dialog*, möchte die Entwicklung hin zu einer wirkungsorientierten, fundierten Evaluationspraxis vorantreiben. Hierfür sammelt sie Daten zur aktuellen Bestandsaufnahme, befasst sich mit offenen Fragen an der Schnittstelle zwischen Forschung und Praxis, entwickelt Tools und Materialien zur Unterstützung von Praktiker*innen in ihren Evaluationsvorhaben und schafft Anlässe und Orte des Austauschs für die beteiligten Stakeholder.

Der vorliegende Bericht ist ein Teil dieser Maßnahmen. Denn um die Evaluationspraxis zu fördern und weiterentwickeln zu können, braucht es zunächst eine Analyse des Status Quo: Wo liegen die Schwerpunkte der Praxis, in welchen Aspekten ist sie fortschrittlich und wo lassen sich Optimierungspotenziale erkennen?

Nach einer ersten Recherche zu den theoretischen und konzeptionellen Grundlagen von Evaluationen in der Wissenschaftskommunikation (*Wissenschaft im Dialog*, 2020) sollte gezielt der Stand in der deutschsprachigen Region betrachtet werden. Eine Säule dieses Vorhabens war eine Online-Befragung von knapp 100 Praktiker*innen (nachfolgend bezeichnet als „Community-Befragung“), die Ende 2019 durchgeführt wurde. Die zweite Säule bildete ein Stakeholder-Workshop im vergangenen Jahr, bei dem die in der Praxis erkannten Herausforderungen und Wünsche für die Evaluationspraxis diskutiert wurden. Die nächste Säule bildet diese Analyse veröffentlichter Evaluationsberichte von Wissenschaftskommunikationsprojekten. Die Auseinandersetzung mit den Verschriftilichungen vergangener Evaluationsprojekte soll dabei eine Ergänzung zu den bisher gesammelten, subjektiven Eindrücken der Praktiker*innen darstellen.

Die Analyse betrachtet insgesamt 55 Evaluationsprojekte aus dem deutschsprachigen Raum, um übliche Praktiken und Trends zu erkennen sowie Stärken und Schwächen auszumachen. Konkret wurde eine Reihe von Fragen verfolgt, welche vornehmlich die Zielsetzungen konkreter Projekte innerhalb der Wissenschaftskommunikation, die Beweggründe für die Evaluation und die Ausge-



staltung der Evaluation (z. B. hinsichtlich der Erhebungsmethoden, herangezogenen Quellen, etc.) betreffen. Außerdem waren Hinweise auf durch die Evaluation inspirierte Projektentwicklungen von besonderem Interesse, wie auch mögliche Bedarfe für bessere Evaluationen.

Methodisches Vorgehen

Zur Bearbeitung der Evaluationsberichte wurde eine Inhaltsanalyse durchgeführt. Hierbei handelt es sich im Allgemeinen um „eine empirische Methode zur systematischen, intersubjektiv nachvollziehbaren Beschreibung inhaltlicher und formaler Merkmale von Mitteilungen, meist mit dem Ziel einer darauf gestützten interpretativen Inferenz auf mitteilungsexterne Sachverhalte“ (Früh, 2017, S. 29). Ähnlich hierzu geht es auch in diesem Vorhaben darum, aus den verfügbaren Texten erste Schlüsse auf nicht-textliche Sachverhalte, sprich die Umsetzung der Evaluationen zu ziehen und diese Schlussfolgerungen auf Basis eines möglichst transparenten, von objektiven Merkmalen geleiteten Verfahren zu treffen.

Konkret wurde angestrebt, in der Stichprobe besondere Merkmale, Schwerpunkte und Muster hinsichtlich des Evaluationsvorgehens auszumachen. Hierfür wurden den Texten durch eine systematische Kodierung Merkmale zugeordnet und deren Häufigkeitsverteilung betrachtet. Darüber hinaus wurden auffällige Zitate und Fallbeispiele gesammelt, um auch tiefere inhaltliche Einblicke und Hinweise zu sammeln. Damit orientiert sich das Vorgehen gewissermaßen an der qualitativen Inhaltsanalyse, die sich Mayring zufolge *zwischen* der qualitativen und quantitativen Denklogik positioniert: „Die Ergebnisse der Analysen werden meist quantitativ weiterverarbeitet (z. B. Kategorienhäufigkeiten), die Intercoderreliabilität spielt eine wichtige Rolle (wenn auch nicht ganz so streng wie in quantitativer Inhaltsanalyse angewandt). Die eigentliche Zuordnung von Textmaterial zu inhaltsanalytischen Kategorien bleibt aber ein (wenn auch durch inhaltsanalytische Regeln kontrollierter) Interpretationsvorgang.“ (2015, S. 8)

An dieser Stelle soll kurz auf für den Bericht relevante Begriffsverständnisse eingegangen werden: Die *Impact Unit* versteht Wissenschaftskommunikation als alle Formen von auf wissenschaftliches Wissen oder wissenschaftliche Arbeit fokussierter Kommunikation, sowohl innerhalb als auch außerhalb der Wissenschaft, inklusive ihrer Produktion, Inhalte, Nutzung und Wirkung. Allerdings wird sich die *Impact Unit* mit ihren Aktivitäten und Produkten an bestimmte Formen der Wissenschaftskommunikation richten, vornehmlich an die institutionelle, externe Wissenschaftskommunikation an eine nicht-wissenschaftliche Allgemeinheit, wobei auch medienvermittelte Kommunikationsmaßnahmen wie der Wissenschaftsjournalismus ausgeklammert werden. Dieser Fokus wird daher auch in dieser Analyse deutlich.

Eine Evaluation wird in diesem Rahmen als systematischer, projektbegleitender Lernprozess verstanden, in dem eine Kommunikationsmaßnahme über eine vorab festgeschriebene Methodik untersucht wird. Sie kann dabei verschiedene Zielorientierungen verfolgen; sei es, die direkten Leistungen und Wirkungen des Projekts zu erfassen, seine Wirkweisen zu verstehen und/oder Abläufe sowie Inhalte des Projekts zu verbessern.

Datensammlung

Um ein möglichst umfassendes Bild von der Evaluationspraxis der Wissenschaftskommunikation im deutschsprachigen Raum zu erhalten, wurde online systematisch nach relevanten Veröffentlichungen gesucht, vornehmlich nach Evaluationsberichten.

Insgesamt wurden 68 einzelne Stichwortsuchen durchgeführt, die sich aus der Kombination von zwei bis drei relevanten Suchbegriffen zusammensetzten. Als Kernbegriffe galten zum einen „Wis-



senschaftskommunikation“ und alternative Formulierungen dieser („Wissenschaft vermitteln“, „Forschung diskutieren“, u. ä.), zum anderen „Evaluation“ und verwandte Stichworte wie „Begleitforschung“ oder „Feedback“. Darüber hinaus wurden in zwei Suchanfragen testweise Erhebungsmethoden in die Suche aufgenommen („Fragebogen“, „Interviews“), was sich allerdings als weniger ertragreich für diese Recherche erwies, da letztendlich nur ein für die Analyse relevantes Dokument entdeckt wurde.

Auf Basis der Erfahrungen in der ersten Recherchephase wurden ausgewählte Formate der Wissenschaftskommunikation als relevante Suchbegriffe aufgenommen. Als Vorlage diente hierbei die Formatesammlung des Online-Portals Wissenschaftskommunikation.de¹. Nach diesen Formatbezeichnungen wurde in Kombination mit dem Stichwort „Evaluation“ gesucht – die Kombination mit alternativen Begriffen wie „Begleitforschung“ und „Feedback“ hatten sich in vorherigen Suchvorgängen als wenig ertragreich herausgestellt. Eine Liste der durchsuchten Begriffskombinationen ist im Anhang C zu finden.

Einer Verzerrung durch personalisierte Suchergebnisse wurde bestmöglich vorgebeugt. So wurde durch die Nutzung eines neuen Standardbrowsers und verschiedene Maßnahmen zum Schutz vor Aktivitätenverfolgung² sichergestellt, damit die Recherche nicht durch vergangene Suchvorgänge oder Seitenpräferenzen beeinträchtigt wird.

Für jede Suchanfrage wurden die ersten 50 Treffer berücksichtigt, exklusive der mit „Ad“ als Werbeanzeigen markierte Treffer. Relevante oder potenziell relevante Links wurden aufgerufen und auf ihre Verwertbarkeit überprüft. Treffer, die bereits bei einer vorangegangenen Suchanfrage aufgetreten waren, wurden nicht erneut berücksichtigt. Ein weiteres Kriterium zur Aufnahme in die Stichprobe war die Deutschsprachigkeit des Dokuments, oder die Durchführung des Projekts im deutschsprachigen Raum. Hinweise auf relevante Dokumente wurden mit bis zu zwei weiteren Links verfolgt. Wenn überzeugende Hinweise auf die Existenz relevanter Dokumente gefunden wurden, so wurden Ansprechpersonen identifiziert und über E-Mail kontaktiert. Wurde auf diese Anfragen mit mehreren Berichten reagiert, die sich in ihren Maßnahmen und Erhebungsmustern glichen, wurden die aktuellsten Berichte in die Analyse einbezogen. Auf diese Weise sollte einer Verzerrung der Stichprobe durch den stärkeren Einbezug vereinzelter Akteure vorgebeugt werden. In zwölf Fällen konnten durch die erwähnten Anfragen zusätzliche Dokumente gesammelt werden. In zwei Fällen wurden Zusammenfassungen der Evaluationsergebnisse mit erläuternden Kommentaren per E-Mail ergänzt. Diese zusätzlichen Informationen wurden in der Analyse berücksichtigt.

Im nächsten Schritt wurden die Dokumente genauer gesichtet. Ein notwendiges Kriterium für den Einbezug in die Stichprobe war es, dass nicht nur die Umsetzung einer Evaluation erwähnt wird, sondern auch ihr Vorgehen nachvollziehbar wird, gegebenenfalls durch Zusatzinformationen der Projektverantwortlichen. Arbeiten, die bei näherer Betrachtung nicht der vorgestellten Definition

¹ Die Formatesammlung ist unter diesem Link zu finden: <https://www.wissenschaftskommunikation.de/formate/>

² Anstelle des Standardbrowsers wurde Firefox zum ersten Mal seit der Neukonfiguration des verwendeten Laptops genutzt. In den Einstellungen des Browsers wurde der Schutz vor Aktivitätenverfolgung aktiviert, etwaige Cookies gelöscht und alle zukünftigen abgelehnt. Des Weiteren wurde das Anlegen einer Chronik deaktiviert. Im Anschluss wurde der Browser neu gestartet und für jede Sucheinheit ein neues privates Fenster geöffnet. Zudem wurde die Suchmaschine Startpage.com genutzt, die zwar auf Google-Suchmechanismen zurückgreift, aber die Identität des Nutzers dabei verbirgt.



von Wissenschaftskommunikation entsprachen, wurden aus der Analyse ausgeschlossen. Nach der Bereinigung standen 51 Dokumente zur Auswertung zur Verfügung.

Untersuchungseinheiten

Um die Bezugspunkte der Aussagen in den folgenden Kapiteln nachvollziehbar zu machen, soll kurz auf das Verhältnis der Untersuchungseinheiten in dieser Analyse eingegangen werden. Denn einige Aussagen können die betrachteten Projekte der Wissenschaftskommunikation betreffen, andere ihre Evaluation und wiederum weitere die Veröffentlichung, in der das Projekt und seine Evaluation auftauchen.

Um Aussagen über die Evaluationsstandards und -praktiken treffen zu können, dienen die 51 durch die Recherche ermittelten Berichte als das physisch vorliegende Untersuchungsmaterial. Wenn im Folgenden von „Berichten“ die Rede ist, bezieht sich dies auf die Gesamtheit der Stichprobe.

Einige der Fragen, denen in dieser Recherche nachgegangen wurde, beziehen sich auf die Merkmale der jeweiligen Wissenschaftskommunikationsprojekte (so etwa Projektziele), andere beziehen sich wiederum auf die Ausgestaltung der Evaluation der Projekte (zum Beispiel das jeweilige Design oder ihre Erhebungsmethoden). Damit bilden die behandelten Wissenschaftskommunikationsprojekte³ und die dazugehörige Evaluation innerhalb der Berichte die zentralen Einheiten zur konkreten Analyse. Die Unterscheidung zwischen Projekten und ihren Evaluationen sowie der Berichte, in denen sie zu finden sind, ist insofern von Bedeutung, da einige Berichte die Evaluation verschiedener Projekte behandeln oder aber das gleiche Projekt in verschiedenen Berichten evaluiert wird.⁴ Bezieht sich ein Bericht auf zwei voneinander unabhängige Projekte, so wurde jener Bericht dupliziert und für jedes Projekt separat kodiert. Lagen zwei aufeinander aufbauende Berichte zu dem gleichen Projekt vor, etwa eine Zwischenevaluation und eine damit verbundene Abschlussevaluation, so wurden diese beiden Berichte zusammengefasst und als ein Fall kodiert.⁵ Auf diese Weise konnten zudem Hinweise darauf, ob sich ein Projekt im Laufe der Zeit verändert hat und die Evaluation Weiterentwicklungen anregen konnte, besser in die Kodierung mit einbezogen werden.

Zur Beantwortung der in der Einleitung erwähnten Interessenbereiche ist damit die Fallzahl der evaluierten Projekte der zentrale Richtwert, auf den sich bezogen wird. Die Berichte porträtieren 55 evaluierte Projekte, die folgend analysiert werden.

³ In wenigen Fällen vereinten solche Projekte auch verschiedene Aktivitäten, die innerhalb der Evaluation untersucht werden, so etwa bei einer Großveranstaltung wie dem Wissenschaftsjahr oder einem Programm, das verschiedene Beteiligungsformate im Rahmen eines Projekts umsetzt.

⁴ Es wurde ein Sonderfall identifiziert, in dem zu *derselben* Evaluation desselben Formats zwei Berichte existierten (der ausrichtenden Organisation einerseits, und der externen, beauftragten Evaluierenden andererseits). In diesem Fall wurde ein Dokument wie gewohnt kodiert und um zusätzliche Einsichten aus dem anderen Dokument mit entsprechendem Verweis ergänzt.

⁵ Voraussetzung für dieses Vorgehen ist allerdings, dass die Evaluationen aufeinander aufbauen und keine drastischen Veränderungen vorlagen (beispielsweise durch die Beauftragung neuer Projektverantwortlicher, die ein anderes Evaluationsverständnis vertreten), die für eine Einzelkodierung sprachen, um die Merkmale jener Evaluation nicht durch die Zusammenfassung weiterer Evaluationen zu „verwischen“.

Auswertungsverfahren

Die inhaltliche Auswertung der Dokumente wurde von Mitarbeiterinnen der *Impact Unit* bei *Wissenschaft im Dialog* mit Hilfe der Analysesoftware Atlas.ti durchgeführt. Hierbei wurde auf ein vorab erstelltes Kodebuch zurückgegriffen, das zunächst deduktiv auf Basis verschiedener Hintergrundrecherchen und theoretischer Vorüberlegungen abgeleitet, im Verlauf der Auswertungsphase jedoch in einem iterativen, induktiven Prozess kontinuierlich verfeinert wurde. Zentral war hierbei der ununterbrochene Austausch der zwei Kodiererinnen und innerhalb des erweiterten Projektteams, um ein konsistentes Verständnis der Kodes, sowohl zwischen den Kodiererinnen als auch im Zeitverlauf, zu bewahren. Zu Beginn wurden zu diesem Zweck in einer Probekodierungsphase acht Dokumente von beiden Kodiererinnen zunächst einzeln kodiert und anschließend gemeinsam besprochen. Weiterhin wurden nach Besprechungen und zum Abschluss der Auswertungsphase Schleifen zur Überprüfung und Nachkodierung des bereits kodierten Materials vorgenommen.

Kategorien der Kodierung

Die folgenden Kategorien bilden die Dokumentengruppen, denen ein Bericht im Falle eines vorhandenen Merkmals zugeordnet wurde. Abgesehen von den dichotom strukturierten Formatkriterien bestand bei allen Kategorien die Möglichkeit der Mehrfachkodierung. Projekte konnten also mehrere Ziele und Beweggründe verfolgen sowie eine Kombination von Evaluationsdesigns, Erhebungsmethoden etc. aufweisen.

Formatkriterien des Projekts

Um ein besseres Bild von den evaluierten Projekten in der Stichprobe und weitere Anhaltspunkte für mögliche Gruppenvergleiche zu gewinnen, wurden den Formaten der Projekte formale, dichtochome Kriterien zugeordnet. Diese Kriterien basieren auf einem ersten Systematisierungsansatz der *Impact Unit* zur Beschreibung von Wissenschaftskommunikationsformaten:

- **Online oder offline:** Findet das Projekt in einer realen Umgebung bzw. in analogen Medien oder im digitalen Raum statt?
- **Temporär oder permanent:** Handelt es sich um ein zeitlich begrenztes Projekt oder ein langfristig bzw. dauerhaft angelegtes Projekt?
- **Groß oder klein:** Lag die Zahl der Teilnehmenden unter oder über 100 Personen?
- **Interaktiv oder nicht interaktiv:** Ist im Projektdesign ein wechselseitiger Austausch der Beteiligten vorgesehen oder besteht vielmehr ein einseitiger Kommunikationsfluss? Ein wechselseitiger Austausch kann dabei unter den Teilnehmenden oder zwischen den Teilnehmenden und wissenschaftlichen Akteuren bzw. Projektverantwortlichen stattfinden.

Während der Kodierung zeigte sich, dass die eindeutige Zuordnung der Kriterien nicht immer möglich ist. So werden beispielsweise Großevents evaluiert, die viele verschiedene Aktivitäten vereinen und innerhalb ihres Programms unterschiedliche Ausprägungen bezüglich der genannten Kriterien aufweisen. Letztendlich ließen sich die Kriterien „online oder offline“, „groß oder klein“ sowie „interaktiv oder nicht interaktiv“ am eindeutigsten zuordnen und werden zur Stichprobbeschreibung herangezogen. Diese Erkenntnis wird in der Weiterentwicklung der Systematisierung beachtet.

Ziele der Projektaktivitäten

Die Kodierung der Ziele erfolgte nach einem entworfenen Ansatz der *Impact Unit*, der vier Dimensionen zur Beschreibung der Ziele von externer Wissenschaftskommunikation vorschlägt. Ein Ziel kann dabei auf mehreren Dimensionen abgebildet und entsprechend aus verschiedenen Perspektiven betrachtet werden:

- **Wirkungsdimension:** Auf welcher Wirkungsebene ist die Zielformulierung zu verorten?
 - Outputs: direkte Leistungen (Produkte, Aktivitäten etc.) des Projekts für die Zielgruppe
 - Outcomes: Wirkungen auf Ebene der Zielgruppe
 - Impacts: Wirkungen auf Ebene gesellschaftlicher Gruppierungen über die Zielgruppe hinaus

In Anlehnung an klassische Logik-Modelle der Evaluationspraxis hat sich die Einteilung von Outputs, Outcomes und Impacts (siehe Phineo, 2018) in vielen Projekten etabliert.

- **Gestaltungsdimension:** Inwieweit sollen die teilnehmenden Personen involviert werden? Unterschieden wird zwischen:
 - Information: hiermit gemeint ist die einseitige Wissensvermittlung
 - Dialog: ein Austausch zwischen wissenschaftlichen Vertreterinnen oder Vertretern und Bürgerinnen und Bürgern
 - Beteiligung: Teilnehmende sollen aktiv in den Forschungsprozess einbezogen werden. Darunter können die Gestaltung, Planung, Durchführung, Auswertung und Interpretation von Forschungsergebnissen fallen.
- **Einstellungsdimension:** Auf welche (beobachtbaren oder innerlichen) Prozesse möchte man bei den Teilnehmenden einwirken? Unterschieden wird zwischen:
 - kognitiven Orientierungen: Einwirken auf Wissen und in die Vorstellungen über Wissenschaft und/oder wissenschaftliche Themen der Teilnehmenden
 - evaluativen Orientierungen: Einwirken auf Bewertungen und Beurteilung von Wissenschaft und ihren Themen durch die Teilnehmenden
 - konativer Orientierung: Einwirken auf Verhaltensabsichten der Teilnehmenden
 - tatsächliche Verhaltensänderungen der Teilnehmenden

Da konative Orientierungen und Verhaltensänderungen häufig bei der Zielsetzung von Projekten synonym gebraucht und betrachtet werden, wurden diese Kodierungen bei der Nachkodierung für die Analyse zusammengefasst.

- **Nutzendimension:** Wo ist der durch das Projekt erhoffte Nutzen zu verorten? Unterschieden wird zwischen:
 - persönlichen Nutzen der an der Durchführung beteiligten Forschenden (z. B. Spaß, Reputation, das Verbessern eigener kommunikativer Fertigkeiten oder, im Fall von Beteiligungsprojekten, das Sammeln neuer Erkenntnisse und Daten für die eigene Forschung)
 - institutionellen Nutzen (z. B. die verbesserte Reputation von Forschungseinrichtungen, die das Projekt durchführen)

- Wohl der Wissenschaft (z. B. die gesteigerte Anerkennung von und Wertschätzung für Wissenschaft und Forschung)
- Gemeinwohl (z. B. die Verständigung unterschiedlicher gesellschaftlicher Systeme)

Gründe für die Evaluation

Um mehr über das Erkenntnisinteresse zu erfahren, das die jeweilige Evaluation steuert, wurde die Nennung intrinsischer und extrinsischer Interessen kodiert, die mit der Evaluation verfolgt werden. Die möglichen Ausprägungen wurden aus bisherigen Erfahrungen und theoretischen Überlegungen hergeleitet. In der Literatur und verschiedenen Evaluationsguides finden sich Hinweise auf unterschiedliche Motive, die beispielsweise die Legitimation des Projekts gegenüber außenstehenden Akteur*innen (Friedmann et al., 2008; S. 8; Joubert, 2007; Sanchez-Mora, 2014, S. 1) oder die Effizienzprüfung und -steigerung (Boyette & Ramsey, 2019; Grand & Sardo, 2017, S. 1 f.; Joubert, 2007; Phillips et al., 2018, S. 5) betreffen. Unter Berücksichtigung der diversen Einsatzszenarien von Evaluationen in unterschiedlichen Fachbereichen entwarf Chelimsky (1997, S. 100) eine dreiteilige Systematisierung von generellen Motiven, die sich in der Aufteilung der intrinsischen Interessen spiegelt. Daneben sollten in der Analyse auch extrinsische Impulse bewusst ins Auge gefasst werden, die zu einer Evaluation veranlasst haben könnten. Daher wurden drei weitere Kodes ergänzt, die solche möglichen Szenarien abdecken könnten und die bei zutreffenden Fällen näher untersucht würden. Zusätzlich wurde ursprünglich der Kode „Sonstiges“ ergänzt, der aber tatsächlich nicht in Anspruch genommen wurde.

- **Intrinsische Interessen**

- inhaltliche Neugierde
- Erfolg des Projekts dokumentieren: Die Verantwortlichen möchten Daten sammeln, anhand derer sie den Erfolg des Projekts beurteilen und kenntlich machen.
- Effizienz des Projekts steigern: Die Verantwortlichen möchten den Ablauf oder Aufbau des Projekts beobachten, um diese ggf. zu optimieren und Lektionen für zukünftige Projekte zu ziehen.

- **Extrinsische Interessen**

- Vorgabe oder Impuls durch Vorgesetzte oder ähnliche Akteure: Es wird eine konkrete Person genannt, die zur Evaluation anregte.
- Förderungsbedingung: Es wird kenntlich gemacht, dass die Projektverantwortlichen durch die Förderung zur vorliegenden Evaluation verpflichtet waren.
- Im Rahmen der Begleitforschung: Das Projekt war Gegenstand eines Forschungsprojekts und jene Daten werden für eine Evaluation zweitverwertet.

Evaluationsdesign

Die *Science Foundation Ireland* (SFI) unterscheidet vier Arten von Evaluationen, die unterschiedliche Projektphasen begleiten und entweder eine prozess- oder wirkungsorientierte Ausrichtung aufweisen (2015). Auch wenn diese unabhängig voneinander durchgeführt werden können, besteht die Option, mehrere Evaluationsarten miteinander zu kombinieren bzw. in einem Bericht vorzustellen. Diese vier Arten wurden zur Einordnung des Evaluationsdesigns herangezogen:

- 
- **Formative Evaluation:** Sie findet in der Planungs- und Entwicklungsphase eines Projekts statt, verfolgt somit einen prozessorientierten Ansatz und fokussiert dabei die vorangegangenen Planungsprozesse, die Konzeptionalisierung und Strategie des Projekts. In diesem Zusammenhang prüft sie, ob die Zielsetzung und das Konzept des Formats treffend aufeinander abgestimmt sind und die formale Gestaltung – hierzu mag auch die Evaluation im weiteren Projektverlauf zählen – erfolgversprechend ist. Die Ergebnisse einer formativen Evaluation sollen idealerweise genutzt werden, um das Projektdesign gegebenenfalls frühzeitig anzupassen.
 - **Prozess-Evaluation:** Sie überwacht das bereits angelaufene Projekt in seiner Anfangsphase und prüft, ob die praktische Umsetzung so funktioniert wie angenommen. Ihre Ergebnisse sind naturgemäß prozessorientiert und dienen dazu, Probleme schnellstmöglich zu identifizieren und ihnen rechtzeitig entgegenzuwirken.
 - **Summative Evaluation:** Sie ist wirkungsorientiert und zieht eine Zwischenbilanz des Projekts während seiner Laufzeit. Ihre Ergebnisse stellen eine erste Einschätzung dazu dar, ob die vorab gesteckten Projektziele ansatzweise erreicht werden.
 - **Impact-Evaluation:** Sie ist ebenfalls wirkungsorientiert, ihr Schwerpunkt liegt allerdings in der abschließenden Phase oder im Abschluss des Projekts. Sie ermittelt, ob durch das Projekt die geplanten Leistungen erbracht wurden und die beabsichtigte Wirkung erzielt werden konnte. In diesem Zusammenhang wird in einer Impact-Evaluation auch versucht nachzuvollziehen, worin Verfehlungen von Zielen begründet liegen.

Erhebungszeitpunkt

Als Hinweis auf den Entwicklungsstand des Projekts wurde auch erhoben, zu welchem Zeitpunkt im Projektverlauf⁶ die Erhebung der Evaluationsdaten stattfand.

- **Zu Beginn des Projekts:** Eine Erhebung findet vor oder in der ersten Planungsphase statt, bevor die Aktivität „gelauncht“ wurde.
- **Während des Projekts:** Eine Erhebung findet statt, während das Projekt bereits angelaufen ist bzw. nachdem die Aktivität „gelauncht“ ist.
- **Zum / Nach Abschluss des Projekts:** Eine Erhebung findet frühestens in der Auslaufphase des Projekts bzw. seiner Aktivität statt.

Informationsträger*innen

Diese Kategorie galt der Identifikation der verschiedenen (personenbezogenen) Informationsquellen. Kodiert wurde, welche Personengruppen die Informationen für die Evaluation lieferten – unabhängig davon, ob jene Personen die Informationen persönlich an die Evaluierenden weitergeben (z. B. in einem Interview) oder über Datenträger übermitteln (z. B. indem sie ein Bild malen,

⁶ Entscheidend war bei der Kodierung des Zeitpunkts die Perspektive der Projektverantwortlichen, nicht der Teilnehmenden: Wenn ein Teilnehmer nach dem Rundgang durch eine Ausstellung befragt wurde und dies Teil einer Zwischenevaluation während der Projektlaufzeit war, würde „während des Projekts“ kodiert. Wenn eine Teilnehmerin bei der Nutzung eines Online-Formats beobachtet wurde, das allerdings noch nicht gelauncht wurde, würde „zu Beginn“ kodiert. Allerdings wurden Maßnahmen innerhalb eines Projekts als einzelne Einheiten wahrgenommen. Wenn eine Organisation regelmäßig Schülerlabore durchführt und eines dieser Labore evaluiert, würde es sich nicht automatisch um eine Erhebung „während“ des Projekts handeln, nur weil die Reihe weiterläuft. Hier musste der Kontext beachtet werden.

das inhaltsanalytisch ausgewertet wird). Dabei können sie die Zielgruppe des Projekts darstellen, bei der etwas durch die Aktivität bewirkt werden soll, sie müssen es allerdings nicht. Die vorab angelegten Kodes waren:

- **Teilnehmende Bürger*innen:** Personen, die an den Formaten teilnahmen und auch als „erreichte Zielgruppe“ beschrieben werden können
- **Mitarbeitende des Projekts:** Beispielsweise Moderator*innen in Dialog- und Beteiligungsformaten oder Ausstellungsführer*innen
- **Im Projekt involvierte Forschende:** Wissenschaftliche Akteur*innen, die im Rahmen des Projekts Wissenschaftskommunikation betreiben; häufig auch als Expert*innen beschrieben

Weitere Personengruppen wurden als „Sonstige“ erfasst.

Erhebungsdesign

Wenn im Rahmen einer Evaluation Aussagen über die durch das Projekt bewirkten Veränderungen getroffen werden sollen, ist für die Einordnung der Güte solcher Aussagen das Erhebungsdesign von zentraler Bedeutung. Um nachvollziehen zu können, wie Vergleichswerte zur Prüfung von Wirkungseffekten geschaffen werden, wurden in Anlehnung an die Studie von Fu et al. (2016) einige Merkmale ausgemacht und in folgende Kodes übersetzt:

- **Kein Experimentaldesign:** Es wird eine einmalige Erhebung ohne Vergleichswert vorgenommen.
- **Vorher-Nachher-Vergleich:** Es finden mindestens zwei Erhebungen statt, bei der die evaluierte Personengruppe vor und nach der Aktivität getestet wird.
- **Externer Gruppenvergleich:** Die generierten Daten durch die eigene Untersuchung werden mit externen Benchmarks bzw. Daten anderer Untersuchungen verglichen, um einen Vergleichspunkt herzustellen. Hierbei muss die Zuordnung der Vergleichsgruppen nicht zufällig sein.
- **Kontrollgruppen:** Es findet eine Erhebung mit einer Gruppe statt, die nicht dem Format ausgesetzt war, um einen Vergleichswert zu den Teilnehmenden zu schaffen.
- **Strenge Experimentaldesign:** Ein Experiment nach sozialwissenschaftlichen Standards erfordert 1) zwei Messungen, eine vor und eine nach dem Stimulus (dem Format), 2) eine Kontrollgruppe, die diesem Stimulus nicht ausgesetzt war und 3) eine zufällige Zuordnung der Teilnehmenden in Kontroll- oder Versuchsgruppe, um strukturelle Unterschiede zwischen den beiden Gruppen auszuschließen (Schnell et al., 2018, S. 200).

Auch hier gab es zusätzlich den Kode „Sonstiges“, falls weitere Besonderheiten im Erhebungsdesign auffielen. Evaluationen können mehrere Merkmale (beispielsweise aus verschiedenen Erhebungen mit unterschiedlichen Personengruppen) aufweisen, wobei der Kode „kein Experimentaldesign“ als Default gilt.

Erhebungsmethoden

In dieser Kategorie wird zwischen direkten und indirekten Erhebungsmethoden unterschieden. Während die Evaluierenden bei erstenen die Daten selbst erheben bzw. direkt darauf zugreifen



können, ist bei indirekten Messungen der Einbezug weiterer Personen in der Datenerhebung erforderlich, die Auskunft über die gewünschten Informationen geben. Einen Schwerpunkt in der Evaluationspraxis nehmen dabei häufig sogenannte Selbstreport-Verfahren ein, bei denen Personen ihre Situation, Eindrücke oder ihr Empfinden den Evaluierenden schildern. Sie fallen damit unter die indirekten Erhebungsmethoden. Für beide Messarten wurde jeweils ein offener Kode vergeben, um gegebenenfalls weitere Erhebungsmethoden zu identifizieren und nachträglich ins Kodebuch aufzunehmen.

- **Direkte Messung**

- Webanalyse: Eine Sammlung und Auswertung von Daten bezüglich des Verhaltens von Besucher*innen auf Websites, z. B. Klickzahlen, Verweildauer oder eindeutige Seitenansichten
- Inhaltsanalyse: Die Analyse und Auswertung verschiedener Inhalte, z. B. der medialen Berichterstattung oder Kommentare in sozialen Netzwerken zu einem Projekt
- Beobachtung: Ein mit hoher Aufmerksamkeit verbundenes, systematisches Beobachten eines Ereignisses, oftmals mithilfe von (technischen) Hilfsmitteln
- Sonstige direkte Messung

- **Indirekte Messung**

- Fragebogen: Eine Sammlung von Fragen für eine systematische, schriftliche Befragung von Personen, sowohl online als auch analog, zur Erfassung innerer Prozesse (Wissenssteigerung, Einstellungen, etc.), die von der evaluierten Personengruppe selbstständig durchgeführt wird
- Interview: Die Befragung von Personen im direkten Gespräch, übers Telefon oder über computerunterstützte Kommunikationsmittel
- Gruppengespräch: Eine geführte Diskussion in Form von Fokusgruppen oder Gruppeninterviews über ein Thema
- On-Site-Feedback: Kurze Evaluationsmethoden, bei denen Teilnehmende über Hilfsmittel (z. B. Graffiti-Walls, offenliegende Feedbackkarten oder sogenannte Bean-Polls, bei denen Teilnehmende mithilfe von Tokens Bewertungen abgeben) ein Urteil abgeben
- Narrativer Selbstreport: Eingereichte Tagebucheinträge, Bilder, Foto-Essays oder ähnliches, die zum Ziel haben, mehr über die individuelle Auseinandersetzung mit behandelten Themen, Motivationen und Einstellungen zu erfahren
- Sonstige indirekte Messung

Generalisierbarkeit und Umgang mit Quellen

Bevor die Ergebnisse der Analyse vorgestellt werden, soll vorab auf ihre Generalisierbarkeit und den Umgang mit Quellen in dieser Arbeit eingegangen werden. Die Stichprobengröße und ihre Merkmalsverteilung (hierzu mehr im ersten Kapitel des Ergebnisteils) sprechen dagegen, sie als repräsentativ für die Gesamtheit der Evaluationspraxis in der deutschsprachigen Wissenschaftskommunikation zu bezeichnen. Aus diesem Grund erhebt diese Arbeit keinen Anspruch auf wissenschaftliche Generalisierbarkeit. Trotzdem gibt die systematische Recherche Grund zur Annahme, dass die Stichprobe ein vertrauenswürdiges Abbild solcher Berichte bietet, die für Praktiker*innen des Feldes verfügbar sind, als Vorlage dienen können und den Standard der Praxis für Außenstehende repräsentieren. Sie wurde als hinreichende Grundlage befunden, um Impulse für



aktuelle Herausforderungen, Verbesserungsvorschläge und Inspirationen hinsichtlich der künftigen Leistungsbereiche der *Impact Unit* zu formulieren.

Darüber hinaus lässt die Zuordnung latenter Merkmale, beispielsweise die geschilderten Beweggründe für eine Evaluation oder Projektziele – je nach Formulierung – auch manchmal Raum für Interpretation und unterschiedliche Auslegungen. Daher sind für Leser*innen, die sich selbst ein Bild von der Stichprobe machen möchten oder Inspiration für eigene Evaluationen suchen, die öffentlich zugänglichen Quellen in alphabetischer Reihenfolge im Anhang aufgelistet. Es wurde entschieden, diese im Ergebniskapitel in anonymisierter Form zu beschreiben.

Ergebnisse

In den folgenden Unterkapiteln des Ergebnisteils werden die Häufigkeiten der kodierten Merkmale vorgestellt und teilweise um (subjektive) Eindrücke und Beispiele aus dem Untersuchungsmaterial ergänzt. Formal abgegrenzt (in den kursiven Absätzen) erfolgt eine kurze interpretative Einordnung der Ergebnisse.

Stichprobenbeschreibung

Insgesamt lagen 51 Dokumente zur näheren Untersuchung vor, in denen 55 evaluierte Projekte beschrieben werden. Sofern nicht explizit vermerkt, bezieht sich die Analyse in den nächsten Kapiteln auf jene Projekte und damit auf 55 Fälle. Der Publikationszeitraum erstreckt sich von 2000 und 2019, wobei der Großteil nach 2010 erschien. 21 Berichte wurden zwischen 2015 und 2019 veröffentlicht, weitere 18 Berichte im Zeitraum zwischen 2010 und 2014.

Im Allgemeinen kann zwischen drei Dokumenttypen unterschieden werden. Es lassen sich 29 klassische Projektberichte identifizieren, die allerdings in ihrer Ausführlichkeit stark variieren. Hinzu kommen 14 Forschungsarbeiten, vornehmlich Dissertationen, in denen Formate der Wissenschaftskommunikation vielfach entwickelt und in diesem Rahmen prozessbegleitend, in einigen Fällen auch abschließend hinsichtlich ihrer Wirksamkeit evaluiert wurden. Diese fanden häufig in Kooperation mit den Projektverantwortlichen statt, lagen allerdings hinsichtlich der Umsetzung und Verschriftlichung in der Hand der Forschenden. Acht Dokumente lassen sich weder als klassische Projektberichte, noch als akademische Forschungsarbeiten einordnen. Hierbei handelt es sich etwa um Ergebnispräsentationen oder dem Kodierteam seitens der Projektverantwortlichen bereitgestellte Erhebungsinstrumente mit Zusatzinformationen.

Welche Formate werden in der Stichprobe evaluiert?

Auf Basis der Formatkriterien ließen sich einige generelle Unterscheidungsmerkmale zwischen den evaluierten Projekten feststellen. Wie bereits bei der Vorstellung des Kategoriensystems angemerkt wurde, ergaben sich Ausnahmefälle, die nicht dem einen oder anderen Kriterium eindeutig zugeordnet werden können, weshalb bei den folgenden Vergleichen nicht die Gesamtstichprobe von 55 Projekten vertreten ist.

Zum einen befinden sich überwiegend Offline-Formate in der Auswahl. 48 Projekte konnten eindeutig als „Live“-Formate eingeordnet werden, wohingegen nur vier Projekte ausschließlich digital stattfanden. Außerdem fällt auf, dass ein Großteil der Fälle interaktive Elemente in ihr Projektdesign integriert ($n = 44$) und nur wenige Formate auf solche Elemente verzichten ($n = 7$). Was die Größe der Projekte angeht, besteht eine ausgeglichene Verteilung: 28 Projekte sind für mehr als 100 Personen ausgerichtet, 26 Projekte für weniger als 100 Personen.

Ein erster Blick auf die Stichprobe könnte eine Einteilung⁷ in folgende Formatgruppen nahelegen, in die sich zumindest 53 Projekte grob einordnen ließen:

- **Interaktive Lernorte für Kinder:** Oftmals als dritte Lernorte konzipiert, ergänzen diese Formate den formalen Schulunterricht. Sie setzen einen besonderen Schwerpunkt auf den aktiven Einbezug der Zielgruppe und lassen diese selbstständig experimentieren. Die Rekrutierung der Teilnehmenden erfolgt oft über Kitas oder Schulen, entsprechend besucht die Zielgruppe diese Formate vielfach im Klassenverband. Inhaltlich stehen sehr häufig MINT-Themen im Zentrum, die Nachwuchsförderung in naturwissenschaftlichen Fächern ist ein zahlreich genanntes Ziel der Projekte.
- **Ausstellungen:** Diese Formate werden üblicherweise von Museen betreut, können aber im Fall von Wanderausstellungen aber auch von anderen Vereinigungen umgesetzt sein. Sie richten sich manchmal gezielt an junge Zielgruppen, in anderen Fällen wird keine nähere Beschreibung der gewünschten Teilnehmenden vorgenommen oder es werden Familien als diverse, altersgemischte Zielgruppe anvisiert. Auch hier können interaktive Elemente eine wichtige Rolle spielen, doch besonders deutlich wird das Ziel, konkrete Inhalte zu wissenschaftlichen Themen über Infotafeln und ähnliche Hilfsmittel zu vermitteln.
- **Dialogveranstaltungen:** In diesen Projekten wird der Austausch zwischen Forschenden und Bürger*innen hoch priorisiert, häufig mit dem Ziel, Meinungen von Nicht-Fachkundigen zu allgemeinen wissenschaftlichen Themen zu sammeln, oft auch zu kontroversen Themen mit gesellschaftlicher Tragweite. Dabei können sie entweder für eine kleine Gruppe oder in einem großen Format angelegt sein. In der Regel richten sich diese Formate eher an Erwachsene.
- **Großveranstaltungen:** Hierunter fallen Festivals, Lange Nächte der Wissenschaft bzw. Museen und ähnliche Veranstaltungen, die im Rahmen ihres Programms häufig mehrere Wissenschaftskommunikationsmaßnahmen einschließen, allerdings diese nicht einzeln, sondern im großen Rahmen des Programms evaluieren. Aufmerksamkeit, Neugierde und Interesse für Wissenschaft zu erwecken ist hierbei ein zentrales Anliegen.
- **Vortragsformate:** Bei diesen Formaten, die sich teilweise an Kinder, teilweise an Erwachsene richten, findet eine eher einseitige Vermittlung von Inhalten durch wissenschaftliche Akteur*innen statt.
- **Digitale Formate:** Online-Spiele, Websites und Plattformen, die den selbstständigen Umgang mit wissenschaftlichen Informationen ermöglichen, dabei aber in ihrer Zielgruppenansprache und ihrer Interaktivität stark variieren.

Während die interaktiven Lernorte dieser Einteilung zufolge eindeutig den Schwerpunkt der Stichprobe ausmachen ($n = 26$), sind die anderen Kategorien in geringerer und jeweils ähnlicher Anzahl vertreten ($n = 5$ bis 7), bis auf die digitalen Formate, die durch drei Projekte vertreten sind. Dieses Ungleichgewicht sollte daher beim Lesen der Ergebnisse bedacht werden und wird im Verlauf des Berichts auch zur Einordnung herangezogen. Neben der nicht allzu hohen Fallzahl ist die ungleiche Verteilung der Formate auch ein Grund, warum auf statistische Berechnungen und Vergleichsverfahren sowie verallgemeinernde Aussagen verzichtet wird.

⁷ Wie in den Beschreibungen deutlich wird, sind die obenstehenden Formattypen nicht komplett überschneidungsfrei, sondern teilen auch wichtige Merkmale. Dies bezieht sich nicht nur auf die bereits genannten dichotomen Formatkriterien, sondern auch auf zentrale Eigenschaften wie beispielsweise die Zielgruppe eines Formats. Im Besonderen zu nennen sind dabei Formate, die sich primär oder ausschließlich an Kinder und/oder Jugendliche richten.

Ziele der Wissenschaftskommunikationsprojekte

Im Allgemeinen lässt sich zunächst feststellen, dass unter den 55 Projekten lediglich fünf Fälle auftraten, in denen kein klares Ziel benannt wurde. Dies mag allerdings auch an der Form ihrer Berichte liegen, die zu den eher unkonventionellen Dokumententypen zählen, wie etwa Abschlussberichte in Form von Präsentationen. Die Nennung von Zielen scheint grundsätzlich verankert zu sein, wobei ihre Dokumentation in der Evaluation nicht immer systematisch verläuft. In einigen Fällen werden ganz allgemein Ziele des Formats vorgestellt (z. B. von Schülerlaboren oder Museen), ohne dass noch einmal konkret darauf eingegangen wird, ob das evaluierte Projekt diesem Vorbild in jeder Hinsicht folgt oder eigene Zielsetzungen formuliert hat. Daneben wäre beispielhaft ein anderer Fall zu nennen, in dem Ziele nur beiläufig zwischen den Ergebnissen erwähnt werden, als wären sie den Leser*innen bereits bewusst. Diese nachträglichen Nennungen oder Konkretisierungen lassen allerdings keine klaren Rückschlüsse auf den Planungsverlauf des Projekts zu.

Welche Ziele werden in Evaluationsberichten genannt?

Wie sich die Häufigkeit der genannten Ziele auf die im Methodenteil beschriebenen Zieldimensionen (S. 9-10) verteilt, ist Abbildung 1 zu entnehmen.

Auf der **Wirkungsdimension** wird deutlich, dass die Nennung von Outputs keine Selbstverständlichkeit darstellt. Zwei Projekte listen ausschließlich Output-Ziele auf, in 15 Projekten werden solche Outputs nicht als Ziel in den Blick genommen. Daneben werden in 44 Projekten erwünschte Outcomes und in 25 Fällen Impacts erwähnt. Bei näherer Untersuchung der Fälle, die Impacts nannten, fällt auf, dass diese Ziele in der Einleitung Erwähnung finden, aber im weiteren Verlauf zur Überprüfung nicht unbedingt explizit wieder aufgegriffen werden.

Ziele der Wissenschaftskommunikationsprojekte

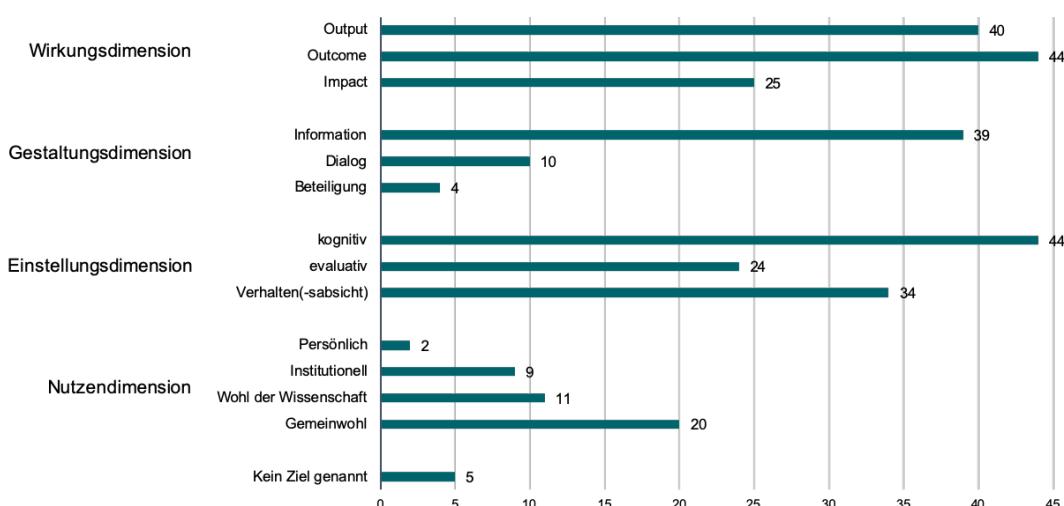


Abbildung 1. Genannte Ziele der Projekte nach den Zieldimensionen der *Impact Unit* (S. 9-10). N = 55; Mehrfachnennung möglich.

Mit Blick auf die **Gestaltungsdimension** zeigt sich, dass die einseitige Informationsvermittlung in stärkerem Maße von den Projekten als Ziel verfolgt wird als dialog- oder beteiligungsorientierte

Ansätze. Dieser Schwerpunkt wird nicht zuletzt von den zahlreichen Formaten für Kinder getragen. Unter diesen Programmen fallen immer wieder Zielaussagen, welche die Vermittlung einer „Grundbildung“ oder allgemeiner Kenntnisse über wissenschaftliche Themen thematisieren.

In der **Einstellungsdimension** zeigt sich ein deutlicher Fokus auf kognitive Orientierungen. Das Ziel, Interesse zu wecken und/oder Wissen zu erweitern, wird hier besonders betont. Betrachtet man diese Zielsetzungen im Detail, so lassen sich weitere Beispiele finden: Oftmals soll ein erster Zugang zur Welt der Wissenschaft geschaffen werden, unter anderem indem die Teilnehmenden dazu motiviert werden, sich näher mit Forschungsinhalten zu beschäftigen. Darüber hinaus gibt es Projekte, welche sich auf die Begabungsförderung konzentrieren oder ferner das Vertrauen der Teilnehmenden in die eigenen akademischen Talente verstärken möchten. Letztere Zielsetzungen werden häufig unter den in der Stichprobe vertretenen MINT-Projekten genannt. Der beachtenswerte Anteil an MINT-Projekten zeigt sich auch in den häufigen Nennungen, die Verhaltensänderungen betreffen. Viele der Projekte, die Verhalten oder im mindesten Verhaltensabsichten beeinflussen möchten, beziehen sich dabei auf die spätere Studien- und Berufswahl von Teilnehmenden. Aber auch der Wunsch, nachhaltiges Verhalten zu inspirieren, wird vereinzelt erwähnt. Das Einwirken auf evaluativen Orientierungen trat in dieser Dimension am seltensten auf. Auffällig ist im Fall evaluativer Zielsetzungen die ungerichtete Formulierung. Anstatt die Beurteilung von Wissenschaft oder wissenschaftlichen Themen in eine bestimmte Richtung lenken zu wollen, ist in den meisten Fällen davon die Rede, eine Förderung des kritischen Denkens zu beabsichtigen und Menschen dazu zu befähigen, ihre eigene Meinung zu bilden. Ausnahmen sind dabei die Verbesserung des Images von Technik oder der Abbau von Vorurteilen zu kontroversen wissenschaftlichen Themen, die allerdings nur in wenigen Fällen auftraten.

Zuletzt soll auf die **Nutzendimension** eingegangen werden: Ziele, welche *nicht* die Auswirkungen des Projekts auf die Teilnehmenden betreffen und stattdessen einen Nutzen für gesellschaftliche Gruppen, beteiligte Forschende oder die Verantwortlichen selbst fokussieren, werden tendenziell weniger genannt. Am häufigsten wurden solche Ziele gezählt, die sich auf das Gemeinwohl beziehen. In diesen Fällen soll beispielsweise durch Wissenschaftskommunikation die Schulbildung ergänzt und verbessert, Bürger*innen in förderpolitische Entscheidungen einbezogen oder auch demokratische Prozesse unterstützt werden, indem Bürger*innen zur individuellen Meinungsbildung befähigt werden. Des Weiteren nennt ein kleinerer Anteil von Projekten Ziele, die das Wohl der Wissenschaft betreffen. Bei näherer Betrachtung der kodierten Fälle fallen auch hier variierende Zielsetzungen auf: die Bedeutung von Wissenschaft und Forschung zu vermitteln, guten wissenschaftlichen Nachwuchs zu finden, die Popularität bestimmter wissenschaftlicher Felder zu steigern oder mit Case Studies zum Wissensstand in der Wissenschaftskommunikationsforschung beizutragen, um einige Beispiele zu nennen. Wie in Abbildung 1 ersichtlich, gehen nur neun Projekte auf den Nutzen für die durchführenden Organisationen ein, etwa auf die gesteigerte Aufmerksamkeit oder die langfristige Bindung der Teilnehmenden an ihre Programme, die sie sich durch das Format erhoffen. Noch seltener, in zwei Fällen, werden positive Effekte für Forschende bzw. die Wissenschaftskommunikator*innen genannt.

Neben der Auszählung verschiedener thematischer Ziele wurde abschließend ein näherer Blick auf die Formulierungen solcher Ziele geworfen. Hierbei fiel auf, dass die genannten Ziele in den seltensten Fällen auf zahlenbasierten Indikatoren beruhen. Es handelt sich eher um allgemeine Beschreibungen, wie „Bewusstsein zu schaffen“, „Interesse“ oder „Gefühle zu erwecken“, „zum Handeln anzuregen“ oder „mehr Besucher zu gewinnen“.

Zusammenfassend legt die Verteilung der Wirkungsdimension den Schluss nahe, dass die Wissenschaftskommunikation sich durchaus wirkungsorientierte Ziele setzt, prägende Erlebnisse für die Teilnehmenden schaffen sowie eine Wirkung auf gesellschaftlicher Ebene erreichen möchte. Obwohl Outputs vermeintlich am leichtesten zu erheben sind, spielen sie dabei eine vergleichsweise untergeordnete Rolle. Doch der fehlende Rückbezug im Fall der Impact-Ziele mag andeuten, dass Impacts nicht zwangsläufig genannt werden, um ihre Erfüllung zu prüfen.

Dass die Informationsvermittlung wesentlich höher priorisiert erscheint als partizipative oder dialogorientierte Ziele, mag wie bereits erwähnt mit dem beachtenswerten Anteil an Formaten für Kinder und Jugendliche zusammenhängen. Eine Vermutung wäre, dass bei jungen Zielgruppen die Informationsvermittlung eine größere Rolle spielt, da eine Wissensgrundlage für Dialoge und Beteiligung zunächst gelegt werden muss. Doch die Fokussierung der Wissensvermittlung (die sich nicht nur in der Gestaltungsdimension, sondern auch in der Einstellungsdimension in Form der stark vertretenen kognitiven Orientierung zeigt) deckt sich auch mit den Erkenntnissen der Community-Befragung, die von der Impact Unit Ende 2019 durchgeführt wurde. Unter den Antworten auf die offene Frage, welche persönlichen Ziele die Teilnehmenden in ihrer Rolle als Kommunikator*innen verfolgen, dominierten Begriffe wie „Information“ und „Wissensvermittlung“ über „Dialog“, „Austausch“ und „Diskurs“ (Abb. 2). Des Weiteren wurden die Praktiker*innen in der Befragung gebeten, aus einer Reihe von sieben Zielen jene anzukreuzen, die sie generell als am Wichtigsten erachten.⁸ Auch hier wurde die öffentliche Teilhabe an Forschung wesentlich niedriger eingeordnet (27%) als das Vermitteln von Informationen (77%).



Abbildung 2. Wordcloud der Ziellennungen in der Community-Befragung der Impact Unit (2019). Die Befragten beantworteten die Frage offen, ihnen standen dafür Felder für fünf Antworten zur Verfügung.

⁸ Es konnten maximal vier Antworten ausgewählt werden. Welche Ziele zur Auswahl standen, ist im Anhang D nachzulesen.



Zuletzt legen die Schwerpunkte der Nutzendimension den Eindruck nahe, dass unter den Projekten die externe, gemeinnützige Zielorientierung vorherrschend ist. Der interne Blick, welche Vorteile sich für beteiligte Forschende oder die Institution selbst ergeben, scheint entweder seltener von Interesse zu sein oder wird schlichtweg seltener veröffentlicht.

Die Zielformulierungen lassen vermuten, dass das Erreichen bestimmter Effektstärken bisher weniger im Fokus der Wirkungsmessung steht als das schlichte Vorhandensein von Veränderungen, da die nähere Spezifizierung ihrer Ausgestaltung meistens fehlt. Diese Neigung zu allgemein gehaltenen Zielen könnte verschiedene Gründe haben: Eine Erklärung wäre, dass in der Stichprobe qualitative Denklogiken und Methodiken dominieren, was allerdings mit Blick auf die Erhebungsmethoden und -designs, die in kommenden Kapiteln vorgestellt werden, nicht der Fall zu sein scheint. Eine andere Deutung wäre, dass Projektverantwortliche zögerlich sind, wenn es darum geht, konkrete Erfolgsversprechen abzugeben, aus Sorge, diese nicht zu erreichen – und sich mit diesem Vorgehen die nachträgliche Einordnung von Erfolgsparametern offenhalten möchten. Vielleicht fehlt es ihnen allerdings auch an Erfahrungswerten, um ihre Erwartungen in Zahlen zu übersetzen.

Wie werden die (definierten und erreichten) Zielgruppen beschrieben?

Im Zuge der Kodierung wurden zusätzlich verschiedene Formulierungen zu Zielgruppen als interessant für die Analyse erachtet und diese daher erfasst, rückwirkend auch in bereits bearbeiteten Berichten aufgenommen und ausgezählt.

Die erwünschte Zielgruppe wird offenbar vorrangig am Alter bzw. dem „Entwicklungs- oder Bildungsgrad“ der Teilnehmenden festgemacht. Der Bildungsgrad erscheint gerade bei Kindern, indem sie durch ihre Schulstufe eingeordnet werden, als relevantes Merkmal zu gelten. Doch selbst wenn Zielgruppen explizit auf konkrete Merkmale heruntergebrochen sind, werden diese Definitionen in manchen Fällen „aufgeweicht“; so etwa in einem Projekt, das seine „hauptsächliche Zielgruppe“ auf eine Sekundarstufe begrenzt.

Neben dem Alter sticht das Geschlecht als zentrale Variable hervor. Gerade im Rahmen von MINT-Förderungsprogrammen werden Mädchen wiederholt als Zielgruppe genannt. Wie in einem Projekt bei der Definition der Zieldimensionen festgestellt wird, zählen Mädchen als besondere Zielgruppe von MINT-Initiativen. In Berichten, die keine explizite Zielgruppe benennen, ist sie häufig impliziert. Ein Beispiel hierfür sind die Schülerlabore, die per Definition ihres Formats einen klaren, wenn auch allgemein gehaltenen Zielgruppenbezug aufweisen.

Unabhängig davon, ob eine Zielgruppe eingangs genannt wird oder nicht, wird immer wieder eine Beschreibung der letztendlich erreichten Personen eines Formats vorgenommen. Die Erhebung soziodemografischer Daten der Teilnehmenden wird in etwa der Hälfte der untersuchten Projektevaluationen beschrieben. Am häufigsten berücksichtigt werden hierbei das Alter ($n = 23$) und das Geschlecht ($n = 22$), die entweder über Selbstreport-Methoden oder, im Fall des zweiten Kriteriums, durch eine Beobachtung festgestellt werden.

Etwas seltener wird der formale Bildungsgrad ($n = 12$) dokumentiert, wobei eine klare Definition hin und wieder ausbleibt. So wird in einem Fall beispielsweise nur von „bildungsnah“ und „bildungsfrem“ gesprochen. Mit zu bedenken ist allerdings, dass der Bildungsgrad gerade bei den zahlreichen Kinderprojekten über die Schulstufe und Schulform einfließen kann. Explizit werden diese in sieben bzw. fünf Fällen erhoben. Der berufliche Hintergrund der Teilnehmenden oder ihrer Eltern ($n = 3$) oder ihr Wohnort bzw. der Ort der Schule ($n = 5$) werden nur in Einzelfällen be-

trachtet.

Die nähere Betrachtung der Stichprobe hinterlässt den Eindruck, dass im Fall einer expliziten Zielgruppendefinition das Erreichen eben dieser in der Evaluation auch geprüft werden soll - auch wenn die anfänglich formulierte Zielgruppe nur recht grob definiert wird. Diese Zielgruppen werden vor allem über basale Personenangaben bestimmt und eher in Ausnahmefällen über aufwendiger zu erhebende Merkmale wie etwa wissenschaftliches Interesse oder soziale Schichten festgemacht.

Beweggründe der Evaluationen

Um die Rolle der Evaluation innerhalb eines Wissenschaftskommunikationsprojekts besser zu verstehen und nachvollziehen zu können, auf welche Motive sich das Erkenntnisinteresse der Evaluation stützt, wurden Hinweise auf extrinsische und intrinsische Beweggründe für ihre Durchführung kodiert.

Extrinsische Gründe wie die Anweisung der Evaluation durch Vorgesetzte oder Fördereinrichtungen wurden in der Stichprobe nie explizit erwähnt. In sehr seltenen Fällen ließ sich trotzdem ein motivierender Impuls von außen erahnen, wenn die Evaluationen im Rahmen einer finanziellen Unterstützung des Projekts angekündigt wurden. An anderer Stelle ließ sich anhand formaler Kriterien erkennen, dass es sich um eine vorgegebene Evaluation für die Förderinstitution handelte. Unter den intrinsischen Beweggründen ließ sich kein klarer Schwerpunkt feststellen. Der Wunsch, die Umsetzung zu bewerten und (hoffentlich) den Erfolg des Projekts zu dokumentieren, wurde ähnlich häufig genannt wie die Aussicht, Verbesserungspotential zu identifizieren – vielfach treten sie gemeinsam auf. Die inhaltliche Neugierde wurde insbesondere bei wissenschaftlichen Arbeitern deutlich.

Extrinsische Motivationen scheinen den Ergebnissen zufolge keine Rolle zu spielen. Es ist allerdings naheliegend, dass in diesem Zusammenhang auch eine wahrgenommene Erwartungshaltung, welche Beweggründe eine Evaluation verfolgen sollte, Einfluss zeigt, weshalb extrinsische Argumente nicht explizit deutlich werden.

Hinsichtlich der intrinsischen Interessen „Erfolg dokumentieren“ und „Effizienz prüfen“ verfestigt sich der Eindruck, dass sie in der Praxis in einem direkten Zusammenhang gesetzt, beinahe synonym verwendet werden. Da das Interesse, das Projekt künftig zu optimieren, eine Prüfung des Gelingens und Misslingens bisheriger Ansätze gewissermaßen voraussetzt, ist diese enge Verbindung durchaus nachvollziehbar. Dass die inhaltliche Neugierde vor allem in Forschungsarbeiten in Erscheinung tritt, ließe sich dadurch erklären, dass in jenen Arbeiten die situationalen Zusammenhänge zum Zweck theorie- oder hypothesengenerierender Erkenntnisse aus einer gewissen Distanz zum Projekt im Fokus stehen.

Ausgestaltung der Evaluationen

Welches Evaluationsdesign wird gewählt?

Wie bereits beschrieben, wurde zur Einordnung des Evaluationsdesigns auf ein Modell der Science Foundation Ireland (2015) zurückgegriffen, die eine vierteilige Kategorisierung vornimmt: Die formative Evaluation und die Prozess-Evaluation (die prozessorientierten Evaluationsdesigns) sowie die summative Evaluation und die Impact-Evaluation (die wirkungsorientierten Evaluationsdesigns).

So eindeutig die Kategorisierung von Evaluationsdesigns der SFI in der Theorie ist, so schwierig gestaltete sich ihre Anwendung in der Praxis. Insbesondere eine klare Abgrenzung zwischen summariver und Impact-Evaluation ist oft nicht möglich. Beide Designs orientieren sich an der Frage nach der erzielten Wirkung eines bestimmten Formats, wenn auch mit unterschiedlichen Schwerpunkten hinsichtlich der Wirkungsdimension ihrer Ziele (da summative Evaluationen beste Voraussetzungen für die Erhebung von Outputs mitbringen und Impact-Evaluationen sich primär für die Messung von Outcomes und Impacts anbieten). Gerade bei permanenten oder regelmäßig wiederkehrenden Formaten kann nicht immer ohne Zweifel geklärt werden, ob eine Evaluation zum Ende einer Veranstaltung als Zwischenbilanz oder Gesamtbilanz des Projekts gilt. Damit bleibt diese Abgrenzung schlussendlich ungenau und ein subjektives Unterfangen.

Bezeichnungen wie summative und formative Evaluation finden zwar in einigen Berichten Verwendung, doch im Rahmen der Analyse wurde diese Terminologie nicht als ausschlaggebend für die Kodierung betrachtet, da sie nicht zwangsläufig deckungsgleich mit der Einordnung des SFI-Modells sein muss. Verwendet werden beispielsweise Umschreibungen wie die der „formative[n] Effektevaluation“ – wobei diese konkrete Evaluation im Rahmen der vorliegenden Analyse am ehesten als Kombination aus Prozess- und Impact-Evaluation zu verstehen ist. Vereinzelte Fälle lassen sich in keine der vier Designs treffend einordnen.⁹

Trotzdem ermöglicht die Kodierung des Untersuchungsmaterials anhand des SFI-Modells eine erste Orientierung in Bezug darauf, welche Funktion die Evaluationen erfüllen, welcher Zielrichtung sie folgen und wie Verteilungen weiterer Kategorien (Erhebungsdesigns und -methode etwa) einzurunden sind.

Häufigkeitsverteilung der Evaluationsdesigns

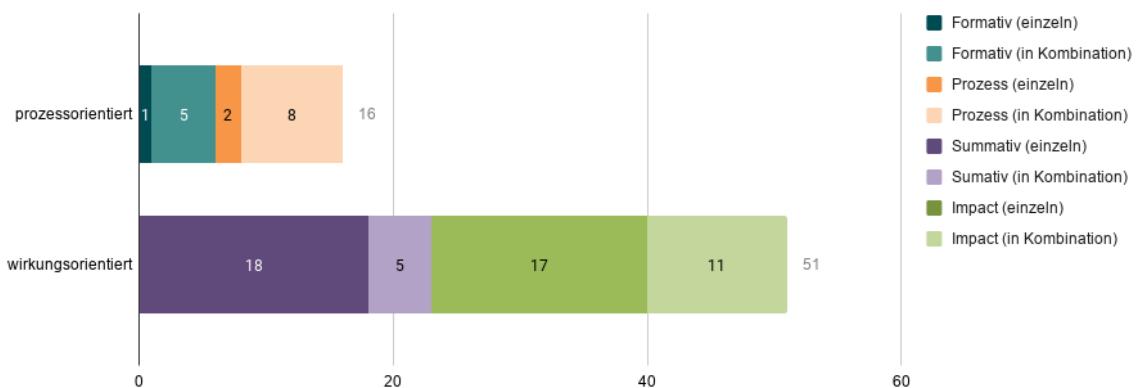


Abbildung 3. Häufigkeitsverteilung der kodierten Evaluationsdesigns, aufgeteilt in prozessorientierte und wirkungsorientierte Designs (Mehrfachkodierung möglich). Dunkel schattiert sind jene Evaluationen, die einzeln auftauchen, die helleren Flächen repräsentieren Evaluationsdesigns, die in Kombination mit weiteren Designs in einem Evaluationsprojekt auftreten. N = 55.

Mit Blick auf die Verteilung der Evaluationsdesigns (Abb. 3) wird deutlich, dass wirkungsorientierte Evaluationen wesentlich häufiger auftreten als prozessorientierte Evaluationen. Zusätzlich tau-

⁹ In einem Fall beschränkt sich das Evaluationsprojekt spezifisch auf die Arbeit der Wissenschaftskommunikator*innen. Das Format selber ist hier zwar der Kontext, nicht aber das primäre Objekt der Evaluation. In dem anderen Fall liegt die Vorstellung eines neuen Formats im Fokus. Ziel ist also, das Format eingehend zu beschreiben und charakterisieren, ohne den Anspruch auf eine Prüfung oder Bewertung des Vorgehens.

chen letztere mit wenigen Ausnahmen ($n = 3$) stets in Kombination mit einer wirkungsorientierten Variante auf. Dagegen finden sich unter den 23 als summativ eingeordneten Evaluationen nur fünf Fälle, die durch eine Prozess-Evaluation oder, in einem Fall, eine Impact-Evaluation ergänzt werden. Auch unter den Impact-Evaluationen werden weniger als die Hälfte mit einer Prozess- oder einer formativen Evaluation kombiniert.

Das größere Auftreten wirkungsorientierter Evaluationen scheint sich auch beim angegebenen Zeitpunkt der Datenerhebung bemerkbar zu machen. Evaluationen, die mit der Erhebung erst zu einem späteren Zeitpunkt im Projekt starten, sind häufiger vertreten (Abb. 4). Formative und Prozess-Evaluationen hingegen sind üblicherweise bereits früher angesetzt. Zudem fällt auf, dass formative Evaluationen vor allem im Rahmen von Forschungsarbeiten durchgeführt werden. Ein häufig auftretendes Beispiel hierfür sind Dissertationen, die Konzepte für Wissenschaftskommunikationsformate entwickeln.

Zeitpunkte der Datenerhebung(en)

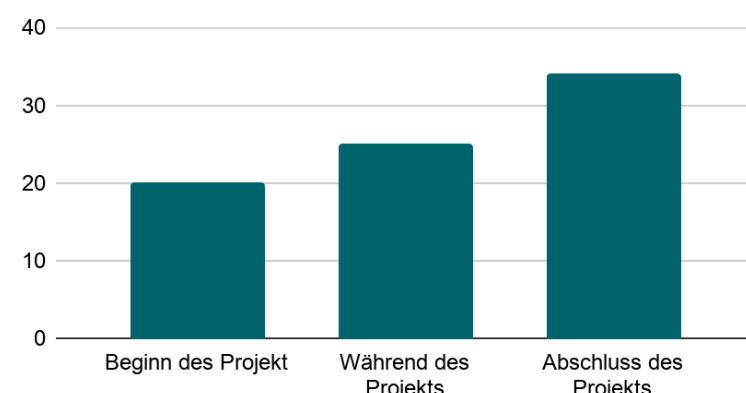


Abbildung 4. Zeitpunkte der Datenerhebungen (Mehrfachkodierung möglich). N = 55.

Ob die Verteilung der Evaluationsdesigns ein generelles Muster in der Evaluationspraxis widerspiegelt oder auch der Tatsache geschuldet ist, dass wirkungsorientierte Evaluationsberichte häufiger veröffentlicht werden, kann an dieser Stelle nicht abschließend geklärt werden. Es wäre jedoch möglich, dass die Ergebnisse prozessorientierter Evaluationen, die letztendlich der direkten Anpassung und Verbesserung des Projektes dienen, eher als interne Informationen gehandhabt werden. Evaluationen, die erzielte Wirkungen präsentieren und Erfolge dokumentieren, würden im Gegensatz dazu eher als Material für die Öffentlichkeit betrachtet.

Eine mögliche Interpretation für den auffälligen Anteil der Forschungsarbeiten unter den formativen Designs wäre, dass diese Art der Evaluation einen guten Rahmen für eine akademische Prüfung schafft, in dem Wissenschaftskommunikationsaktivitäten zunächst konzeptionell entwickelt werden, gegebenenfalls gestützt durch vorausgegangene Theorie und empirische Befunde, um anschließend in verschiedenen Testszenarien in die Praxis überführt zu werden.

Welche Informationsträger*innen werden herangezogen?

Die große Mehrheit der hier betrachteten Evaluationsprojekte beschäftigt sich mit der Untersuchung von Bürger*innen, die an den betreffenden Formaten teilgenommen haben ($n = 49$). Personen, die das Projekt mitgestalten (z. B. eingebundene Forschende oder ähnliche Beteiligte) werden ähnlich häufig als Quellen herangezogen wie diejenigen in der „Sonstigen“-Gruppe, zu der



größtenteils Begleitpersonen wie Eltern oder auch Lehrkräfte zählen. Doch auch diese Gruppen werden oft für Informationen zur möglichen Wirkung auf die teilnehmenden Bürger*innen herangezogen. So werden beispielsweise Lehrkräfte dazu befragt, wie sie die Wirkung des Formats auf ihre Schüler*innen beurteilen, und Kursleiter*innen dazu, wie die Teilnehmenden ihrer Einschätzung zufolge den Kurs erlebt haben oder welche Auswirkungen dieser wohl auf deren Wissen und Kompetenzen hinsichtlich wissenschaftlicher Themen hatte.

Tabelle 1: Häufigkeitsverteilung der Informationsträger*innen

Informationsträger*innen	n
Bürger*innen	49
Forschende	7
Mitarbeiter*innen des Projekts	15
Sonstige (Lehrkräfte, Eltern, externe Kooperationspartner*innen...)	25

N = 55; Mehrfachkodierung möglich.

Betrachtet man die „Vielfältigkeit“ der Quellen, so fokussieren sich 20 Projekte in ihrer Evaluation auf eine Personengruppe, wobei es sich dann meistens um die Bürger*innen handelt (n = 16). Mehr als zwei Gruppen von Informationsträger*innen in der Evaluation zu berücksichtigen, gehört eher zur Seltenheit. Sofern weitere Gruppen hinzugezogen werden, findet die Erhebung ihrer Daten nur einmal statt und nicht zu mehreren Zeitpunkten.

*Die Wahl der Informationsträger*innen scheint damit ein starkes Interesse der Evaluierenden an Wirkungsfragen abzuzeichnen. Im Fokus stehen dabei die Teilnehmenden des Wissenschaftskommunikationsprojekts, während weitere Quellen scheinbar zusätzliche Informationen liefern, anstatt selbst hinsichtlich auf durch das Projekt verursachte Wirkungen geprüft zu werden.*

Welches Erhebungsdesign wird angewendet?

An dieser Stelle soll auf die möglichen Designmerkmale in der Stichprobe eingegangen werden, die für einen experimentellen oder nicht-experimentellen Aufbau sprechen. Diese Unterscheidungsmerkmale sind insofern interessant, als dass sie eine bessere Einschätzung ermöglichen, ob Aussagen über die Wirkung eines Projekts auf einer ausreichenden Datenlage tatsächlich getroffen werden können. Des Weiteren zeigen sie auf, ob die erfassten Effekte entsprechend sozialwissenschaftlicher Gütekriterien und Standards mit großer Sicherheit nicht nur aus Zufall entstanden, sondern auch außerhalb der Untersuchung in der Zielgruppe zu erwarten wären und folglich, ob die Evaluationen gesicherte und aussagekräftige Erkenntnisse liefern können.

Tabelle 2: Häufigkeitsverteilungen der Erhebungsdesign-Merkmale

Design-Merkmal	Erklärung	n
Nicht-experimentell	Einmalige Erhebung einer Zielgruppe	51
Vorher-Nachher-Tests	Gleiche Zielgruppe vor und nach der Aktivität untersucht	20
Externer Gruppenvergleich	Vergleich mit einer weiteren, fremden Stichprobe	9
Kontrollgruppen	Vergleich mit Gruppe, die nicht vom Experiment betroffen	7
Strenge Experiment	Mehrere Messpunkte und randomisierte Zuordnung in Versuchs- und Kontrollgruppe	1

N = 55; Mehrfachkodierung möglich.

31 Evaluationen arbeiten *ausschließlich* mit nicht-experimentellen Designs und verzichten auf weitere Design-Merkmale. Hierbei handelt es sich meistens um eine einmalige Erhebung mit einer bestimmten Personengruppe. In 20 Projekten werden Daten vor und nach der Aktivität erhoben. Üblicherweise werden diese Fälle zusätzlich durch eine nicht-experimentelle Erhebung, etwa zur Datenerfassung mit einer weiteren Personengruppe, ergänzt.¹⁰ Von den Evaluationen, die neben nicht-experimentellen Erhebungen weitere Design-Merkmale integrieren, führen die meisten zusätzlich Vorher-Nachher-Tests durch (n = 17). Externe Gruppenvergleiche kommen insgesamt neunmal, und die Unterscheidung von Versuchs- und Kontrollgruppen siebenmal in der Stichprobe vor. Der Versuch, ein nach sozialwissenschaftlichen Standards strenges Experiment durchzuführen, wird nur in Zusammenhang mit einem Evaluationsprojekt beschrieben. Hierfür wäre neben dem Vergleich mehrerer Erhebungszeitpunkte eine zufällige Zuteilung in eine Versuchsgruppe und eine Kontrollgruppe nötig, um strukturelle Unterschiede zwischen den beiden Gruppen zu minimieren. Die restlichen Projekte mit Kontrollgruppendesign nehmen diese Einteilung ohne Zufallsprinzip vor, wobei in einem Fall zumindest eine randomisierte Zuordnung von Klassenverbänden in die eine oder andere Gruppe vorgenommen wird.

Auch unter den 28 vorliegenden Impact-Evaluationen (für deren Beurteilung Vergleichswerte eine zentrale Rolle spielen) führt ein Viertel ausschließlich einmalige Erhebungen durch. Wie in solchen Fällen Auswertungsstrategien entwickelt und Wirkungsaussagen getroffen werden, sollte daher näher betrachtet werden. Die Evaluierenden entschieden sich offenbar in jenen Projekten dafür, die Effekteinschätzung an die Teilnehmenden selbst zu übertragen: Anstatt die Unterschiede bestimmter innerlicher Prozesse, beispielsweise die Intensität des Interesses, vorher und nachher zu prüfen, wird etwa im Anschluss des Projekts auf einer Skala erfragt, ob das Interesse gestiegen

¹⁰ Auch im Fall der Erhebungsdesigns waren mehrfache Kodierungen eines Projekts möglich. Gerade bei der Unterscheidung einer einmaligen Erhebung und eines Vorher-Nach-Vergleichs möchten wir unseren Umgang kurz erläutern: Erhebungsdesign-Merkmale wurden dann mehrfach kodiert, wenn es sich um eine zusätzliche Methode handelt oder sich die verschiedenen Designs auf unterschiedliche Personengruppen beziehen. Ein Beispiel wäre, im Falle eines Schülerprogramms etwa, ein einmaliges Interview mit Lehrkräften neben einem Vorher-Nachher-Vergleich zum Lerneffekt bei Schüler*innen. Auf diese Weise sollte vermieden werden, dass jede kleinste Fragen-Unterscheidung im Pre-Post-Design als zusätzliche einmalige Erhebung gewertet wird.



sei. In einem anderen Fallbeispiel werden beobachtende Dritte zu ihrer Einschätzung hinsichtlich der Bewusstseinsänderung in der primären Zielgruppe befragt.

Während der Sichtung der Stichprobe fiel zudem ein Fallbeispiel auf, das die Herausforderungen experimenteller Vorhaben besonders veranschaulicht. Hierbei wurden nach mehreren Testdurchläufen des Formats Schüler*innen einer Klasse in einer letzten Testrunde randomisiert in Versuchs- und Kontrollgruppe aufgeteilt. Während der Versuchsgruppe die Inhalte in dem neuartigen Format vermittelt werden sollten, musste die Kontrollgruppe diese im Schulunterricht lernen. Die unbeabsichtigten Nebeneffekte des Designs, vorrangig die Enttäuschung der Teilnehmenden in der Kontrollgruppe und folglich ihre fehlende Motivation, dem Unterricht zu folgen, gefährdeten die Güte der Ergebnisse, sodass das Experiment abgebrochen werden musste. Auch die Sorge der Eltern, dass die Kontrollgruppe Nachteile tragen könnte oder auch die Frustration der Lehrer*innen, halbe Klassen unterrichten zu müssen, wurden als Hürden experimenteller Vorhaben in der realen Umgebung erkannt.

Die Stichprobe deutet damit an, dass zumindest Vorher-Nachher-Tests bereits in einigen Fällen Verwendung finden und vereinzelte Evaluationen mit Kontrollgruppen oder externen Gruppenvergleichen arbeiten. Doch der umfassenden Etablierung komplexer Erhebungsdesigns, die zuverlässige Aussagen über Wirkungszusammenhänge ermöglichen, stehen bisher auf der einen Seite pragmatische Herausforderungen im Weg, auf der anderen Seite erfordern sie Zeit, Ressourcen und Kenntnisse von den Evaluierenden, die diese bisher nicht investieren können oder möchten. Bei der Einordnung der vergleichsweise hohen Zahl an nicht-experimentellen Designs sollte bedacht werden, dass nicht alle Evaluationen zwangsläufig die Ermittlung von Wirkungen anvisieren. Doch gerade die isolierte Betrachtung der Impact-Evaluationen deutet an, dass Umgehungsversuche entwickelt werden, um Wirkungen zu erahnen, deren Aussagekraft allerdings aus Validitätsgründen nicht unbedingt vertraut werden kann.

*Die Ergebnisse der Community-Befragung (2019, Anhang D) bestätigen die Tendenz zu einmaligen Erhebungen: Nur 27 Prozent der Praktiker*innen gaben damals an, dass mindestens eines ihrer vergangenen Projekte in den letzten zwei Jahren schon einmal mithilfe einer Vorher-Nachher-Untersuchung evaluiert wurde. Sieben Prozent berichten davon, dass Versuchs- und Kontrollgruppen verwendet wurden.*

Welche Erhebungsmethoden werden eingesetzt?

Knapp die Hälfte der Evaluationsprojekte greift auf eine Kombination von zwei bis drei Erhebungsmethoden zurück. Dass mehr als drei Erhebungsmethoden eingesetzt werden, kommt in dieser Stichprobe nur bei Projekten vor, die eine Reihe an Maßnahmen umfassen, wie etwa Großveranstaltungen. In 17 Fällen wird nur eine einzige Erhebungsmethode eingesetzt, vornehmlich der Fragebogen. Dementsprechend kam in mehr als zwei Dritteln der Fälle eine Kombination mehrerer Methoden zum Einsatz.

Was die konkrete Wahl der Erhebungsmethode betrifft, so werden Selbstreport-Methoden am häufigsten eingesetzt. In 44 von 55 Evaluationsprojekten werden die Daten mittels Fragebögen erhoben. Wenn kein Fragebogen verwendet wird ($n = 11$), wird stattdessen fast durchgehend auf Interviews zurückgegriffen. Diese stellen dann auch mit 32 Projekten die zweitbeliebteste Methode dar. Die am weitesten verbreitete direkte Erhebungsmethode stellt die Beobachtung dar ($n = 21$), die gerade für größere Veranstaltungen eingesetzt werden oder wenn sich die evaluierte Gruppe nicht zur Befragung eignet (beispielsweise kleine Kinder). Die Beobachtung wird aber im

Rahmen des Untersuchungsmaterials stets in Kombination mit mindestens einer auf Selbstreport basierenden Erhebungsmethode eingesetzt. Weniger zum Einsatz kamen narrative Methoden, direkte Auszählungen, Webanalysen oder auch schnelle Feedback-Methoden (die vor Ort in Form von Feedbackkarten oder Bean Polls stattfinden können).

Tabelle 3: Häufigkeitsverteilung der Erhebungsmethoden

Erhebungsmethode	n
Fragebogen	44
Interview	32
Beobachtung	21
Direkte Auszählungen	11
Inhaltsanalyse	12
Fokusgruppen	11
Narrative Methoden	5
On-Site-Feedback	2
Webanalyse	2

N = 55; Mehrfachkodierung möglich.

*In der vorliegenden Analyse zeigen sich einige Abweichungen von den Ergebnissen der Community-Befragung (2019, Anhang D). In dieser gaben die befragten Praktiker*innen an, dass ihre Projekte nur selten oder nie mit mehr als einer Erhebungsmethode evaluiert werden. Eine mögliche Erklärung für diese Diskrepanz wäre, dass aufwendige Projektevaluationen eher in Berichten dokumentiert und öffentlich zugänglich gemacht werden und somit zu einem größeren Anteil in die Analyse einfließen, allerdings nicht unbedingt die Norm darstellen. Zudem gelten Webanalysen in der Community-Befragung nach den Fragebögen als zweithäufigste Erhebungsmethode. Dass sich diese Verteilung nicht in der Stichprobe widerspiegelt, in der nur zweimal Webanalysen eingesetzt wurden, ist wohl auch der geringen Repräsentation von Online-Formaten (n = 3) geschuldet.*

Ebenso ist es denkbar, dass die Auswertung von Web-Statistiken nicht immer als eine Form der Evaluation verstanden wird, die in einem Evaluationsbericht integriert wird. Das gleiche könnte für die On-Site-Feedback-Methoden, wie beispielsweise Bean Polls, gelten, die möglicherweise nicht als anspruchsvolle Evaluationsmethodik verstanden und als „berichtenswert“ erachtet werden. Zwar sind On-Site-Feedbackmethoden der Community-Befragung (ebd.) zufolge in der Praxis durchaus üblich und hinter der Webanalyse auf Platz drei der beliebtesten Methoden vertreten, in der Stichprobe sind sie aber nahezu ungenutzt (n = 2).

Es deutet sich damit an, dass der Methoden-Mix in der Evaluationspraxis bereits Zuspruch findet, allerdings insgesamt vornehmlich die gleichen Erhebungsmethoden, vor allem der Selbstreport, eingesetzt werden.

Projektentwicklung auf Basis der Evaluationen

Um ein endgültiges Urteil darüber fällen zu können, ob eine Evaluation einen nachhaltig positiven Effekt auf künftige Vorhaben hatte, reicht nicht der Blick in einen Evaluationsbericht, sondern bedarf es auch Dokumentationen darauffolgender Maßnahmen, um diese Entwicklungen zu prüfen. Leider ergab sich diese Möglichkeit nur in Einzelfällen, die in diesem Kapitel thematisiert werden. Daneben wurde Ausschau gehalten nach Hinweisen, dass vergangene Evaluationen eine Veränderung in den vorliegenden Projekten zu verantworten haben. Auch die Formulierung von Handlungsempfehlungen für künftige Projekte, die etwas häufiger in den Berichten behandelt wurden, wurde in Betracht gezogen.

Beste Beispiele für eine direkte Projektentwicklung innerhalb der vorliegenden Evaluationen lassen sich, wie bereits behandelt, in formativen Evaluationen festmachen. Hier können Inhalte des Formats oder Abläufe oftmals nach ersten Tests direkt für den nachfolgenden Prototypen angepasst werden. Tatsächlich ist gelegentlich von diesen Anpassungen die Rede, wenn auch nicht immer in sehr konkreter Form. In einem Beispiel, während der Entwicklung eines Schülerlabors, wird im Detail beschrieben, wie die ersten Lehrerfortbildungen direkte Verbesserungen der beinhalteten Aufgaben und Materialien nach sich zogen. Ähnliches wird in zwei weiteren Projekten berichtet, bei denen erste Fokusgruppengespräche mit der Zielgruppe, aber auch Beobachtungen im Rahmen von Pilottests positive Auswirkungen auf die Gestaltung des Projektverlaufs und Aufgabenstellungen hatte.

Doch nicht nur formative Evaluationen sind hierbei zu nennen. Ein anschauliches Beispiel ergab sich im Fall einer langfristigen Museumsausstellung, zu der zwei Evaluationsberichte im Abstand eines Jahres vorlagen. Es zeigt auf, wie auf Basis der Ergebnisse nicht nur Methoden der späteren Evaluation, sondern auch konkrete Elemente der Ausstellung neue Formen annahmen. So wurden auf Basis der ersten Befragungen Veränderungen in der Ausstellungsroute und ihrer Beschriftung vorgenommen, um auf die Kritik der Befragten einzugehen, welche die fehlende Orientierung im Museum bemängelten. Zudem wurden die Interviewleitfäden entsprechend angepasst, um die umgesetzten Veränderungen gezielt auf ihre Effektivität hin zu prüfen. Auf den Erkenntnissen der Evaluation aufbauend wurden also nicht nur Verbesserungen am Format, sondern auch Anpassungen an der Evaluationsmethodik vorgenommen.

Wie eine Entwicklung auf noch längere Sicht aussehen kann, zeigt das Beispiel einer Kinder-Uni, zu der ein Bericht nach vier Jahren und ein Bericht nach acht Jahren vorlag. Mit dem Ziel, neben dem praktischen Beitrag für die Wissenschaftsvermittlung auch zur wissenschaftlichen Begleitforschung beizutragen, wurden hier unabhängige Begleitforschungsprojekte etabliert. In diesen Berichten werden zahlreiche Maßnahmen benannt, zu denen die Evaluationsergebnisse angeregt hätten. Diese betrafen nicht nur strategische Belange wie den Ausbau von Kooperationen und Kommunikationsmaßnahmen zur Bewerbung der Programme, sondern auch die inhaltliche Ausweitung künftiger Projekte.

Wie eingangs angesprochen sind derartige Hinweise auf eine Projektentwicklung im Untersuchungsmaterial eher selten. Nichtsdestotrotz spricht eine Vielzahl der Evaluierenden Handlungsempfehlungen aus, die sich auf künftige Projekte beziehen. Diese variieren allerdings deutlich in ihrem Detailgrad und werden manchmal als offene Vorschläge, manchmal aber auch als fest vorgenommene Veränderung präsentiert: So etwa in einer Forschungsarbeit, dessen Autorin zu dem Schluss kommt, dass „aufgrund der durchgeföhrten Befragung Handlungsbedarf bei der Anleitung zur Vor- und Nachbereitung der Angebote im Unterricht [besteht], speziell bei Schülern der Sekundarstufe I und Schülern der Sekundarstufe II ... , sowie bei der Einbindung der Inhalte und Kontexte in den Unterricht, die durch vorherige Absprachen mit den entsprechenden Lehrern stärker forcierter werden muss.“



Thematisch beziehen sich diese Handlungsempfehlungen häufig auf inhaltliche Aspekte wie die Themensetzung und -darstellung, die Ausgestaltung von Materialien, neue Programmideen oder auch den richtigen Umgang mit den Teilnehmenden in partizipativen Formaten. Des Weiteren werden formale Verbesserungen angeregt, etwa bezüglich der Strukturierung des Programms, der reibungslosen und verständlichen Umsetzung digitaler Formate, der Einbindung von Partner*innen sowie Teilnehmenden oder auch schlichtweg organisatorische Details wie die Wahl neuer Veranstaltungsorte. An einigen Stellen werden auch strategische Handlungsbedarfe formuliert, die zum Beispiel die Verwertung der Ergebnisse und die Kommunikation und Bewerbung der Angebote betreffen.

Hinsichtlich der Frage, inwieweit Evaluationen zur Projektentwicklung anregen, kann aufgrund der begrenzten Datenlage aufeinanderfolgender Dokumentationen kein definitives Urteil gefällt werden. Es zeichnet sich allerdings ab, dass Evaluationen zu sehr verschiedenen inhaltlichen, gestalterischen und strategischen Verbesserungen beitragen können, nicht zuletzt durch die erkannten Handlungsempfehlungen in diversen Berichten. Zudem deutet sich anhand der Fallbeispiele an, dass eine früh einsetzende Evaluation bereits viele zunächst unerwartete Verbesserungspotenziale erkennen lässt, bevor das Programm offiziell „gelauncht“ wird.

Auch in der Community-Befragung (2019, Anhang D) wurde die Rolle von Evaluationen zur Verbesserung künftiger Vorhaben thematisiert. Hier waren nur 39 Prozent der Befragten der Meinung, dass ihre Daten in die Projektentwicklung einfließen, während 23 Prozent unentschlossen waren. 12 Prozent machten keine Angabe an dieser Stelle. Es zeigte sich also bei den Befragten eine gewisse Unsicherheit, ob die Evaluationen einen Beitrag zur Weiterentwicklung der entsprechenden Projekte leisten. Die Wahl der Quellen in der vorliegenden Analyse sollte bei der Einordnung der optimistischen Wahrnehmung von Evaluationen folglich nicht unterschätzt werden. Es ist erwartbar, dass in öffentlich zugänglichen Evaluationsberichten eine positive Einschätzung des Potenzials von Evaluationen vermittelt wird, die den allgemeinen Enthusiasmus innerhalb der Praxis übersteigt.

Genannte Bedarfe für bessere Evaluationen

Zu guter Letzt stellt sich die Frage, wie sich die Evaluationspraxis unterstützen lässt, um ihre Qualität weiter zu steigern. Gedanken aus dieser Makroperspektive werden nur selten in Evaluationsberichten geäußert. In vereinzelten Fällen lassen sich allerdings Hinweise darauf finden, was die Evaluierenden als Einschränkungen wahrnehmen. Hier stehen eindeutig Ressourcenmängel im Vordergrund, entweder zeitlicher, finanzieller oder personeller Art.

So wurde in einem Bericht erwähnt, dass der ausführliche Datensatz nur teilweise ausgewertet werden konnte, da eine „systematische Auswertung den Budget- sowie den Projektrahmen von zwei Jahren sprengen würde“. In einem weiteren Projekt, in dem bis zur Projekteröffnung mehrere Evaluationsphasen geplant waren, stellten die Verantwortlichen bei der Vorbereitung ihrer summativen Evaluation fest, dass „der Umfang und die Qualität der Evaluierung durch die Verfügbarkeit betreuender Personen begrenzt ist.“ – Personen, die in diesem Fall unter anderem Tracking-, Beobachtungs- und Befragungsdaten auswerten sollten. Um diesem Problem zu begegnen, wurden zwei Maßnahmen ergriffen: Zum einen die Etablierung einer Tracking-App, welche die Auswertung beschleunigen würde und zum anderen die Zusammenarbeit mit einem universitären Institut, das die Erhebung und Datenauswertung nach wissenschaftlichen Kriterien ermöglichen würde und „über das nötige methodische Wissen verfüge“, das in der Praxisgemeinde vielerorts fehlt.



Doch auch im Rahmen von Forschungsarbeiten nennen die Evaluierenden die Herausforderungen durch Zeit- und Personalbegrenzungen – erst Recht, wenn eine theoretische Vorarbeit der Erhebung vorangeht.

*Insgesamt bestätigen diese Fälle jene Bedarfe, die auch in anderen Zusammenhängen von Praktiker*innen geäußert werden. Bei der Community-Befragung kamen die gleichen Mängel zur Sprache. Auf die gezielte Frage, durch welche Art von Unterstützung die Aussagekraft von Evaluation verbessert werden könnte, schlügen die Teilnehmenden der Community-Befragung (2019, Anhang D) Lösungsansätze vor, die entweder die Entwicklung von Hilfsmitteln und Tools betrafen, den Zugang zu Expertise und Informationsmaterialien oder Veränderungen, die sich auf kulturelle Prägungen der Evaluationspraxis beziehen, namentlich eine weniger starke Fixierung auf Metriken oder eine Entbürokratisierung des Berufsfelds.*

Zusammenfassung und abschließende Gedanken

Wie bereits erläutert wird die Stichprobe aufgrund der begrenzten Datenlage und ihrer Format-schwerpunkte nicht als repräsentatives Bild der deutschsprachigen Evaluationspraxis verstanden. Trotzdem werden einige Tendenzen deutlich, welche die *Impact Unit* als besonders interessant erachtet: In den allermeisten Fällen bezieht sich eine Evaluation auf definierte Projektziele. Diese beziehen sich sehr häufig auf Veränderungen bei den Teilnehmenden bzw. Bürger*innen und setzen einen Schwerpunkt auf die Informationsvermittlung. Zudem fällt auf, dass Zielgruppen am ehesten anhand ihres Alters, Geschlechts und Bildungsstands charakterisiert werden. Hinsichtlich der Beweggründe für die Evaluation werden keine extrinsischen Motivationen genannt, dafür tauchen Gründe wie inhaltliche Neugierde, die Dokumentation von Erfolgen sowie die Effizienzprüfung des Projekts in ähnlicher Häufigkeit auf. Passend zu der verstärkten Wirkungsorientierung in den Zielsetzungen sind auch die wirkungsorientierten Evaluationsdesigns wesentlich häufiger vertreten als prozessorientierte Evaluationsdesigns. Zudem werden überwiegend die Teilnehmenden an den Projekten als Informationsträger*innen zur Datenerhebung herangezogen, wobei auch deren Begleitpersonen und die Mitarbeiter*innen des Projekts immer wieder einbezogen werden. Auch wenn Vorher-Nachher-Vergleiche hin und wieder eingesetzt werden, werden mehrheitlich nicht-experimentelle Erhebungsdesigns mit einmaliger Datenerhebung durchgeführt. Selbstreport-Verfahren wie Fragebögen und Interviews werden am meisten eingesetzt, erst danach folgen Beobachtungen und Inhaltsanalysen als direkte Messmethoden. Auffällig ist außerdem, dass in zwei Dritteln der Fälle mehrere Erhebungsmethoden eingesetzt werden.

Insgesamt lassen sich also einige positive Merkmale feststellen, gerade was die generelle Nennung von Zielen, das rege Interesse an Wirkungen auf Individual- oder auch gesellschaftlicher Ebene und den Einsatz kombinierter Erhebungsmethoden betrifft. Dennoch lassen sich auf Basis der betrachteten Fälle einige aktuelle Herausforderungen feststellen, deren Verbesserung im Sinne einer wirkungsorientierten und aussagekräftigen Evaluationspraxis besonders relevant erscheint.

Ein solcher Optimierungsbedarf wird in der **Etablierung von tatsächlich wirkungsorientierten Erhebungsdesigns** deutlich. Um gesicherte Aussagen über die Wirkung von Wissenschaftskommunikation zu generieren, sind gerade bei quantitativen Erhebungen Vergleichswerte notwendig. Die Einschätzung der erreichten Wirkung rückwirkend den Teilnehmenden oder beobachtenden Dritten zu überlassen, wie es in einigen Fällen vorkam, lässt im Gegensatz zum Vorher-Nachher-Vergleich viel Spielraum für subjektive Abweichungen und



Verzerrungen und ist daher nicht empfehlenswert. Auch wenn die pragmatischen Gründe für die Umgehung komplexer Erhebungsdesigns in der Analyse deutlich wurden, ist die Weiterentwicklung der quantitativen Erhebungsstandards innerhalb der Evaluationspraxis durch das Heranziehen von Vergleichswerten von zentraler Bedeutung. Vorher-Nachher-Vergleiche sind hierfür die ideale Variante, doch auch alternative Methoden wie etwa externe Vergleichsgruppen aus ähnlichen Projekten könnten künftig die Ergebniseinordnung bereichern. Eine weitere Stärkung der Aussagekraft bzw. Absicherung der Ergebnisse könnte erreicht werden, wenn die **Kombination verschiedener Erhebungsmethoden**, insbesondere die Ergänzung der Selbstreport-Methoden um direkte Messungen, weiter etabliert wird.

Zudem zeigte sich bei der Betrachtung der Zieldimensionen, dass die genannten Ziele nicht geradlinig auf die Ausgestaltung der Projekte schließen lassen. Besonders deutlich wird dies bei der vergleichsweise seltenen Nennung von dialog- oder partizipationsorientierten Zielen und dem verstärkten Einsatz interaktiver Elemente, den die Verteilung der Formatkriterien andeutet. 44 Formate integrieren interaktive Elemente in ihr Projektdesign, doch in nur 14 Evaluationen werden Dialog und Beteiligung als Ziele konkret deutlich. Der bemerkenswerte Anteil interaktiver dritter Lernorte in der Stichprobe, die allerdings informierende Ziele in den Mittelpunkt ihrer Projekte stellen, wird zu diesem Verhältnis beigetragen haben. Die Verteilung wirft das Licht auf die verschiedenen Funktionen solcher Gestaltungselemente wie der Interaktion (etwa im Lernprozess), stellt allerdings auch die Frage in den Raum, welche weiteren Faktoren neben der Zielseitung des Projekts in die strategische Projektplanung einfließen – und inwieweit Trends in der Wissenschaftskommunikation, der Spaß am Programm oder Besonderheiten der Zielgruppe mit jenen Projektzielen in diesem Prozess konkurrieren. **Eine bewusste Auseinandersetzung mit der Überführung von Projektzielen in die strategische Projektplanung**, auch im Sinne der Ressourcenschonung, erscheint daher angebracht.

Ein weiterer Punkt, der den Umgang mit Zielen betrifft, bezieht sich auf die Funktion von Impact-Zielen in den Evaluationsberichten, die zwar anfangs erwähnt, aber im Verlauf des Berichts vielfach nicht konkret weiterverfolgt oder reflektiert werden. Die Prüfung von nachhaltigen, gesellschaftlichen Effekten und ihre Rückführung auf ein isoliertes Projekt erweist sich als äußerst schwierig, was die fehlende Nachverfolgung erklären mag. Scheinbar werden Impact-Ziele dennoch formuliert, um die Vision des Projekts zu verdeutlichen, seine Gemeinnützigkeit zu betonen und die notwendige Relevanzbegründung zu liefern. Um solche Spekulationen zu vermeiden, wäre eine **explizite Trennung von Zielen, die zur Einordnung von Erfolgen und Misserfolgen genutzt werden und Aussagen, die eher visionären Charakter haben**, sinnvoll, um Missverständnisse hinsichtlich der abschließenden Beurteilung von Projekten zu vermeiden. Denn solche Fehlinterpretationen können in der Diskussion über die Potenziale und Leistungen von Wissenschaftskommunikation eher hinderlich als förderlich sein.

Des Weiteren ließen sich nur vereinzelte Beispiele für prozessorientierte Evaluationen finden. An dieser Stelle bedarf es möglicherweise noch weiterer Studien, welchen Stellenwert formative Evaluationen und Prozess-Evaluationen in der Wissenschaftskommunikation einnehmen und inwie weit die niedrigeren Zahlen in der Stichprobe mit der Fokussierung auf veröffentlichte Evaluationsberichte zusammenhängen. Unabhängig der Gründe für ihre geringe Repräsentation wären weitere Fallbeispiele in diesem Bereich interessant. Die vereinzelten vorliegenden Fälle zeigen zumindest auf, welche Vorteile es mit sich bringt, **frühzeitig einsetzende Evaluationen** im Verlauf eines Wissenschaftskommunikationsprojekts zu integrieren: Sie können einen Beitrag dazu leisten, dass die Projekte wirksamer und effizienter gestaltet werden, da Probleme und Verbesserungspotenziale früher erkannt werden.

Die **Ermutigung, mehr Evaluationen zu veröffentlichen**, soll sich nicht nur an Projekte mit bestimmten Evaluationsdesigns richten. In der Recherche für diese Analyse wurden 3400 Google-



Treffer betrachtet. Letztendlich waren über die Stichworte „nur“ 55 Evaluationen in der Wissenschaftskommunikation auffindbar, die den angemessenen Detailgrad für die Analyse mitbrachten und unterschiedliche Projekte abbildeten. Um die Fortschritte der Evaluationspraxis nachvollziehen und nicht zuletzt voneinander lernen zu können, ist die Transparenz hinsichtlich der eigenen Praktiken ein wichtiger erster Schritt. Ein Dank geht an dieser Stelle an jene, die ihre Evaluationen bereits online gut auffindbar bereitstellen sowie die Ansprechpartner*innen, die dieses Projekt durch weitere Evaluationsberichte und Informationen bereichert haben, und damit diese Offenheit vorleben.

Dies sind nur einige Aspekte, die mit Blick auf die Ergebnisse aktuell besonders notwendig erscheinen. Auf längere Sicht ließe sich weiter diskutieren, wie Zielgruppen über basale Informationen hinaus näher spezifiziert werden können oder wie die Erfassung nicht-intendierter Effekte in den Evaluationen bedacht werden kann. Um diese Punkte fundiert und ausführlich diskutieren zu können, bedarf es allerdings noch einer näheren, strukturierten Betrachtung der aktuellen Praktiken, da sie in diesem Bericht nicht im Fokus standen.

Insgesamt nimmt die *Impact Unit* aus der Analyse neue Inspirationen mit für weiterführende Forschungsfragen, Inhalte für Weiterbildung und Hilfsmittel sowie Gesprächsbedarfe zwischen Forschung, Praxis und Förderinstitutionen. Die Ergebnisse zeichnen das Bild einer noch jungen Evaluationspraxis, die zwar an vielen Stellen hinsichtlich der wissenschaftlichen Aussagekraft ihrer Daten Nachholbedarf hat, allerdings das Interesse an wirkungsorientierten Fragestellungen und Evaluationsvorhaben mitbringt, was hoffentlich eine gute Ausgangsposition für die gemeinsame Adressierung der ausgemachten Problemfelder darstellt.

Anhang

A. Literaturverzeichnis

BMBF (2019). Grundsatzpapier des Bundesministeriums für Bildung und Forschung zur Wissenschaftskommunikation.

https://www.bmbf.de/upload_filestore/pub/Grundsatzpapier_zur_Wissenschaftskommunikation.pdf

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>

Chelimsky, E. (1997). Thoughts for a New Evaluation Society. *Evaluation*, 3 (1), S. 97-109.
<https://doi.org/10.1177/135638909700300107>

Deutscher Bundestag (2020). *Wissenschaftskommunikation im Fokus. Bildung, Forschung und Technikfolgenabschätzung/Anhörung - 27.05.2020 (hib 547/2020)*. <https://www.bundestag.de/presse/hib/698158-698158>

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

Früh, W. (2017). *Inhaltsanalyse* (9., überarbeitete Auflage). Konstanz und München: UVK.

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>

Grand, A. & Sardo, A. M. (2017). What Works in the Field? Evaluating Informal Science Events. *Frontiers in Communication*, 2 (22). <https://doi.org/10.3389/fcomm.2017.00022>

Joubert, M. (2007). *Evaluating science communication projects*.
<https://www.scidev.net/global/communication/practical-guide/evaluating-science-communication-projects-1.html>

Mayring, P. (2015). *Qualitative Inhaltsanalyse. Grundlagen und Techniken* (12., überarbeitete Auflage). Weinheim/Basel: Beltz Verlag.

Phillips, T., Porticella, N., Constas, 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>

Phineo (2018). *Kursbuch Wirkung: Das Praxishandbuch für alle, die Gutes noch besser tun wollen*.
<https://www.phineo.org/kursbuch-wirkung>

Sanchez-Mora, C. (2014). Evolution of evaluation in science museums and centres. *Journal of Science Communication*, 13 (1). <https://doi.org/10.22323/2.13010302>

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/>

Wissenschaft im Dialog (2020). *Evaluationen in der Wissenschaftskommunikation: Ein Zwischenbericht zu theoretischen und konzeptionellen Grundlagen*. https://www.wissenschaft-im-dialog.de/fileadmin/user_upload/Projekte/Impact_Unit/Dokumente/200930_Zwischenbericht_Toolkit.pdf

B. Stichprobe

Im Folgenden werden jene Quellen aus der Stichwortsuche aufgelistet, die online zugänglich waren.

acatech; Universität Stuttgart (2011). „*Forscherinnen-Camps*“. Eine Woche Ingenieurin sein. [https://www.schulewirtschaft.de/www/schulewirtschaft.nsf/res/91FC32C6BE8B2F37C12579B100484938/\\$file/forscherinnenCamp.pdf](https://www.schulewirtschaft.de/www/schulewirtschaft.nsf/res/91FC32C6BE8B2F37C12579B100484938/$file/forscherinnenCamp.pdf)

Baden-Württemberg Stiftung (2015). *Nachhaltigkeit Lernen - Kinder gestalten Zukunft. Ergebnisse der Evaluation des Programms*. <https://www.bwstiftung.de/nachhaltigkeitlernen/>

Beck, C. (2012). „*Vom Wetterfrosch zum Klimafuchs*“ Konzeption und Umsetzung einer interaktiven Erlebnisausstellung für Kinder. <https://www.dbu.de/OPAC/ab/DBU-Abschlussbericht-AZ-29053.pdf>

Bittner, L. (2018). *Guerilla-Wissenschaftskommunikation in der Kneipe: Untersuchung am Projekt “Plötzlich Wissen!”*. https://wmk.itz.kit.edu/downloads/SIP_Arbeitsberichte_4_WEB.pdf

Budke, M. (2019). *Entwicklung und Evaluation des Projektes GreenLab_OS - Empirische Studie zu Effekten und stationären und mobilen Schülerlaborangeboten*. <https://repositorium.ub.unisnabrueck.de/handle/urn:nbn:de:gbv:700-201905101527>

com.X; inscico (2015). *Bericht Evaluation ZukunftsForen 2015 des BMBF*. <https://docplayer.org/150830453-Bericht-evaluation-zukunftsforen-2015-des-bmbf.html>

com.X; inscico (2017). *Wissenschaftsjahr 2015 - Zukunftsstadt. Ergebnisse der Begleitforschung*. [Broschüre](#) und [Langfassung](#).

Damerau, K. (2012). *Molekulare und Zell-Biologie im Schülerlabor. Fachliche Optimierung und Evaluation der Wirksamkeit im BeLLBio (Bergisches Lehr-Lern-Labor Biologie)*. <http://elpub.bib.uni-wuppertal.de/servlets/DerivateServlet/Derivate-3530/dc1231.pdf>

Decristoforo, B.; Hopmann, S.; Katschnig, T.; Seebauer, L.; Swertz, C. (Hg.). *Hands-On im Technischen Museum Wien. Konzeption und Evaluation der Mitmachausstellung In Bewegung*. Wien: New Academic Press. <https://services.phaidra.univie.ac.at/api/object/o:725874/diss/Content/get>

Felt, U.; Müller, A; Schober, S. (2001). *Evaluierung der ScienceWeek @ Austria 2001. Analyse eines Experiments der Wissenschaftskommunikation im österreichischen Kontext*. <https://repository.fteval.at/290/>

Flämig-Wolak, A.; Sträter, C.; Galateanu, R. (2018). *TEDxDresden Feedback Report 2018*.
<https://tedxdresden.com/wp-content/uploads/2019/01/2018-Feedback-Report.pdf>

Garner, N., Siol, A., Eilks, I. (2014). *Nachhaltigkeit und Chemie im Schülerlabor - Erwartungen und Erfahrungen*. Conference Paper.
https://www.researchgate.net/publication/266374182_Nachhaltigkeit_und_Chemie_im_Schulerlabor_-_Erwartungen_und_Erfahrungen

Geyer, C.; Lewalter, D. (2012). *Abschlussbericht der Besucherbefragung in der Ausstellung zur Nano- und Biotechnologie im Juni/Juli 2012*. https://www.deutsches-museum.de/fileadmin/Content/010_DM/050_Forschung/030_Forschung_Projekte/Besucherstruktur_Nanoausstellung_Sept2012_2.pdf

Geyer, C.; Noschka-Roos, A.; Lewalter, D. (2005). *Abschlussbericht der summativen Evaluation der Internetseiten des ZNT des Deutschen Museums München*. https://www.deutsches-museum.de/fileadmin/Content/010_DM/050_Forschung/030_Forschung_Projekte/zntweb11.pdf

Gruber, S.; Thenius, C. (2014). *Draw a Scientist - Wer macht Wissenschaft? Ein Science-Awareness-Projekt für Kinder. Endbericht*. https://www.rat-fte.at/files/rat-fte-pdf/publikationen/2014/1407_Draw%20a%20scientist-Wer%20macht%20Wissenschaft.pdf

Goldschmidt, R.; Scheel, O.; Renn, O. (2012). *Zur Wirkung und Effektivität von Dialog- und Beteiligungsformaten*. <https://elib.uni-stuttgart.de/handle/11682/5569>

Heinzerling, P. (2016). *Das EAM Schülerlabor Chemie. Ein Konzept zur Heranführung von Gymnasiasten an Inhalte und Studiengänge eines Exzellenzclusters*.
https://www.researchgate.net/publication/331430459_Das_EAM_Schulerlabor_Chemie_-Ein_Konzept_zur_Heranfuehrung_von_Gymnasiasten_an_Inhalte_und_Studiengange_eines_Exzellenzclusters

Landesverband Hamburg e.V. (2017). *Projekt Wald verstehen*.
https://www.bfn.de/fileadmin/BfN/gesellschaft/Dokumente/Endbericht-Wald-verstehen_bf.pdf

Langner, M. (2018). *Story Maps in der Wissenschaftskommunikation: Ergebnisse der Sturzflutforschung öffentlichkeitswirksam präsentieren*.
https://www.bgu.tum.de/fileadmin/w00bjl/hydrologie/Christiane1/Lehre/Studentische_arbeiten/fertige_Arbeiten/B81_Kaiser_Langer_Story_Map.pdf

Mathematikum Gießen e.V. (2017). *Abschlussbericht im Rahmen der finanziellen Unterstützung der Konzeption und Umsetzung einer Wanderausstellung zum Thema Virtuelles Wasser*.
<https://www.dbu.de/OPAC/ab/DBU-Abschlussbericht-AZ-31850.pdf>

Pfuhl, N.; Lewalter, D. (2008). *Abschlussbericht: Studie zum Ausstellungsbereich Gläsernes Forscherlabor*.
<https://www.deutsches-museum.de/fileadmin/Bericht-Gl%C3%A4sernesForscherlabor.pdf>

Riemenschneider, K. (2012). *Das Maritime Schülerlabor Ostsee (MariSchool) am Leibniz-Institut für Ostseeforschung Warnemünde. Konzeption und erste Evaluationsergebnisse*. http://rosdok.unirostock.de/file/rosdok_disshab_0000000948/rosdok_derivate_0000005023/Dissertation_Riemenschneider_2013.pdf

Pawek, C. (2009). *Schülerlabore als interessefördernde außerschulische Lernumgebungen für Schülerinnen und Schüler aus der Mittel- und Oberstufe.*

https://www.dlr.de/schoollab/en/Portaldata/24/Resources/dokumente/Diss_Pawek.pdf

Röllke, K. (2019). *Was kann ein Schülerlabor leisten? Konzeptionierung des Schülerlabors teutolab-biotechnologie als Lehr-Lern-Labor mit Angeboten zur Breiten- und zur Begabtenförderung von Schülerinnen und Schülern.* <https://pub.uni-bielefeld.de/record/2935577>

Schelhowe, H.; Kümmel, A. (2007). *Ergebnisse der wissenschaftlichen Begleitforschung des Projekts "Rober-ta goes EU". Abschlussbericht.* http://dimeb.informatik.uni-bremen.de/documents/Abschlussbericht_Evaluation_2007.pdf

Schmitt-Sody, B. (2013). *NESSI-FÖSL. Konzeption und Evaluation eines Schülerlabors für Förderschüler aus chemiedidaktischer Perspektive.* <https://d-nb.info/1075478588/34>

Science Center Netzwerk (2009). *Begleitforschung zur Ausstellung "Erlebnis Netz(werk)e" am Standort Wien. Endbericht.* <https://www.yumpu.com/de/document/view/52245237/begleitforschung-zur-ausstellung-erlebnis-netzwerke-am-standort-wien/39>

Science Center Netzwerk (2010). *Forschend lernen - Partnerschaften zwischen Volksschulen und Science Center Einrichtungen.* https://www.science-center-net.at/wp-content/uploads/2018/12/Forschend-Lernen_EB_4-Anhang.pdf

Science Center Netzwerk (2014). *Bedarfserhebung und wissenschaftliche Begleitung von Qualifizierungsmaßnahmen für interaktive Wissenschaftskommunikation. Endbericht.* <https://www.science-center-net.at/wp-content/uploads/2018/06/Bedarf-und-wissenschaftliche-Begleitung-von-Qualifizierungsma%C3%9Fnahmen-f%C3%BCr-interaktive-Wissenschaftskommunikation.pdf>

Schneider, K.; Makles, A.; Diepers, B.; Frank, C. (2013). *Evaluation der Junior Uni. Zentrale Befunde aus vier Jahren Begleitforschung.* https://www.wib.uni-wuppertal.de/fileadmin/wib/documents/publications/Bericht_Junior_Uni_Evaluation.pdf

Schulz, M.; Hiller, S.; Keierleber, V.; Renn, O.; Pfenning, U. (2011). *Stuttgarter Projektergebnisse zum Thema technisch-naturwissenschaftliche Wissensvermittlung an Kinder und Jugendliche.* <https://elib.uni-stuttgart.de/handle/11682/5570>

Specht, I.; Lewalter, D. (2011). *Abschlussbericht der Besucherbefragung in der Ausstellung zur Nano- und Biotechnologie im Deutschen Museum München.* https://www.deutsches-museum.de/fileadmin/Content/010_DM/050_Forschung/030_Forschung_Projekte/Bericht_ZNT_Besucherstruktur_Februar_2011.pdf

Spielvogel, N. (2014). *Begreifen - Erfassen - Durchsteigen. Über die Produktion von Körperwissen im Kindermuseum.* https://uol.de/f/3/inst/materiellekultur/Studien_Mat_Kult_Band_01-30/Band11_Spielvogel_Koerperwissen_2014_01.pdf

Tajmel, T.; Salzmann, I. (2016). *From Stage to Classroom - the Transfer of Knowledge though the Festival "Science on Stage".* https://www.science-on-stage.eu/images/download/Boston_Tajmel_Presentation_MRS_2016.pdf

Thänert, S. (2000). *Die „Lange Nacht der Museen“ als Form von Eventmarketing für Museen?* https://archiv.ub.uni-heidelberg.de/artdok/1038/1/Thaenert_Die_Lange_Nacht_der_Museen_2000.pdf

Gerstenberg, F.; Wedekind, H.; Weßels, H. (2014). *Wind bringt's / Einblicke in die wissenschaftliche Evaluation der Arbeit des Kinderforschungszentrums.* https://www.helleum-berlin.de/fileadmin/user_upload/Dokumente/Alice_28_Web_NawiLT.pdf

van Heukelum, M. (2015). *Erstellung und Evaluation eines Schülerlabors zum Magnetismus für Grundschüler.* <http://www.thomas-wilhelm.net/arbeiten/SLmagnetismus.pdf>

Wissenschaft im Dialog (2013). *ZukunftsWerkStadt. Evaluation der Bürgerbeteiligung.* <https://www.wissenschaft-im-dialog.de/medien/publikationen-fotos-videos/zukunftswerkstadt/>

Wissenschaft im Dialog (2016). *Junior Science Café Evaluation.* https://juniorsciencecafe.de/media/filer_public/bb/fa/bbfa849d-936e-49e8-a48c-704c534572a5/evaluation_juniorsciencecafe.pdf

Wissenschaftsladen Bonn e.V.; Technische Universität Dresden; the Good Evil GmbH (2017). *Schlussbericht „Serena – Serious Game Erneuerbare Energien zu technischen Ausbildungsberufen für Mädchen“.* <https://www.tib.eu/de/suchen/id/TIBKAT:1028477589/>

Wissenschaftsladen Bonn e.V.; Technische Universität Dresden; Technische Universität Berlin; the Good Evil GmbH (2019). *Schlussbericht „MitEffekt - Mittelfristige Effekte eines Serious Games in Abhängigkeit der didaktischen Einbettung und des konkreten Spielverhaltens“.* <https://www.tib.eu/de/suchen/id/TIBKAT:1687213909/MitEffekt-Mittelfristige-Effekte-eines-Serious?cHash=2eeb4673381690b7cb6f488b16906dca>

Wuppertaler Institut für bildungökonomische Forschung (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

Zehren, W. (2009). *Forschendes Experimentieren im Schülerlabor.* <https://publikationen.sulb.uni-saarland.de/handle/20.500.11880/22651>

Zürcher Hochschule für Angewandte Wissenschaften (2013). *Schlussbericht Evaluation der Doppelseite Wissen in 20 Minuten. Wahrnehmung von Wissenschaftskommunikation bei Jugendlichen.* <https://digitalcollection.zhaw.ch/handle/11475/7846>

C. Liste der durchsuchten Stichworte

Stichwort Kombination	Treffer bei Google	Relevante Treffer*
Wissenschaftskommunikation Evaluation	51100	11
Wissenschaftskommunikation Begleitforschung	4110	19
Wissenschaftskommunikation Fragebogen	15500	3
Wissenschaftskommunikation Interview	88200	0

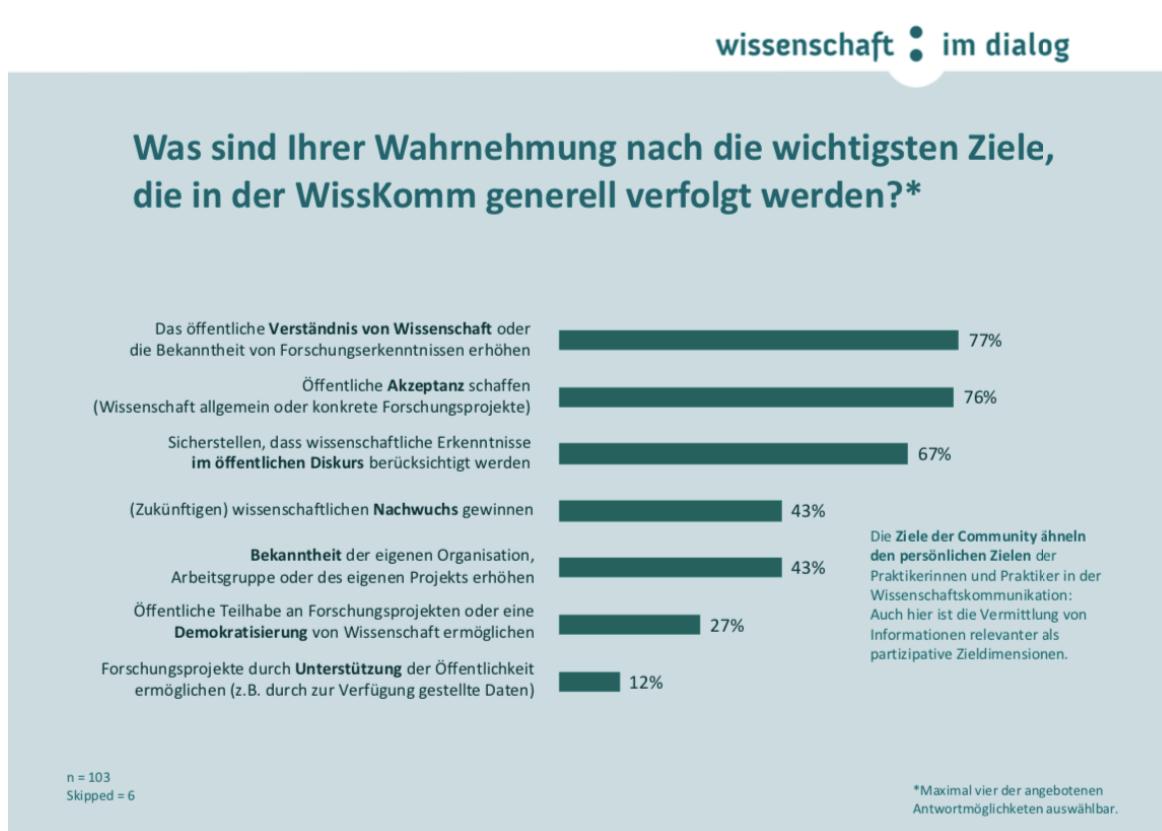
Wissenschaftskommunikation Feedback	36400	2
Wissenschaft vermitteln Evaluation	2320000	4
Wissenschaft erklären Evaluation	1400000	7
Wissenschaft Beteiligung Evaluation	1930000	3
Wissenschaft diskutieren Evaluation	3710000	0
Wissenschaft vermitteln Begleitforschung	55200	1
Wissenschaft erklären Begleitforschung	49100	0
Wissenschaft Beteiligung Begleitforschung	92100	0
Wissenschaft diskutieren Begleitforschung	71900	0
Wissenschaft vermitteln Feedback	3110000	1
Wissenschaft erklären Feedback	2950000	0
Wissenschaft Beteiligung Feedback	2820000	0
Wissenschaft diskutieren Feedback	7470000	1
Forschung vermitteln Evaluation	5130000	0
Forschung erklären Evaluation	8000000	0
Forschung Beteiligung Evaluation	4980000	0
Forschung diskutieren Evaluation	8480000	0
Forschung vermitteln Begleitforschung	76700	1
Forschung erklären Begleitforschung	49900	0
Forschung Beteiligung Begleitforschung	99600	0
Forschung diskutieren Begleitforschung	82300	0
Forschung vermitteln Feedback	2390000	1
Forschung erklären Feedback	2590000	0
Forschung Beteiligung Feedback	760000	0
Forschung diskutieren Feedback	2700000	0
Barcamp Evaluation	187000	1
Citizen Science Evaluation	138000000	2
Bürgerwissenschaft Evaluation	4210	0
Fishbowl Evaluation	455000	0
Forschungsbörse Evaluation	2050	0

Hackday Evaluation	177000	0
Haus der Wissenschaft Evaluation	5020000	0
Kindermuseum Evaluation	32000	2
Kinderuni Evaluation	635000	5
Lange Nacht Evaluation	1330000	9
Meet the Scientist Evaluation	56200000	0
Postersession Evaluation	3930000	1
Posterslam Evaluation	5810	0
Pub-Science Evaluation	338000000	0
Pubquiz Evaluation	822000	0
Schülerlabor Evaluation	48200	15
Schülerparlament Evaluation	13200	0
Schülerwettbewerb Evaluation	21100	4
Science Café Evaluation	135000000	1
Science Fair Evaluation	301000000	0
Science Festival Evaluation	49500000	1
Science Slam Evaluation	3670000	0
Science Rallye Evaluation	9630000	1
Science Show Evaluation	950000000	0
Science Speeddating Evaluation	1520000	0
Science Tweetup Evaluation	144000	0
Scrollytelling Evaluation	8970	1
Soapbox Evaluation	2850000	0
Taschenlampenführung Evaluation	367	0
TEDx Event Evaluation	469000	2
Unterhausdebatte Evaluation	161	0
Wanderausstellung Evaluation	51500	7
Wissensbuffet Evaluation	303	0
Wissenschaftscomedy Evaluation	145000	0
Wissenschaftsfilmfest Evaluation	122	0

Wissenschaftsladen Evaluation	6070	2
Wissenschaftspuppentheater Evaluation	110	0
Wissenschaftssong Evaluation	8	0
Wissenschaftstheater Evaluation	563	0

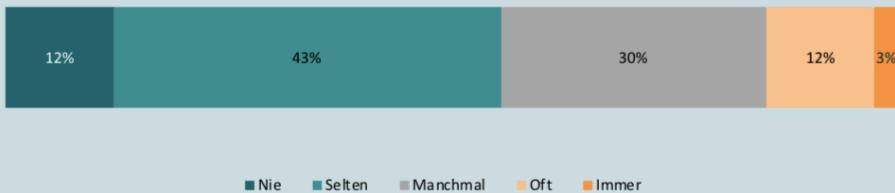
*Diese Dokumente wurden begutachtet, allerdings aufgrund fehlender notwendiger Kriterien möglicherweise aus der Analyse ausgeschlossen.

D. Ergebnisse der Community-Befragung der Impact Unit (2019)



Wie häufig werden bei Projekten, an denen Sie beteiligt sind/waren, mehrere Methoden oder Instrumente zur Evaluation eines Formats angewendet?

Passend zu den Ergebnissen hinsichtlich der Methodenwahl, die für eine möglichst pragmatische, wenig aufwendige Evaluationspraxis sprechen, sind Mehr-Methoden-Designs eher unüblich.

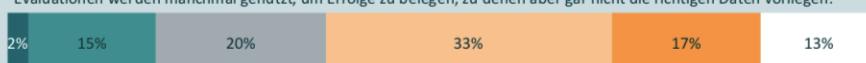


n = 74
Skipped = 35

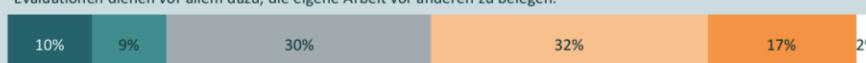
Inwiefern stimmen Sie vor dem Hintergrund Ihrer Erfahrungen den folgenden Aussagen zu?

Kritische Einschätzungen werden deutlich, wenn es um die Durchführung, Motivation und sinnvolle Verwertung von Evaluationen geht. 16 % zweifeln daran, ob sich aufwendige Evaluationen lohnen.

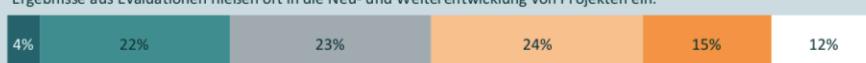
Evaluationen werden manchmal genutzt, um Erfolge zu belegen, zu denen aber gar nicht die richtigen Daten vorliegen.



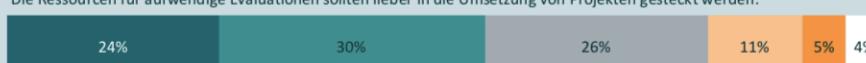
Evaluationen dienen vor allem dazu, die eigene Arbeit vor anderen zu belegen.



Ergebnisse aus Evaluationen fließen oft in die Neu- und Weiterentwicklung von Projekten ein.



Die Ressourcen für aufwendige Evaluationen sollten lieber in die Umsetzung von Projekten gesteckt werden.



n = 82
Skipped = 27

Welche Art von Unterstützung könnte die Aussagekraft von Evaluationen verbessern?

Für die Zukunft wünschen sich die Befragten einen besseren Zugang zu Expertenwissen, Hilfsmittel für komfortable Methodendurchführung und zur Interpretation (und Vergleichbarkeit) von Daten sowie einen kulturellen Wandel bezüglich des Stellenwerts von Evaluationen für die Praxis der Wissenschaftskommunikation.

