



Experimentelle Herangehensweisen in der Evaluation von Maßnahmen der Wissenschaftskommunikation

Marc Stadtler und Corinna Schuster

Zusammenfassung

In unserem Beitrag zeigen wir Potenziale und Limitationen von Experimenten in der Evaluation von Maßnahmen der Wissenschaftskommunikation auf. Dabei arbeiten wir zunächst den Unterschied zwischen experimentellen und nicht-experimentellen Studien heraus und zeigen deren spezifische Erkenntnismöglichkeiten auf. Die Logik von Experimenten wird anhand zahlreicher Beispiele aus der Literatur verdeutlicht. Das Kapitel schließt mit einem Blick auf methodische Herausforderungen und deren Bewältigung, die beim Experimentieren in der angewandten Forschung der Wissenschaftskommunikation besonders verbreitet sind.

1 Einführung

Lässt sich mit authentischen Ausstellungsstücken die Aufmerksamkeit von Museumsbesucher:innen besonders gut gewinnen (vgl. Schwan et al. 2016)? Führen Citizen-Science-Projekte zu einem verbesserten Wissenschaftsverständnis bei den teilnehmenden Bürger:innen (vgl. Bonney et al. 2015)? Fördert die Dar-

M. Stadtler (✉) · C. Schuster
Institut für Erziehungswissenschaft, Ruhr-Universität Bochum, Bochum, Deutschland
E-Mail: marc.stadtler@rub.de

C. Schuster
E-Mail: corinna.schuster@rub.de

stellung wissenschaftlicher Sachverhalte in Comics das Themenverständnis bei jungen Menschen (vgl. Lin et al. 2015)?

Studien, in denen Maßnahmen der Wissenschaftskommunikation evaluiert werden, rücken oft die Frage nach der Wirkung spezifischer kommunikativer Maßnahmen in den Mittelpunkt. Wünschenswerte Effekte in Bereichen wie Interessensförderung, Vertrauensbildung oder Wissenschaftsverständnis sollen in einem kausalen Sinne auf die zu evaluierende Maßnahme zurückgeführt werden. Mit diesem Erkenntnisinteresse kommen Studiendesigns, die Reaktionen von Bürger:innen, Museumbesucher:innen oder Podcasthörer:innen ausschließlich beobachten, ohne in den Rezeptionsprozess einzugreifen, an ihre Grenzen. Stattdessen sind Experimente die Methode der Wahl, die ihren Ursprung in den Naturwissenschaften haben und im Fokus dieses Beitrags stehen. Der Beitrag gibt eine praxisnahe Einführung in die Logik des Experiments, skizziert seine verschiedenen Erscheinungsformen und zeigt Potenziale und Limitationen des Experimentierens im Kontext der Wissenschaftskommunikation auf.

2 Grundbegriffe des Experimentierens

Stellen Sie sich vor, ein Team von Forscher:innen möchte herausfinden, ob der Besuch von Science Slams zu einer positiven Einstellung gegenüber Wissenschaft führt. Die Forschenden besuchen eine ganze Reihe solcher Veranstaltungen, erheben von den anwesenden Besucher:innen Daten zur Häufigkeit des Besuchs von Science Slams und messen ihre Einstellung zur Wissenschaft. Zur Freude des Forscher:innenteams zeigen die statistischen Analysen einen positiven Zusammenhang der Variablen: Je öfter Personen Science Slams besuchen, desto positiver ihre Einstellung gegenüber der Wissenschaft. Lässt sich hieraus ableiten, dass Unterschiede in den Einstellungen kausal auf die Häufigkeit des Besuchs von Science Slams zurückzuführen sind? Tatsächlich ist die im fiktiven Beispiel beobachtete Kontiguität zwischen Ursache und Wirkung, also ihr gemeinsames Auftreten, eines von mehreren Kriterien, die erfüllt sein müssen, um auf einen kausalen Zusammenhang zwischen zwei Variablen schließen zu können (für einen Überblick der erkenntnistheoretischen Grundlagen von Kausalschlüssen, siehe Field und Hole 2003). Allerdings ist die Kontiguität nur ein notwendiges, kein hinreichendes Kriterium. Um auf eine kausale Beziehung zwischen zwei Variablen zu schließen, muss die Ursache zudem zeitlich *vor* der Wirkung liegen. Dieses Kriterium zeigt ein erstes Problem der oben skizzierten querschnittlichen Studienanlage auf, denn es wäre auch denkbar, dass die positiven Wissenschaftseinstellungen der Science-Slam-Enthusiast:innen bereits vor dem Besuch der

Science Slams vorlagen, mithin nicht als Wirkung infrage kommen. Schließlich könnte es auch eine dritte, von Bauer und Kolleg:innen (2014) als „vergessene Variable“ bezeichnete Größe geben, die *sowohl* Wissenschaftseinstellung *als auch* Häufigkeit der Besuche von Science Slams beeinflusst. So könnte die Nähe des von den Versuchspersonen ausgeübten Berufs zum Wissenschaftsbetrieb sowohl einen Einfluss darauf ausüben, wie häufig man Science Slams besucht, als auch darauf, wie positiv man der Wissenschaft gegenübersteht. Eine solche Konfundierung kann nur ausgeschlossen werden, wenn ein Studiendesign gewählt wird, in dem die Forschenden die mutmaßlich verursachende Variable isolieren und sie in einer Untersuchungsbedingung präsentieren, während diese in einer anderen Bedingung, die ansonsten vergleichbar ist, nicht dargeboten wird. Das Ausmaß der Differenz zwischen den Bedingungen, in denen die Ursache präsent beziehungsweise nicht präsent ist, nährt laut Field und Hole das Vertrauen darin, dass es sich um einen kausalen Zusammenhang zweier Variablen handelt.

Genau das ist die Grundidee des wissenschaftlichen Experimentierens: Veränderungen in der Ausprägung einer *abhängigen Variable* werden durch gezielte Manipulationen einer anderen Variable, die als *unabhängige Variable* bezeichnet wird, verursacht. Anders als im weiter oben dargestellten korrelativen Design, greift der:die Forscher:in aktiv in den Rezeptionsprozess ein und manipuliert eine (oder mehrere) Variable(n), während er:sie andere kontrolliert. Durch das hohe Maß an Kontrolle steigt die eindeutige Interpretierbarkeit der Ergebnisse (man spricht von einer hohen ‚internen Validität‘), weil Veränderungen der abhängigen Variable mit nur einer Ursache plausibel erklärt werden können. Das Gegenstück zur internen Validität ist die externe Validität, also die Verallgemeinerbarkeit der Ergebnisse eines Experiments über die Untersuchungssituation hinaus. Hier zeigen Experimente mitunter Defizite, weil ein hohes Maß an experimenteller Kontrolle zu artifiziellen Untersuchungssituationen führen kann, die sich von der natürlichen Auseinandersetzung der Bürger:innen mit Wissenschaftsinhalten deutlich unterscheiden. Ein konstruktiver Umgang mit diesem Problem könnte darin bestehen, zunächst sorgfältig im Labor getestete Wirkhypothesen zunehmend auch in authentischeren Settings zu überprüfen, um das Beste aus beiden Welten zu vereinen.

3 Experimentaldesigns für die Evaluation von Maßnahmen der Wissenschaftskommunikation

Wer eine Maßnahme der Wissenschaftskommunikation mithilfe eines Experiments evaluieren will, muss im Rahmen der Studienplanung Entscheidungen über das Experimentaldesign treffen. Hiermit bezeichnet man die Auswahl von

abhängigen und unabhängigen Variablen sowie die Anordnung der Messzeitpunkte innerhalb einer Studie. Die Wahl eines Designs will wohlüberlegt sein, hat sie doch gewichtige Konsequenzen für die mit dem Experiment verbundenen Erkenntnismöglichkeiten, den benötigten Stichprobenumfang und – hierüber vermittelt – den mit der Studiendurchführung verbundenen Aufwand.

Kennzeichnend für sogenannte Messwiederholungsdesigns (engl.: repeated measures design) ist, dass jede:r Untersuchungsteilnehmende alle Experimentalbedingungen durchläuft. Ein solches Design wurde zum Beispiel von Scharrer und Kolleg:innen (2017) in ihrer Laborstudie zum Einfachheitseffekt der Wissenschaftspopularisierung (Scharrer et al. 2017) verwendet. In dem Experiment lasen medizinische Lai:innen authentische Wissenschaftsartikel aus dem Internet, in denen medizinische Themen erläutert wurden. Die Artikel richteten sich entweder an ein Expert:innenpublikum (z. B. auf www.aerzteblatt.de) und waren sprachlich komplex oder an ein Lai:innenpublikum (z. B. auf www.manshealth.com) und verwendeten eine stark vereinfachte Sprache. Weil die Proband:innen nacheinander Artikel (zu verschiedenen Themen) aus beiden Kategorien lasen, handelt es sich um ein Messwiederholungsdesign (hier mit dem Messwiederholungsfaktor „Sprachliche Einfachheit“). Scharrer und Kolleg:innen (2017) vermuteten unter anderem, dass die Untersuchungsteilnehmer:innen Behauptungen aus den sprachlich einfacheren Lai:innentexten mehr Glauben schenken als den Behauptungen aus den sprachlich komplexeren Artikeln, die sich an ein Expert:innenpublikum richteten. Die Ergebnisse bestätigten die Hypothese der Studienautor:innen.

Weil die mit statistischen Verfahren zu vergleichenden Mittelwerte aus den Angaben derselben Versuchspersonen gebildet werden, benötigen Messwiederholungsdesigns vergleichsweise wenige Proband:innen und sind damit ökonomischer als Zwischensubjektdesigns, in denen die Mittelwerte basierend auf den Angaben *verschiedener* Versuchspersonen gebildet werden. Zu den Nachteilen von Messwiederholungsdesigns zählt jedoch, dass Proband:innen die mit der experimentellen Manipulation verbundene Intention erkennen und sich zum Beispiel möglichst kooperativ im Sinne der vermuteten Hypothese verhalten könnten (auch das Gegenteil ist denkbar), was einem verzerrten Antwortverhalten gleichkäme. Zudem besteht die Gefahr, dass Unterschiede zwischen den Experimentalbedingungen im Sinne von Kontrasteffekten überzeichnet und Effekte der experimentellen Manipulation überschätzt werden. So ist es in der Studie von Scharrer und Kolleg:innen denkbar, dass die einfachen Themendarstellungen besonders einfach wirken, wenn sie direkt nach einer komplexen Darstellung rezipiert werden und umgekehrt. Anders ausgedrückt: Die Messungen können potenziell durch Erfahrungen beeinflusst werden, die Proband:innen in

anderen Versuchsbedingungen gemacht haben. Solche Phänomene, die man auch als Spill-Over-Effekte bezeichnet, gilt es zu vermeiden, weil sie die Interpretierbarkeit der Studienergebnisse einschränken.

Zwischensubjekt designs begegnen diesem Problem, indem den Stufen einer unabhängigen Variable distinkte Subgruppen der Gesamtstichprobe zugeordnet werden. So wurden zum Beispiel in einer Studie von Lin und Kolleg:innen (2015) erwachsenen Proband:innen Informationen zum Thema Nanotechnologie dargeboten. Die Hälfte der Proband:innen erhielt die Information in Form eines Sachtexts, während die andere Hälfte der Proband:innen dieselben Informationen als Comic aufbereitet rezipierten. Es zeigte sich, dass die Teilnehmer:innen der Comic-Bedingung mehr Freude beim Lernen erlebten und ein größeres Interesse an Wissenschaft bekundeten, als die Proband:innen der Sachtextbedingung, wohingegen keine Unterschiede in den Bereichen Wissenserwerb und Einstellungen zur Wissenschaft gefunden wurden.

Der Vorteil von Zwischensubjekt designs besteht darin, dass die im Zusammenhang mit Innersubjekt designs diskutierten Nachteile (Hypothesenraten und Kontrasteffekte) keine bzw. eine untergeordnete Rolle spielen. Sie sind durch die größeren Stichprobenumfänge allerdings aufwendiger in der Durchführung. Zudem gilt, dass selbst bei randomisierter Zuweisung der Versuchspersonen zu Experimentalbedingungen das oben bereits diskutierte Drittvariablenproblem nicht vollständig ausgeschlossen werden kann. Hier gilt es bereits im Vorfeld, potenziell konfundierende Variablen wie Wissenschaftsinteresse oder thematisches Vorwissen zu identifizieren und diese Konstrukte als Kontrollvariablen zu erfassen. Dies gibt den Forschenden zumindest die Möglichkeit, den Effekt der konfundierenden Variablen statistisch mit dem Verfahren der Kovarianzanalyse zu kontrollieren (für eine detaillierte Darstellung dieses Verfahrens, siehe Döring und Bortz 2016).

In der Evaluation von Maßnahmen der Wissenschaftskommunikation sind neben den beiden vorgestellten Grundformen vor allem gemischte Designs mit Zwischensubjekt- und Innersubjektfaktor angezeigt. Dabei werden verschiedene Varianten einer Maßnahme als Stufen eines Zwischensubjekt factors miteinander verglichen. Die interessierenden Variablen werden zu mindestens zwei Messzeitpunkten erhoben. In der bereits vorgestellten Studie von Lin und Kolleg:innen (2015) wurde das Wissen über Nanotechnologie, die Einstellungen zur Nanotechnologie und das Interesse am Thema zum Beispiel unmittelbar vor dem Bearbeiten der Untersuchungsmaterialien (Prätest) und unmittelbar danach (Posttest) erfasst. Auf diesem Wege sind die Autor:innen in der Lage, Veränderungsmessungen vorzunehmen und auszuschließen, dass die unmittelbar nach dem Lesen erfassten Unterschiede bereits vor der Rezeptionssituation bestanden.

Idealerweise würde das Design um einen Messzeitpunkt ergänzt, der zeitlich deutlich verzögert liegt und Aussagen über die zeitliche Stabilität der Befunde zulässt (Verzögerter Posttest). So zielen viele kommunikative Maßnahmen darauf ab, möglichst nachhaltige Veränderungen herbeizuführen, die sich nicht bloß auf rasch verblässende Eindrücke beschränken. Zudem kennt die Lernpsychologie auch sog. Sleeper-Effekte, bei denen die Vorteile einer Lernform erst nach einer Weile sichtbar werden (Köller 2015). Dies ist zum Beispiel der Fall, wenn nicht die Teilnahme an einer Maßnahme der Wissenschaftskommunikation selbst, sondern die in der Folge auftretende vermehrte Auseinandersetzung mit dem Thema zu einem substanziellen Lernzuwachs führt. Solche Effekte würden ohne Hinzufügen eines verzögerten Posttests übersehen und die Wirksamkeit einer Maßnahme systematisch unterschätzen.

4 Zur Wahl eines geeigneten Benchmarks in Evaluationsstudien oder „Was macht eigentlich die Kontrollgruppe?“

Soll die Wirkung einer Maßnahme der Wissenschaftskommunikation evaluiert werden, geht damit in der Regel die Entscheidung über eine geeignete Kontrollgruppe einher. Eine Taxonomie der Handlungsmöglichkeiten bietet die Klassifikation von Hager (2008), die ihren Ursprung in der Evaluation pädagogisch-psychologischer Interventionen hat und zwischen isolierter, vergleichender und kombinierter Evaluation unterscheidet.

Bei der *isolierten Evaluation* geht es darum, einen grundsätzlichen Wirknachweis einer Maßnahme zu erbringen. Die zu testende Maßnahme wird experimentell mit einer anderen Intervention von ähnlicher Komplexität und Dauer verglichen, die jedoch andere Ziele verfolgt. Beispielsweise könnten die Effekte eines Schülerlaborworkshops zu den Ursachen des Klimawandels mit einem in Dauer, Didaktik und Zielgruppe vergleichbaren Schülerlaborworkshop zu einem anderen Thema verglichen werden. Es würde erwartet, dass bei den Proband:innen der Interventionsgruppe, nicht jedoch bei denen der Kontrollgruppe, Veränderungen in der abhängigen Variable (z. B. Wissen über Ursachen des Klimawandels) auftreten. Hager (2008) spricht nur dann von einer „echten“ Kontrollgruppe, wenn diese ebenfalls eine Maßnahme durchläuft. Ist dies nicht der Fall, kann nicht ausgeschlossen werden, dass Veränderungen in der abhängigen Variable ein methodisches Artefakt sind, das auf der erfahrenen pädagogischen Zuwendung und nicht auf den spezifischen Lernhandlungen beruht.

Bei der *vergleichenden Evaluation* werden zwei oder mehr konkurrierende Maßnahmen, die mit unterschiedlichen Mitteln dieselben Ziele erreichen wollen, hinsichtlich ihrer Wirksamkeit verglichen. Diese Variante testet folglich, ob eine neue Maßnahme ähnlich gute oder gar bessere Wirkungen erzielt als eine bereits bekannte, etablierte Maßnahme. Dieses Vorgehen bietet sich insbesondere dort an, wo mit einer isolierten Evaluation bereits ein Wirknachweis einer Maßnahme erbracht worden ist. So könnte zum Beispiel getestet werden, ob ein durch die Nutzung digitaler Apps unterstützter Museumsbesuch eine größere Interessensförderung bei jungen Museumsbesucher:innen hervorruft als der alleinige Museumsbesuch. Vergleichenden Evaluationen können grundsätzlich drei Hypothesen zugrunde liegen: Bei der Äquivalenzhypothese erwarten die Forschenden, dass die zu testende Maßnahme genauso gute Ergebnisse erzielt, wie eine bereits etablierte (aber vielleicht aufwendigere) kommunikative Maßnahme. Die Überlegenheitshypothese geht hingegen davon aus, dass mit der zu testenden Maßnahme bessere Ergebnisse erzielt werden als mit der etablierten Variante, während die „Nicht-Unterlegenheitshypothese“ annimmt, dass die zu testende Maßnahme mindestens ebenbürtig ist.

Die *kombinierte Evaluation* führt schließlich beide Evaluationsformen zusammen. Sie kommt zum Beispiel zum Einsatz, wenn sich in einer vergleichenden Evaluation zunächst zwei Interventionen als gleichwertig erwiesen haben und die Wirkungsfrage nun in einer Folgestudie unter Verwendung des Designs der isolierten Evaluation beantwortet werden soll.

5 Ausgewählte methodische Herausforderungen bei der Evaluation von Maßnahmen der Wissenschaftskommunikation mit Experimenten

Eine qualitativ hochwertige Evaluation von Maßnahmen der Wissenschaftskommunikation geht mit einer Vielzahl methodischer Herausforderungen einher, die eine sorgfältige Studienplanung unausweichlich machen. Hierzu gehört neben der Wahl eines geeigneten Studiendesigns auch der Umgang mit fehlenden Daten, die aufgrund des Drop-Outs von Versuchspersonen in Designs mit mehreren Messzeitpunkten ein besonderes Augenmerk verlangen (Köller 2015). Eine weitere Herausforderung ist die geschachtelte Datenstruktur in Studien, an denen Versuchspersonen in intakten Gruppen (z. B. Schulklassen in Schülerlaborworkshops) teilnehmen. Die Datenstruktur gilt es bei der Auswahl geeigneter statistischer Verfahren wie der Berechnung von Intraklassenkoeffizienten oder Mehrebenenanalysen zu berücksichtigen (Lüdtke und Köller 2010). Weil in

diesem Beitrag die experimentellen Methoden der Evaluation im Vordergrund stehen, liegt der Fokus der methodischen Überlegungen im Folgenden auf der Unterscheidung von experimentellen und quasiexperimentellen Ansätzen und dem mit quasiexperimentellen Ansätzen verbundenen Problem der Sicherung der internen Validität.

5.1 Was ist der Unterschied zwischen Experimenten und Quasiexperimenten?

Das Kernmerkmal echter Experimente ist die zufällige (fachsprachlich: „randomisierte“) Zuteilung von Untersuchungsteilnehmer:innen zu Versuchsbedingungen, sodass jede Versuchsperson die gleiche Chance hat, in Experimental- oder Kontrollbedingung zu gelangen. Die Randomisierung sichert bei hinreichend großer Stichprobe, dass sich die Versuchsbedingungen nur in Bezug auf die manipulierte Variable, nicht jedoch in Bezug auf personenbezogene Drittvariablen wie Alter, Bildungshintergrund oder Themeninteresse unterscheiden.

Bei vielen Evaluationsstudien, die unter natürlichen Bedingungen (man sagt auch „im Feld“) durchgeführt werden, dürfte die Randomisierung allerdings eine große Herausforderung darstellen¹. Schließlich entscheiden Bürger:innen selbst, ob sie an einem Citizen-Science-Projekt teilnehmen, oder welchen Bürger:innendialog sie besuchen. Auch durch Ausstellungen in Science-Centern schlendern die Menschen üblicherweise in Gruppen (Lewalter und Noschka-Roos 2009) und können demzufolge nicht einfach durch die Experimentalleitung zufällig auf verschiedene Varianten der Ausstellung verteilt werden. Studien, in denen eine unabhängige Variable manipuliert wird, ohne dass eine randomisierte Zuteilung von Versuchspersonen zu Versuchsbedingungen erfolgt, werden als Quasiexperiment bezeichnet. Diese sind im Feld zwar leichter zu realisieren als reine Experimente, sie kommen aber oft mit der methodischen Einschränkung der Konfundierung von Versuchsbedingungen mit personenbezogenen Drittvariablen daher. Dies schränkt die interne Validität der Studie ein, weil Bedingungsunterschiede in den Ausprägungen der abhängigen Variable nicht mehr eindeutig auf die experimentelle Manipulation zurückgeführt werden können.

¹Es sei jedoch darauf hingewiesen, dass die Forschung zur Wissenschaftskommunikation auch zahlreiche „reine“ Experimente hervorgebracht hat, in denen Forscher:innen unter kontrollierten (Labor-)Bedingungen Gesetzmäßigkeiten nachspüren (z. B. Anderson et al. 2014; Bearth et al. 2022; Hendriks et al. 2016).

5.2 Maßnahmen zur Sicherung der internen Validität in Quasieexperimenten

In Quasieperimenten gilt es bereits vor Studiendurchführung verschiedene Maßnahmen zu ergreifen, um die interne Validität zu erhöhen (für einen umfassenden Überblick, siehe Döring und Bortz 2016):

- *Erfassung konfundierender Variablen:* Um Konfundierungen überhaupt aufzudecken, ist es unausweichlich, dass Forschende vor Studienbeginn potenziell konfundierende Variablen identifizieren und diese in ihrer Studie ebenfalls erfassen. Die Identifikation solcher Variablen erfordert sowohl theoretische Überlegungen als auch Kenntnisse der einschlägigen empirischen Befundlage. Nur wenn konfundierende Variablen bekannt sind und erfasst wurden, können diese zum Beispiel mit der Kovarianzanalyse statistisch kontrolliert werden. Hierbei wird regressionsanalytisch geschätzt, welchen Wert die Versuchspersonen auf der abhängigen Variable einnehmen, wenn alle Proband:innen denselben Wert auf der Kovariate hätten (Döring und Bortz 2016). Für die anschließenden Mittelwertvergleiche werden sodann die geschätzten, um Unterschiede auf der Kovariaten bereinigten Werte verwendet.
- *Einführung eines Prätests:* Um auszuschließen, dass etwaige Bedingungsunterschiede nicht bereits vor Beginn der Studie existierten, können Forscher:innen zusätzlich zum Posttest einen Prätest durchführen. Im Idealfall unterscheiden sich die Bedingungen im Prätest nicht, wohingegen im Posttest Bedingungsunterschiede in der erwarteten Richtung auftreten. Die Ausgangswerte können bei der Datenanalyse auf verschiedene Weise berücksichtigt werden. Fließen sie als weiterer Messzeitpunkt in eine gemischte Varianzanalyse (siehe oben) ein, würde sich der erwartete Effekt in einer Interaktion zwischen dem Bedingungsfaktor und dem Messwiederholungsfaktor äußern. Eine weitere Möglichkeit besteht darin, den Effekt des Ausgangsniveaus statistisch mit der Methode der Kovarianzanalyse zu kontrollieren. Eine vergleichende Diskussion der verschiedenen analytischen Vorgehensweisen bietet Hager (2008).
- *Parallelisieren von Untersuchungsbedingungen:* Sind potenziell konfundierende Variablen bekannt, sollten diese nach Möglichkeit über alle Untersuchungsbedingungen konstant gehalten werden oder zumindest ähnlich ausgeprägt sein. Wer zum Beispiel eine Studie zur Wirkung authentischer Ausstellungsobjekte in einem Museum macht, sollte die Datenerhebungen stets zu denselben Wochentagen durchführen, um sicherzustellen, dass Faktoren

wie Besucher:innenstruktur oder Besucher:innendichte in allen Bedingungen vergleichbar ausgeprägt sind. In ähnlicher Weise könnten Forscher:innen, die Workshops in einem Schülerlabor durchführen, vorab vorliegende Informationen über die Schulform und den Sozialindex der entsendenden Schule nutzen, um sicherzustellen, dass Variablen wie Lernvoraussetzungen und sozioökonomische Herkunft über die Versuchsbedingungen hinweg vergleichbar ausgeprägt sind.

6 Fazit

Die Evaluation von Maßnahmen der Wissenschaftskommunikation ist ein wichtiges Instrument zur Qualitätssicherung solcher Initiativen und zugleich Ausdruck der zunehmenden Professionalisierung eines ‚boomenden‘ Tätigkeitsfeldes. Sie liefert wichtige Erkenntnisse über die Randbedingungen gelingender Wissenschaftskommunikation und erfüllt damit nicht nur eine innerwissenschaftlich relevante, sondern auch eine gesellschaftlich bedeutsame Funktion.

Im vorliegenden Beitrag lag der Fokus auf experimentellen Herangehensweisen in der Evaluation von Maßnahmen der Wissenschaftskommunikation. Diese stellen einerseits hohe methodische Anforderungen an die Studienplanung und -durchführung, bieten andererseits aber das Potenzial, belastbare Aussagen über Kausalzusammenhänge zu liefern. Dies mag überall dort erstrebenswert sein, wo Forschende vom Konkreten abstrahieren und Aussagen mit einem weiteren Geltungsbereich treffen wollen. Es geht dann nicht mehr primär um die Rekonstruktion der subjektiven Eindrücke von zufällig zu einem bestimmten Zeitpunkt an einem Ort befindlichen Personen. Diese sind in der hier vorgestellten Forschungstradition vielmehr Stellvertreter:innen einer größeren Grundgesamtheit, fachsprachlich eine Stichprobe aus einer Population, aus der man beliebig viele weitere Stichproben ziehen könnte. Das Ziel ist, Gesetzmäßigkeiten zu erforschen, die in einem genaueren zu explizierenden Anwendungsbereich Gültigkeit besitzen und im Idealfall zu replizierbaren Effekten führen, was wiederum einen großen praktischen Nutzen hat.

Beim Design experimenteller Studien, die einen solchen Erkenntnisgewinn versprechen, können Evaluator:innen mittlerweile auf einen großen Fundus an Erkenntnissen und Best-Practice-Beispielen zurückgreifen, die in den letzten Jahrzehnten in benachbarten Disziplinen wie der empirischen Bildungsforschung gewonnen wurden. Hieraus sind verschiedene Lehrbücher und zahlreiche Fachartikel hervorgegangen, in denen das hier vorgestellte Kernproblem in der Regel deutlich ausführlicher behandelt wird, als es im Kontext dieses

Herausgeber:innenbands möglich und sinnvoll ist. Einige dieser Werke sind im vorliegenden Beitrag referenziert und bieten der interessierten Leser:innenschaft damit Anknüpfungspunkte, um tiefer in das Thema der Evaluation mit experimentellen Ansätzen einzusteigen.

Literatur

- Anderson AA, Brossard D, Scheufele DA, Xenos MA, Ladwig P (2014) The “nasty effect:” online incivility and risk perceptions of emerging technologies. *J Comput-Mediat Commun* 19(3):373–387. <https://doi.org/10.1111/jcc4.12009>
- Bauer TK, Gigerenzer G, Krämer W (2014) Warum dick nicht doof macht und Genmais nicht tötet. Über Risiken und Nebenwirkungen der Unstatistik. Campus, Frankfurt
- Beathar A, Kaptan G, Kessler SH (2022) Genome-edited versus genetically-modified tomatoes: an experiment on people’s perceptions and acceptance of food biotechnology in the UK and Switzerland. *Agric Hum Values* 39(3):1117–1131. <https://doi.org/10.1007/s10460-022-10311-8>
- Bonney R, Phillips TB, Ballard HL, Enck JW (2015) Can citizen science enhance public understanding of science? *Public Underst Sci* 25(1):2–16. <https://doi.org/10.1177/0963662515607406>
- Döring N, Bortz, J (2016) Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften (5. Aufl.). Springer, Berlin. <https://doi.org/10.1007/978-3-642-41089-5>
- Field A, Hole G (2003) How to design and report experiments. Sage, London. <https://doi.org/10.7748/nr.11.1.83.s13>
- Hager W (2008) Evaluation von pädagogisch-psychologischen Interventionsmaßnahmen. In: Schneider W, Hasselhorn M (Hrsg) *Handbuch der Pädagogischen Psychologie*. Hogrefe, Göttingen, S 721–732
- Hendriks F, Kienhues D, Bromme R (2016) Disclose your flaws! Admission positively affects the perceived trustworthiness of an expert science blogger. *Stud Commun Sci* 16(2):124–131. <https://doi.org/10.1016/j.scoms.2016.10.003>
- Köller O (2015) Evaluation pädagogisch-psychologischer Maßnahmen. In: Wild E, Möller J (Hrsg) *Pädagogische Psychologie*. Springer, S 329–342. https://doi.org/10.1007/978-3-642-41291-2_14
- Lewalter D, Noschka-Roos A (2009) Museum und Erwachsenenbildung. In: Tippelt R, von Hippel A (Hrsg) *Handbuch Erwachsenenbildung/Weiterbildung*. VS Verlag, Wiesbaden, S 527–541. https://doi.org/10.1007/978-3-531-91834-1_32
- Lin S-F, Lin H-S, Lee L, Yore LD (2015) Are science comics a good medium for science communication? The case for public learning of nanotechnology. *Int J Sci Educ Part B* 5(3):276–294. <https://doi.org/10.1080/21548455.2014.941040>
- Lüdtke O, Köller O (2010) Mehrebenenanalyse. In: Rost DH (Hrsg) *Handwörterbuch Pädagogische Psychologie* (4. Aufl.). Beltz/PVU, Weinheim, S 530–535
- Scharrer L, Rupieper Y, Stadler M, Bromme R (2017) When science becomes too easy: science popularization inclines laypeople to underrate their dependence on experts. *Public Underst Sci* 26(8):1003–1018. <https://doi.org/10.1177/0963662516680311>

Schwan S, Bauer D, Kampschulte L, Hampp C (2016) Representation equals presentation? Photographs of objects receive less attention and are less well remembered than real objects. *J Media Psychol* 29(4):176–187. <https://doi.org/10.1027/1864-1105/a000166>

Marc Stadler ist Professor für Kompetenzentwicklung und Kompetenzmodellierung am Institut für Erziehungswissenschaft der Ruhr-Universität Bochum. In seiner Forschung beschäftigt er sich mit Fragen der kritischen Nutzung von Wissenschaftsinformationen in digitalen Informationsumwelten, insbesondere mit der Förderung der Kompetenz zur kritischen Bewertung von Wissenschaftsinformationen bei Kindern, Jugendlichen und jungen Erwachsenen.

Corinna Schuster ist Akademische Rätin a. Z. im Arbeitsbereich Kompetenzentwicklung und Kompetenzmodellierung am Institut für Erziehungswissenschaft der Ruhr-Universität Bochum. In ihrer Forschung beschäftigt sie sich mit dem Transfer metakognitiver Strategien beim selbstregulierten Lernen sowie der Akkuratheit von metakognitiven Verstehensurteilen beim Lesen in digitalen Informationsumwelten.

Open Access Dieses Kapitel wird unter der Creative Commons Namensnennung 4.0 International Lizenz (<http://creativecommons.org/licenses/by/4.0/deed.de>) veröffentlicht, welche die Nutzung, Vervielfältigung, Bearbeitung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden.

Die in diesem Kapitel enthaltenen Bilder und sonstiges Drittmaterial unterliegen ebenfalls der genannten Creative Commons Lizenz, sofern sich aus der Abbildungslegende nichts anderes ergibt. Sofern das betreffende Material nicht unter der genannten Creative Commons Lizenz steht und die betreffende Handlung nicht nach gesetzlichen Vorschriften erlaubt ist, ist für die oben aufgeführten Weiterverwendungen des Materials die Einwilligung des jeweiligen Rechteinhabers einzuholen.

