

# Fallbeispiele: Wisskomm evaluieren

Inspirationen für die Evaluationsgestaltung

Hier klicken, um das Tool zu starten

GEFÖRDERT VOM



Bundesministerium  
für Bildung  
und Forschung

**Veröffentlichung**

Oktober 2023

**Konzeption und Redaktion**

Imke Hedder, Liliann Fischer, Julia Panzer, Eliah Groß



Dieses Tool steht unter folgender Lizenz zur Verfügung: [CC BY-SA 3.0](https://creativecommons.org/licenses/by-sa/3.0/)



# Fallbeispiele: Wisskomm evaluieren

## Inspirationen für die Evaluationsgestaltung

Eine Evaluation zu planen erfordert viele Entscheidungen. Aus welchen Gründen fällt die Wahl auf das eine oder andere Design, diese oder jene Methode? Das sollen die *Fallbeispiele: Wisskomm evaluieren* veranschaulichen.

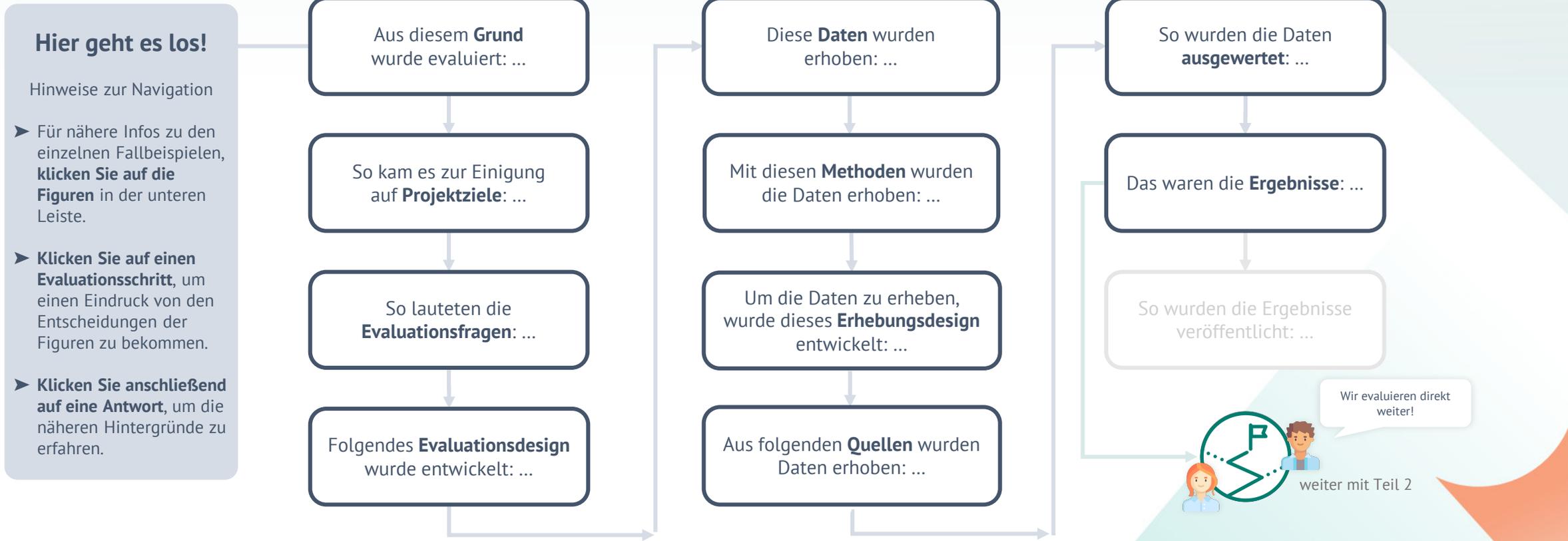
In dieser interaktiven PDF können Sie die Evaluationen unserer Fallbeispiele begleiten und dabei die Entscheidungen, Fehlritte und Erfolge der Evaluierenden mitverfolgen.

Falls Ihnen unbekannte Evaluationsbegriffe begegnen, lohnt sich ein Blick in unseren [Entscheidungsbaum zur Evaluation von Wissenschaftskommunikation](#).

- ▶ Alle Fallbeispiele sind fiktiv.
- ▶ Die Fallbeispiele sollen nicht als „Goldstandard“ verstanden werden. Sie veranschaulichen, wie Evaluationsentscheidungen ausfallen können, die Projektziele und (herausfordernde) Rahmenbedingungen der Evaluation berücksichtigen.
- ▶ Diese Erkundungsversion ermöglicht einen direkten Vergleich der Fallbeispiele und der Entscheidungen, die getroffen wurden. Vorstellungen jedes einzelnen Fallbeispiels finden Sie [hier als Videoreihe!](#)

Weiter

# Der Evaluationspfad



Wir evaluieren direkt weiter!

weiter mit Teil 2

# Der Evaluationspfad (Teil 2)

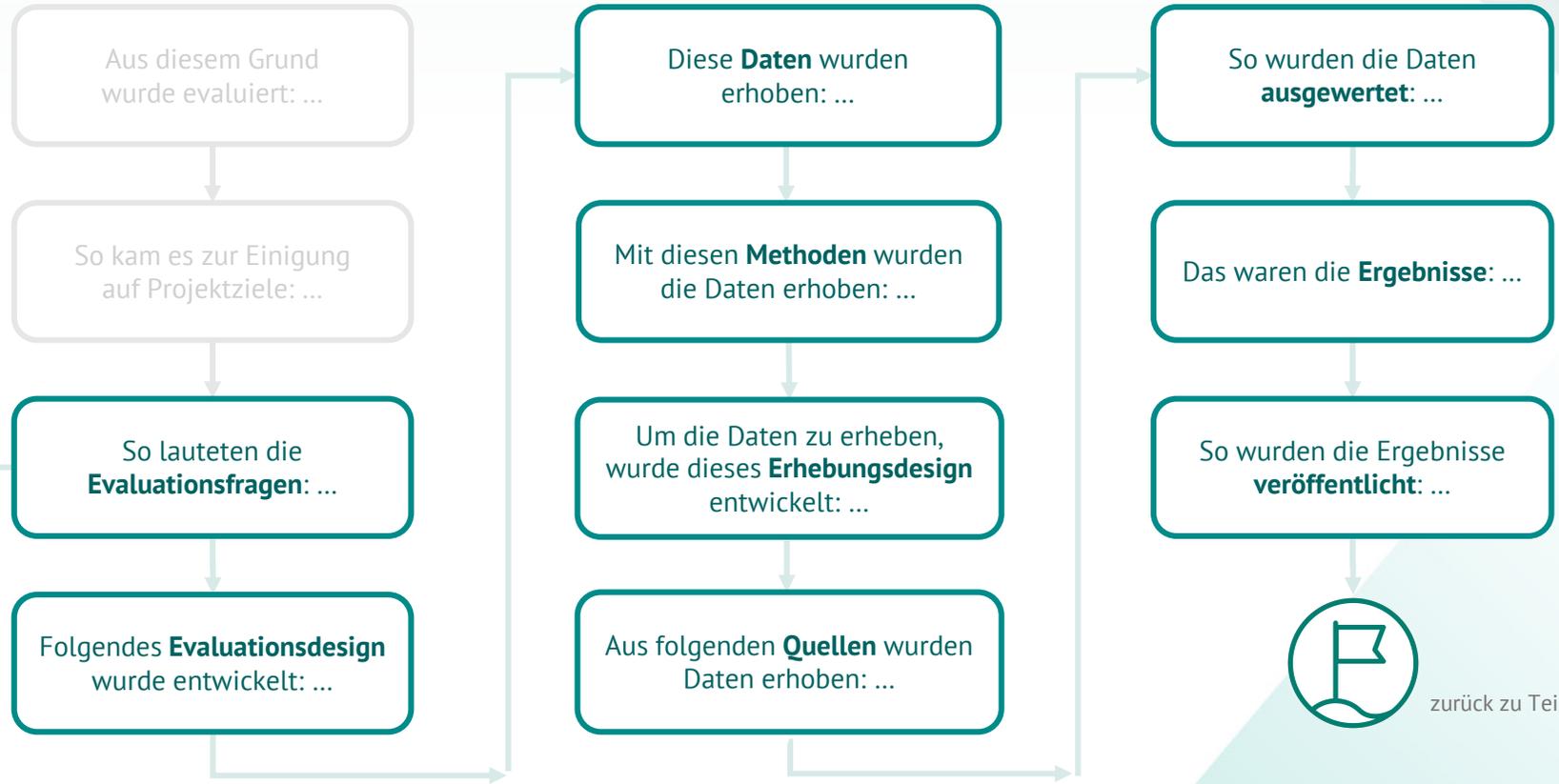


Für uns ergaben sich direkt neue Ideen für eine weitere Evaluationsphase!

## Hier geht es weiter!

Hinweise zur Navigation

- Für nähere Infos zu den einzelnen Fallbeispielen, **klicken Sie auf die Figuren** in der unteren Leiste.
- **Klicken Sie auf einen Evaluationsschritt**, um einen Eindruck von den Entscheidungen der Figuren zu bekommen.
- **Klicken Sie anschließend auf eine Antwort**, um die näheren Hintergründe zu erfahren.



# Kai: Das neue Format

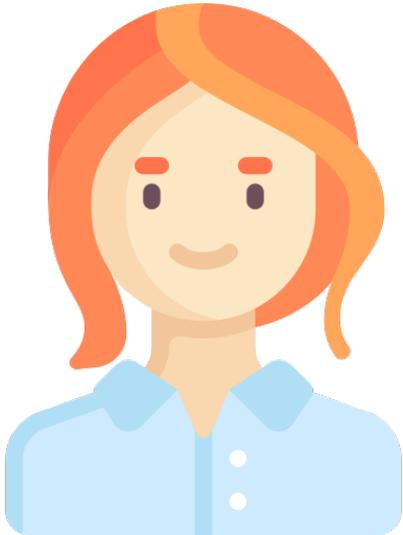


Kai ist Doktorand eines Graduiertenkollegs im Bereich Mathematik. Er belegte ein zweisemestriges **Seminar zu Wissenschaftskommunikation**, in dem er und die anderen Doktorand\*innen des Kurses als Leistungsnachweis ein Wisskomm-Format planten. Sie einigten sich auf die Entwicklung eines Knobelspiels für Jugendliche, das den Spaß an Mathematik anregen und ihr Selbstbewusstsein in dem Fach stärken soll. Kai und zwei weitere Doktorand\*innen kümmerten sich verstärkt um die Evaluation und Dokumentation des Prozesses und bildeten somit das Evaluationsteam.

- **Format:** Knobelspiel
- **Zielgruppe:** Jugendliche zwischen 14 und 16 Jahren mit dem mathematischen Grundlagenwissen der ersten Sekundarstufe; kein gezieltes Interesse für Mathematik nötig
- **Evaluationsanforderung:** anfängertauglich, kostengünstig



## Tabea: Das weit entfernte Ziel



Tabea war für die Evaluation eines **Verbundprojekts verschiedener Hochschulen** verantwortlich. Das Ziel dieses Projekts war es, die Attraktivität „kleiner Studienfächer“ – von Kultur- und Sprachwissenschaften bis hin zu Literatur- und Kunstwissenschaften – an diesen Hochschulen zu erhöhen. Als **externe Evaluatorin** sollte Tabea regelmäßig berichten, ob die Aktivitäten des Projekts zu diesem Ziel beitragen konnten. Bei den Aktivitäten handelte es sich um eine **Social-Media-Kampagne** und regelmäßige **Science-Watch-Partys an Hochschulen**, bei denen populäre Filme gezeigt und anschließend von Wissenschaftler\*innen vor dem Hintergrund ihrer Expertise kommentiert und mit dem Publikum diskutiert wurden.

- **Format:** Social-Media-Kampagne / Science-Watch-Partys
- **Zielgruppe:** Studierende und Studieninteressierte
- **Evaluationsanforderung:** mehrere Aktivitäten gleichzeitig evaluieren



# Rahel: Die unsichtbare Evaluation



Rahel arbeitet in einem **Science Center**, das eine **Ausstellung zu Robotik** entwickelt hat. Die Ausstellung sollte zeigen, wie Roboter künftig verschiedene Lebensbereiche, von der Industrie bis zur Pflege, verändern werden. Mit den Inhalten der Ausstellung sollten Sorgen und Bedenken hinsichtlich des Einsatzes von Robotik adressiert werden. Da das Science Center vornehmlich ein Anlaufpunkt für Familienausflüge ist, sollten die Exponate auch für Kinder und Jugendliche geeignet sein und ihre Faszination für Robotik entfachen.

- **Format:** Ausstellung
- **Zielgruppe:** Familien – zentrale Botschaften der Ausstellung richten sich an Erwachsene, die wenig informiert, skeptisch oder besorgt bzgl. Robotik sind
- **Evaluationsanforderung:** unauffällig und unterhaltsam



# Mark: Die kritische Prüfung



Mark ist Teil des Kommunikationsteams für das **Exzellenzcluster eines Universitätsverbunds**. Das Cluster betreibt Grundlagenforschung im Bereich Gendermedizin. Ein langjähriger Favorit unter ihren Wisskomm-Formaten ist ihre Pub-Science-Reihe: Forschende des Clusters sowie eingeladene Wissenschaftler\*innen treffen dabei in Bars der Stadt auf Bürger\*innen. Dort zeigen sie anhand ihrer Studien sowie interessanter Fälle aus der Medizingeschichte die Bedeutung geschlechterbezogener Unterschiede in Krankheitsbildern auf.

- **Format:** Pub-Science-Reihe
- **Zielgruppe:** Erwachsene jungen und mittleren Alters, interessiert an Medizin; insbesondere Personengruppen, die von geschlechterbasierten Fehldiagnosen betroffen sind
- **Evaluationsanforderung:** standardisierbar, niedrighschwellige Teilnahme





# Start des Evaluationspfads



# Motivation zur Evaluation

Klicken Sie auf eine Antwort, um die näheren Hintergründe zu erfahren.



## Einigung auf Projektziele

Klicken Sie auf eine Antwort, um die näheren Hintergründe zu erfahren.



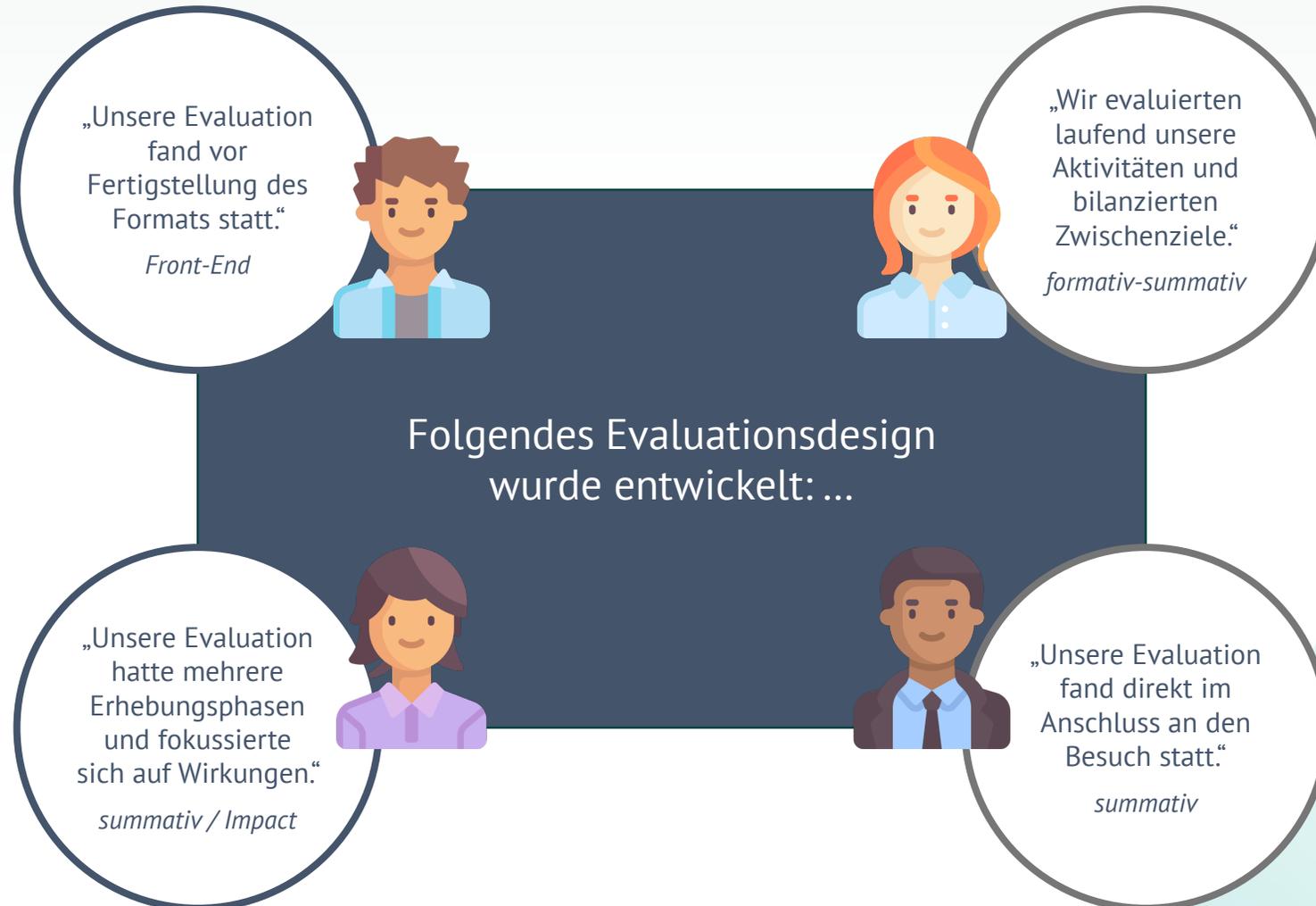
# Evaluationsfragen

Klicken Sie auf eine Antwort, um die näheren Hintergründe zu erfahren.



# Evaluationsdesign

Klicken Sie auf eine Antwort, um die näheren Hintergründe zu erfahren.



## Identifikation nötiger Daten

Klicken Sie auf eine Antwort, um die näheren Hintergründe zu erfahren.



## Methoden und Instrumente

Klicken Sie auf eine Antwort, um die näheren Hintergründe zu erfahren.



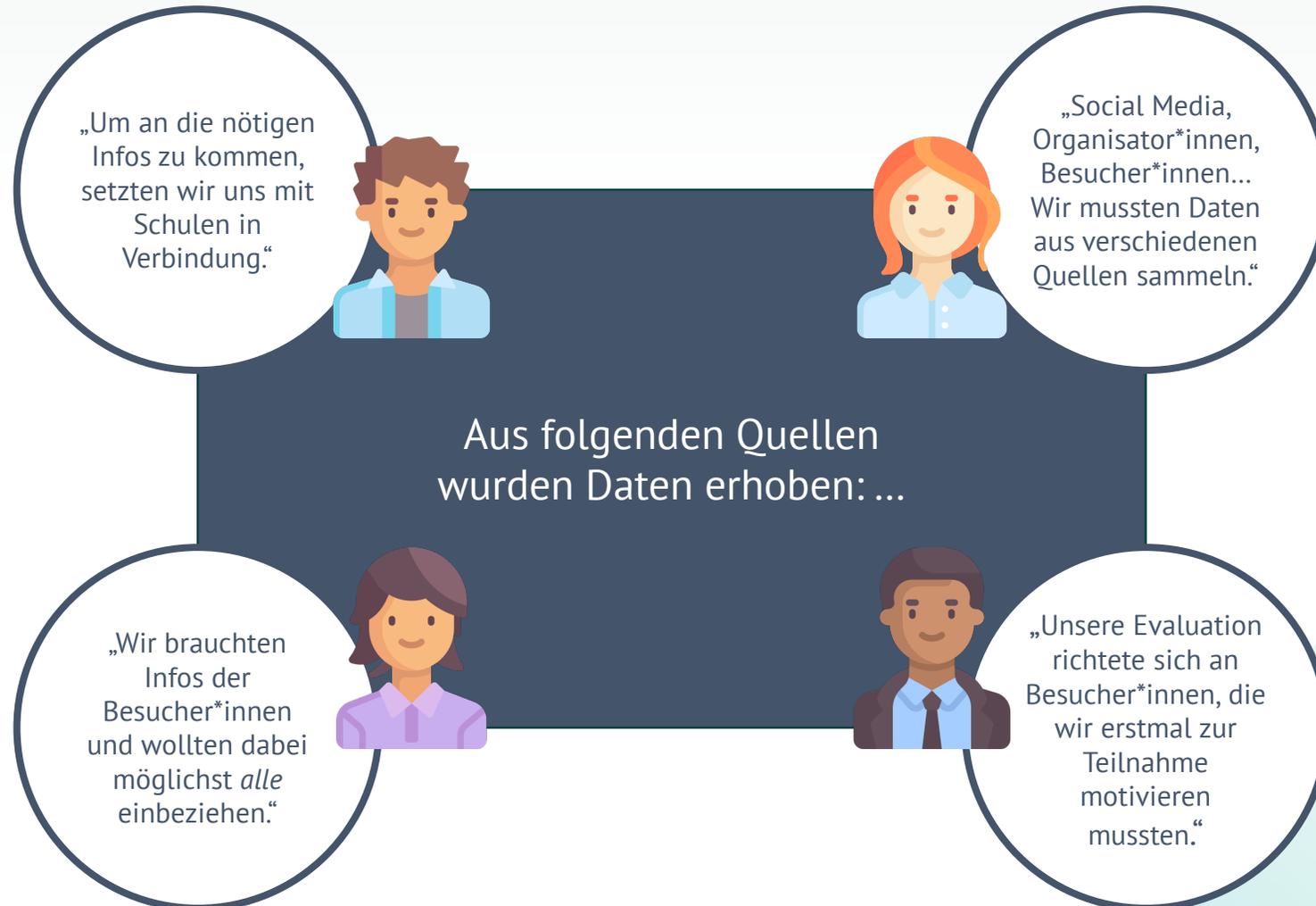
# Erhebungsdesign

Klicken Sie auf eine Antwort, um die näheren Hintergründe zu erfahren.



# Datenquellen und Datenauswahl

Klicken Sie auf eine Antwort, um die näheren Hintergründe zu erfahren.



# Auswertung

Klicken Sie auf eine Antwort, um die näheren Hintergründe zu erfahren.



## Besprechung der Ergebnisse

Klicken Sie auf eine Antwort, um die näheren Hintergründe zu erfahren.





## Erste Evaluationsphase abgeschlossen.

Für Tabea und Kai geht es aber noch weiter!



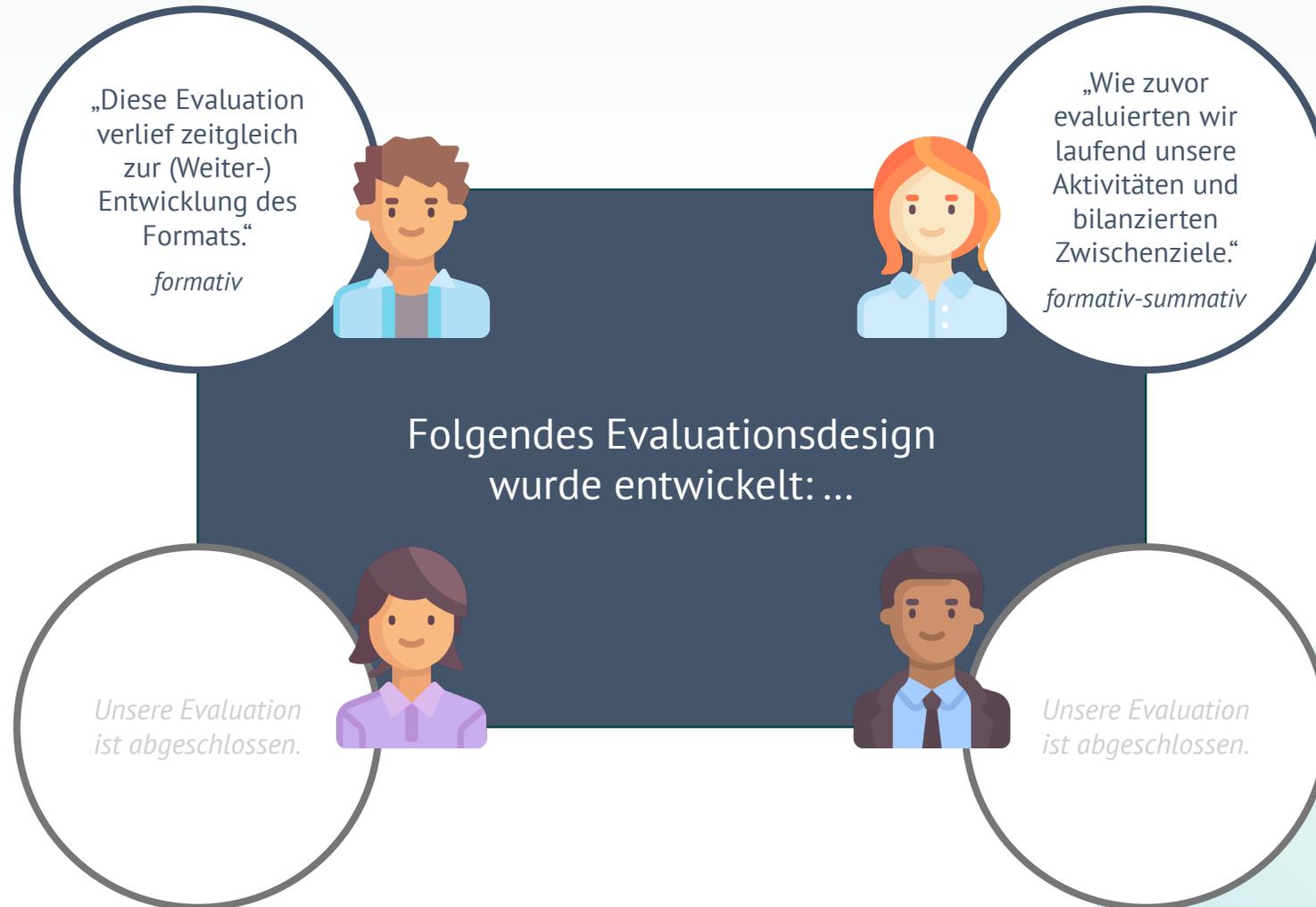
# Evaluationsfragen

Klicken Sie auf eine Antwort, um die näheren Hintergründe zu erfahren.



# Evaluationsdesign

Klicken Sie auf eine Antwort, um die näheren Hintergründe zu erfahren.



## Identifikation nötiger Daten

Klicken Sie auf eine Antwort, um die näheren Hintergründe zu erfahren.



## Methoden und Instrumente

Klicken Sie auf eine Antwort, um die näheren Hintergründe zu erfahren.



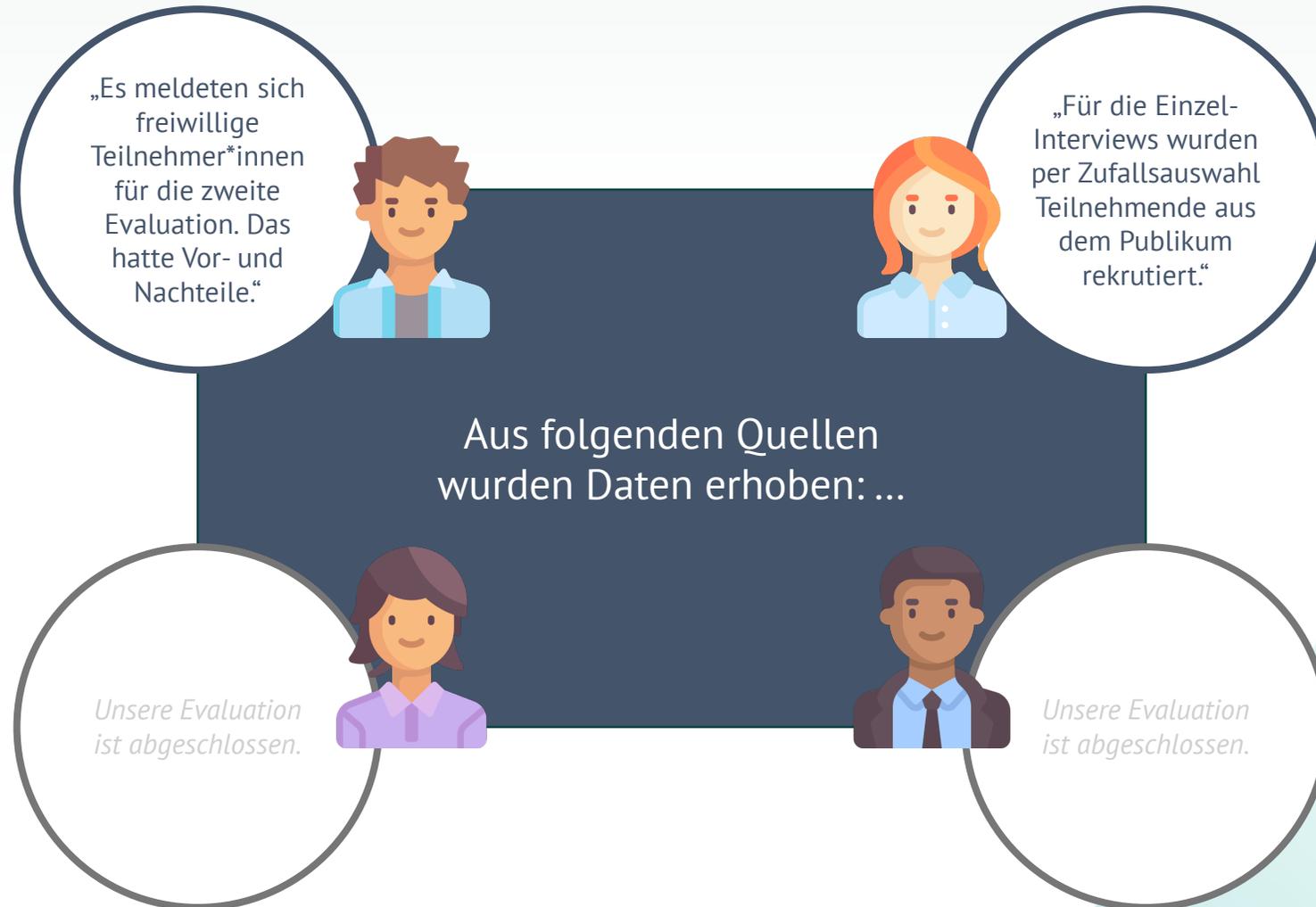
# Erhebungsdesign

Klicken Sie auf eine Antwort, um die näheren Hintergründe zu erfahren.



# Datenquellen und Datenumfang

Klicken Sie auf eine Antwort, um die näheren Hintergründe zu erfahren.



# Auswertung

Klicken Sie auf eine Antwort, um die näheren Hintergründe zu erfahren.



## Besprechung der Ergebnisse

Klicken Sie auf eine Antwort, um die näheren Hintergründe zu erfahren.



## Veröffentlichung der Ergebnisse

Klicken Sie auf eine Antwort, um die näheren Hintergründe zu erfahren.





# Das neue Format

Evaluationsphase 1



# Begleitende Qualitätssicherung

Kai und sein Team hatten bisher **keine Erfahrung** mit Wisskomm oder Spielentwicklung. Sie wollten sichergehen, dass ihre Idee in der Praxis auch wirklich funktioniert und zielgruppengerecht gestaltet ist. Die Evaluation war ein Weg, **die Zielgruppe besser kennenzulernen**, damit das Spiel sie anspricht.

**Unbeabsichtigte Effekte wollte das Team unbedingt vermeiden**; zum Beispiel, dass das Spiel zu kompliziert oder langweilig wird und die Jugendlichen stattdessen entmutigt, sich mit Mathematik zu. Indem sie während der Entwicklung des Formats die Meinungen und Reaktionen der Jugendlichen einbezogen und Testdurchläufe organisierten, wollten sie diese Risiken eingrenzen.



# Herleitung aus der Forschung

Bevor sich das Team an die Arbeit machte, wurde viel darüber diskutiert, welche **Ziele sich realistischerweise** mit einem Spiel erreichen lassen (→ [Worauf kann Wisskomm wirken?](#)). Der Überblick zum Forschungsstand verdeutlichte ihnen, dass übergeordnete Ziele (wie die häufigere Wahl von Mathematik als Vertiefungsfach) von vielen Faktoren abhängen: zum Beispiel von Voreinstellungen und emotionalen Empfindungen zum Fach. **Deshalb entwickelten sie zunächst Annahmen**, wie ein Spiel Gedanken und Gefühle bezüglich Mathematik positiv beeinflussen könnte. Daraus entstanden die erhofften Outputs, Outcomes und Impacts des Projekts, die logisch aufeinander aufbauen und somit eine „**Wirkungskette**“ bilden sollten.



Output <i>(erhoffte Leistungen des Spiels)</i>	Outcome <i>(erhoffte Wirkung bei der Zielgruppe)</i>	Impact <i>(erhoffte gesellschaftliche Wirkung)</i>
<ul style="list-style-type: none"> <li>• Spielende <i>üben mathematische Prinzipien</i> und Denkweisen</li> <li>• Das Spiel schafft <i>Erfolgslebnisse</i> mit Mathematik</li> <li>• Das Spiel macht <i>Spaß</i>, dank einnehmendem, witzigem Storytelling</li> </ul>	<ul style="list-style-type: none"> <li>• Spielende merken, <i>dass sie dazu fähig sind</i>, mathematische Rätsel zu lösen</li> <li>• Spielende verbinden Mathematik mit einem <i>spassigen Erlebnis</i></li> <li>• Spielende erkennen, wie sie mathematische <i>Denkweisen</i> außerhalb des Matheunterrichts <i>einsetzen</i> können</li> <li>• Spielende entwickeln <i>positivere Einstellungen</i> zum Fach</li> </ul>	<ul style="list-style-type: none"> <li>• Mathematik verliert bei Schüler*innen den <i>Ruf</i>, ein langweiliges Fach zu sein</li> <li>• Mehr Schüler*innen wählen <i>Mathematik als Vertiefungsfach</i></li> </ul>

# Die Zielgruppe kennenlernen

Aus Erinnerungen an die eigene Schulzeit wussten Kai und das Team zwar, dass viele Jugendliche Mathematik langweilig oder schwer verständlich finden. **Die konkreten Herausforderungen dieser Zielgruppe mit dem Fach kannten sie allerdings nicht.**

Auch waren sie unsicher, mit welchen Arten von Spielen man Jugendliche begeistern kann. Das führte sie zu folgenden Evaluationsfragen:

- *Welche Themenschwerpunkte im Fach Mathematik fallen unserer Zielgruppe leicht und welche machen ihnen Spaß?*
- *Welche Themenschwerpunkte im Fach Mathematik fallen ihnen schwer und welche langweilen sie?*
- *Welche Vorstellungen über das Fach Mathematik bringen sie mit?*
- *Welche Art von Spielen machen ihnen Spaß?*



# Evaluation während der Projektplanung

Damit die Evaluationsergebnisse rechtzeitig im Spieldesign berücksichtigt werden konnten, z. B. um den richtigen Schwierigkeitsgrad der Rätsel festzulegen, wurden die **Erhebungen vor der Fertigstellung der ersten Spielversion** durchgeführt. Damit entschied sich Kais Kurs für eine **Front-End-Evaluation**; also eine Evaluation, die in der Projektentwicklung bzw. aus Sicht der Zielgruppe vor dem Start der Maßnahme stattfindet.



# Einstellungen der Zielgruppe

Um die Evaluationsfragen zu beantworten, mussten Kai und seine Kolleg\*innen versuchen, sich in die **Köpfe und Gefühlswelten ihrer Zielgruppe** hineinzusetzen, denn die Fragen drehten sich besonders um ihre **Präferenzen und Beurteilungen** von Mathematik und ihr Verhältnis zu dem Fach.

Das war allerdings leichter gesagt als getan: Die große Herausforderung des Teams war es, Vorstellungen von Mathematik zu definieren und **sicherzugehen, dass alle Mitglieder im Wisskomm-Kurs darunter dasselbe verstehen** und nicht unwissentlich unterschiedliche Ziele verfolgen. Denn nur, wenn das Projektteam sich darüber einig ist, welche Vorstellungen relevant sind und wie Daten ihre Annahmen widerlegen oder bestätigen können, kann die Evaluation ihren Zweck erfüllen. Sie brachen den Begriff herunter auf drei für sie relevante Aspekte: spontane Assoziationen mit Mathematik (die u. a. Vorurteile abbilden könnten), die Bewertung des Fachs und die Selbsteinschätzung der eigenen Fähigkeiten in dem Fach.



# Qualitative Interviews und „Lautes Denken“

Da das Team wenig Vorwissen zu den Einstellungen und Empfindungen der Zielgruppe hatte, wollten sie als Erstes mit Jugendlichen ins Gespräch kommen. Eine weitere Idee war, erste Rätsel mitzubringen und gemeinsam zu spielen, um zu verstehen, wie Jugendliche den mathematischen Problemen begegnen und wo die Herausforderungen in den Rätseln liegen. Dabei nutzte das Team Methoden, die auch häufig in sozialwissenschaftlicher Forschung Anwendung finden: So entschieden sie sich für **Einzelinterviews** mit offenen, vorab festgelegten Fragen (sog. qualitative Interviews), die es ermöglichen, an spannenden Stellen des Gesprächs **näher einzuhaken**. Der zweite Teil des Gesprächs sollte das **Test-Rätsel** beinhalten, bei dem die Interviewenden die Teilnehmenden **beim Spielen beobachten** und sie **laut beschreiben lassen**, welche Gedanken ihnen kommen (sog. „Lautes-Denken-Methode“).



# Die Auswahl des richtigen Zeitpunkts im Projektplan

Während einige Kursmitglieder damit begannen, die Spielerätsel zu entwickeln, **plante das Evaluationsteam in den ersten Wochen nach dem Kursbeginn bereits die Interviews**. Etwa zehn Wochen nach dem Start des Kurses sollten die Gespräche stattfinden. Das ließ dem Zeitplan zufolge noch dreieinhalb Monate bis zur Fertigstellung der ersten Spielversion. Das Team einigte sich darauf, dass die **Evaluationsergebnisse spätestens zwei Wochen nach den Erhebungen vorliegen sollten**, damit sie direkt eingearbeitet werden können.



# Kriterien für eine vielfältige Stichprobe

Mit der Projektidee ließen sich leicht Schulen gewinnen. **Motivierte Mathematik- und Informatik-Lehrkräfte** waren besonders hilfreich für die erste Kontaktaufnahme zu möglichen Interviewpartner\*innen. Kai schätzte mit Blick auf den Interviewleitfaden ab, dass die drei Interviewer\*innen **in einem zweistündigen Zeitfenster jeweils fünf Interviews führen** könnten.

**Dem Team war es wichtig, möglichst unterschiedliche Personen zu befragen.** Schließlich wollten sie nicht (versehentlich) ein Spiel nur für eine bestimmte Gruppe von Jugendlichen entwickeln. Für die Stichprobenauswahl (→ [Teil 3 der How-To-Reihe: Wisskomm evaluieren](#)) einigten sie sich auf **ein paar zentrale Kriterien**: Jede\*r Interviewer\*in sollte an einer unterschiedlichen Schule befragen, die gewählten Schulen sollten in verschiedenen Bezirken ihrer Region liegen und es sollten ähnlich viele Mädchen wie Jungen befragt werden. So sollten unterschiedliche soziale Gruppen erreicht und die Folgen möglicher gesellschaftlicher Vorurteile gegenüber den mathematischen Fähigkeiten von Mädchen berücksichtigt werden.

Basierend auf diesen Kriterien entstand eine Liste möglicher Interviewpartner\*innen. Dann entschied ein **Losverfahren**. Da es sich bei den Schüler\*innen um Minderjährige handelt, wurden aus **Datenschutzgründen** neben den Gelosten auch ihre Eltern um Zustimmung gebeten.



# Kategorisierung und Vergleich der Aussagen

Alle Interviewer\*innen nahmen die **Audiospur ihrer Gespräche** (inklusive der Beschreibungen aus der „Lautes-Denken-Übung“\*) auf und **tippten die Antworten nachträglich** ab. Stichworte reichten ihnen aus, weil dem Team die Zeit für eine tiefergehende Analyse der genauen Wortwahl fehlte. Dann wurden die Antworten verglichen, um **Gemeinsamkeiten und personenspezifische Besonderheiten** herauszuarbeiten. An einigen Stellen wurden die genannten Inhalte ausgezählt, z. B. welche Themenschwerpunkte am häufigsten als langweilig, schwierig, interessant oder leicht erschließbar genannt wurden. Die Assoziationen mit dem Fach wurden **nach positiven und negativen Gefühlen bzw. Wertungen gruppiert**. So bekam das Evaluationsteam einen Eindruck davon, welche Urteile über Mathematik hervorstachen, auch wenn nicht alle Personen 1-zu-1 dasselbe sagten.

\* siehe „Mit diesen Methoden wurden die Daten erhoben“, Evaluationsphase 1 (S. 36)



# Grundlage für die nächste Evaluation

Die Ergebnisse wurden im Kurs diskutiert. Die Interviews zeigten auf, mit welchen Themen die Schüler\*innen häufig Schwierigkeiten hatten und welche sie leicht meisterten. Gleichzeitig lernte der Kurs die Lieblingsspiele der Teilnehmenden kennen, die neben Multiplayer-Videospielen auch Escape Games und Strategiespiele umfassten, und was ihnen an diesen Spielen Spaß machte.

**Die Evaluation half ihnen dabei, die Umsetzung des Formats auf die Zielgruppe anzupassen:** Das Spiel müsse den Wert des logischen Denkens hervorheben und eine einnehmende Geschichte erzählen. Es sollte in Teams oder gemeinsam statt gegeneinander gespielt werden, um aktuelle Spielrends aufzugreifen und keinen Leistungsdruck auszuüben. **Die Ideen wurden in die weitere Entwicklung integriert.** Kai und sein Team einigten sich direkt darauf, **eine weitere Evaluationsphase zu planen:** Die ersten Testspiele sollten beobachtet werden, um zu sehen, ob diese Spielelemente die gewünschten Reaktionen auslösten oder nicht.





# Das neue Format

## Evaluationsphase 2

Weiter



# Reaktionen auf das Spiel

Diesmal interessierten das Team die Reaktionen der Teilnehmenden beim Spielen und die allgemeine Frage, inwieweit die aktuelle Version die **anvisierten Outputs** erreichte, welche die **Grundvoraussetzungen für die erhoffte Wirkung** wären.

Daraus ergaben sich konkret folgende Evaluationsfragen:

- *Wie reagieren die Teilnehmenden auf das Spiel?*
- *Wie bewerten die Teilnehmenden das Spiel? Macht es ihnen Spaß?*
- *Ist das Spiel verständlich oder entstehen Fragen und Missverständnisse?*
- *Ist das Spiel lösbar? Erscheint es uns plausibel, dass das Spiel die Erfolgserlebnisse bietet, die es schaffen soll?*



<b>Output</b> (erhoffte Leistungen des Spiels)	<b>Outcome</b> (erhoffte Wirkung bei der Zielgruppe)	<b>Impact</b> (erhoffte gesellschaftliche Wirkung)
<ul style="list-style-type: none"> <li>• Spielende <i>üben mathematische Prinzipien</i> und Denkweisen</li> <li>• Das Spiel schafft <i>Erfolgserlebnisse</i> mit Mathematik</li> <li>• Das Spiel macht <i>Spaß</i>, dank <i>einnehmendem, witzigem Storytelling</i></li> </ul>	<ul style="list-style-type: none"> <li>• Spielende merken, dass sie dazu <i>fähig sind</i>, mathematische Rätsel zu lösen</li> <li>• Spielende <i>verbinden Mathematik mit einem spaßigen Erlebnis</i></li> <li>• Spielende erkennen, wie sie <i>mathematische Denkweisen außerhalb des Matheunterrichts einsetzen können</i></li> <li>• Spielende <i>entwickeln positivere Einstellungen zum Fach</i></li> </ul>	<ul style="list-style-type: none"> <li>• Mathematik <i>verliert bei Schüler*innen den Ruf</i>, ein langweiliges Fach zu sein</li> <li>• <i>Mehr Schüler*innen wählen Mathematik als Vertiefungsfach</i></li> </ul>

# Evaluation während der Weiterentwicklung

In der zweiten Evaluationsphase sollte der Spielentwurf des Wisskomm-Kurses getestet werden. Damit setzte die Evaluation an, **als die erste Spielversion von der Zielgruppe ausprobiert wurde**, mit dem Ziel, diesen „Prototypen“ noch zu verbessern. Dieser Ansatz lässt sich als **formative Evaluation** beschreiben, bei der ein Projekt in seiner Weiterentwicklung evaluiert wird, um ggf. Anpassungen vorzunehmen, welche die Zielerreichung wahrscheinlicher machen. In Kais Fall lag der Fokus auf der Prüfung, ob die aktuelle Spielversion Erkenntnisse aus der ersten Evaluationsphase bereits erfolgreich umsetzen konnte. Das würde die Chance erhöhen, dass die anfangs entwickelte Wirkungskette\* tatsächlich durchlaufen wird.

\* siehe „So kam es zur Einigung auf Projektziele“, Evaluationsphase 1 (S. 32)



# Spontane Reaktionen

Kai und sein Team waren vor allem auf die **Reaktionen** und **Kommentare** der Teilnehmenden beim Spielen gespannt, die Hinweise auf das Erzielen oder Verfehlen der erhofften Wirkung geben: alles von **Gesichtsausdrücken, Lachen, Verständnisfragen** bis hin zu **Ausrufen**, die für Spaß und das Erleben von Erfolgsmomenten oder Unverständnis und Frustration sprechen können.



# Beobachtung und Feedbackrunde

Da das Spiel noch nie getestet wurde und sie selbst keine Vorkenntnisse hatten, wie es bei der Zielgruppe ankommen würde, erschien dem Team ein qualitativer, ergebnisoffener Ansatz auch diesmal sinnvoll. In diesem Fall sollte eine **Beobachtung** im Zentrum stehen. Vor Ort waren 25 Schüler\*innen, die sie in fünf Spielgruppen aufteilten\*. Jede Gruppe wurde von einem Kursmitglied beobachtet. **Kai kündigte in den Gruppen an, dass er und andere Kursmitglieder sich während des Spiels Notizen machen** würden, sie hielten sich aber im Hintergrund auf und kamen nur bei Fragen zur Hilfe. Die Beobachtungssituation schien nach wenigen Minuten bei den Jugendlichen vergessen zu sein, auch wenn sich nicht ausschließen lässt, dass sich die Beobachtung auf das Spielverhalten der Teilnehmenden auswirkte.

Das Testspiel dauerte 45 Minuten und endete mit einer zehnmütigen **Feedback-Runde**. Hier fragte das Team nach Verständnisschwierigkeiten und bat um die Bewertung des Spiels und Verbesserungshinweise.

\* Auf die Auswahl der Schüler\*innen wird im weiteren Verlauf eingegangen („Aus folgenden Quellen wurden die Daten erhoben“, Evaluationsphase 2, S. 47).



# Erhebungen während und nach der Maßnahme

Da sich das Team für die direkte, spontane Reaktion der Jugendlichen beim erstmaligen Spielen interessierte, wurden alle Daten direkt **beim Spieltreffen** gesammelt. Die Beobachtung erfolgte dabei **während der Maßnahme** (dem Testspiel) und die Feedbackrunde fand direkt **im Anschluss an das Testspiel** statt.



# Folgen von Freiwilligkeit

Als Dank für ihre Unterstützung in den Interviews wurden zunächst die Interviewpartner\*innen aus der ersten Evaluationsphase zum Spielen eingeladen. Das Interesse an dem Projekt war allerdings in den Klassen größer geworden, sodass das Team kurzerhand zwei Spieltreffen **mit allen Freiwilligen** veranstaltete. Damit waren die Teilnehmenden weniger systematisch ausgewählt als in der ersten Evaluationsphase. Da sie jedoch für die erste Kritik an dem Spiel besonders investierte Spieler\*innen brauchten, nahmen Kai und das Team **mögliche damit einhergehende Verzerrungen** in Kauf.

Jedes Testspiel endete mit einer Feedbackrunde, allerdings fanden **bei nur einem Termin gezielte Beobachtungen statt, da das Team nicht die Kapazitäten hatte, Notizen mehrerer Treffen auszuwerten**. Bereits der erste Beobachtungstermin bot sehr viel Material zum Auswerten.



# Auswertung der Beobachtungsbogen

Während des Spiels machten sich die Beobachter\*innen Notizen in ihren Beobachtungsbogen. Der **Beobachtungsbogen beinhaltete drei Kategorien:** „Auffällige Reaktionen“ (Gesichtsausdrücke, Ausrufe und Kommentare o. ä.), „Fragen der Spielenden“ und „Sonstiges“, worunter zum Beispiel eine Beobachterin Gruppendynamiken notierte, ein anderer Beobachter verschiedene Spielstrategien, die er in den Teams erkannt hatte. Zwei Personen **dokumentierten während der Feedbackrunde stichpunktartig die Diskussion.**

Das Evaluationsteam las alle Beobachtungsnotizen, **grupperte die Inhalte nach Themen** (ähnlich wie bei den Interviews) und sortierte sie in diesen Themenbereichen nach positiven bzw. erhofften Reaktionen sowie negativen bzw. ungewünschten Reaktionen, die auf Verbesserungspotenzial hindeuten könnten. Diese Vermutungen verbanden sie mit dem abschließenden Äußerungen aus den Feedbackrunden, die zumindest einige ihrer Interpretationen während der Beobachtungen bestätigten.



# Letzte Anpassungen und Tipps für neue Projekte

Das Evaluationsteam **stellte die Ergebnisse nochmals dem Kurs und der Kursleitung vor**. Auch wenn sie an vielen Stellen zufrieden mit dem Feedback und den Beobachtungen der Testspiele waren, lief letztendlich nicht alles wie in der Theorie geplant. Nicht alle Gruppen schafften es, das Spiel zu Ende zu spielen, und auch die Rahmenhandlung zeigte einige Lücken auf, worauf die Jugendlichen in der Feedbackrunde aufmerksam machten. Trotzdem war das Team im Allgemeinen zufrieden: Das Spiel machte einem Großteil der Jugendlichen Spaß und forderte ihr logisches, mathematisches Denken.

**In der Ergebnisdiskussion entstanden viele Ideen, wie das Spiel verbessert werden könnte**, um die Outputs gezielter zu erreichen. Nicht alle von ihnen konnten im Rahmen des Kurses noch umgesetzt werden, da sie umfangreichere Überarbeitungen verlangt hätten. Während der restliche Kurs die möglichen Änderungen noch umsetzte, kümmerte sich Kai um die Fertigstellung des **Projektberichts**, der die gesammelten Erfahrungen und neuen Ideen dokumentieren und für jene verfügbar machen sollte, die ähnliche Projekte verfolgen.



# Bericht, Blog und Tagungsbeiträge

Die **Veröffentlichung des Projektberichts und der offen lizenzierten Spielanleitung** wurde mit einem **Blogbeitrag auf der Universitätswebsite** begleitet. Damit erhoffte sich das Team zunächst, Interessierte in der Wissenschaft und Wissenschaftskommunikation zu erreichen. Vor allem die Problemanalyse, was Jugendliche im Mathematikunterricht entmutigt und wie das Spiel diese Probleme zu adressieren versuchte, stieß dabei auf großes Interesse. Deshalb reichte Kais Team das Spiel und die Evaluationserkenntnisse für verschiedene **Tagungen zur Vermittlung von MINT-Fächern** ein – in der Hoffnung, damit auch Lehrer\*innen zu erreichen, die diese Erkenntnisse in ihrem Unterricht nutzen könnten.





# Das weit entfernte Ziel

Evaluationsphase 1



# Evaluation als Förderbedingung

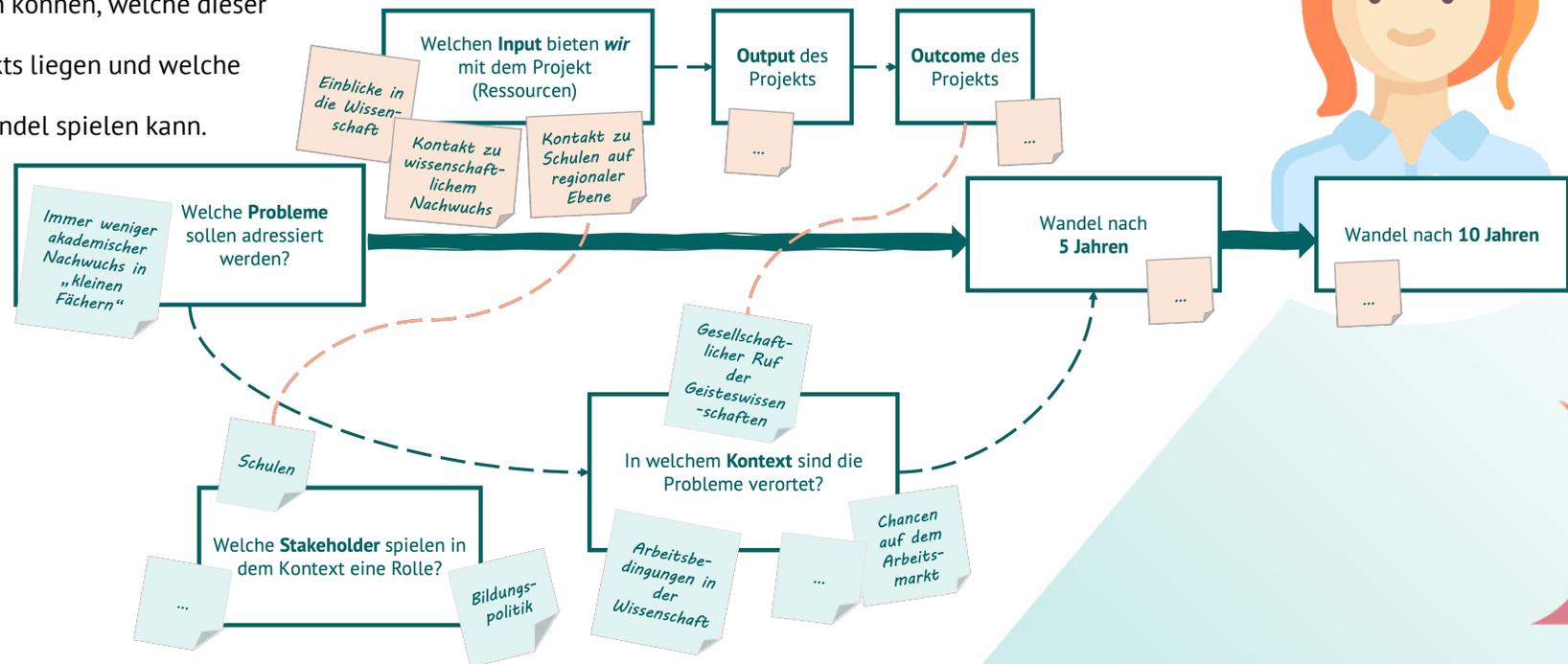
Das Projekt wurde von einer Stiftung finanziert, die sich in verschiedenen Projekten der Förderung von Geisteswissenschaften widmet. **Zu den Bedingungen für die Projektförderung gehörte ein jährlicher Evaluationsbericht darüber, inwieweit „Kleine Fächer stärken“ zur Erreichung dieses Stiftungsziels beiträgt.** Tabea wurde angestellt, um die Evaluation durchzuführen und die Ergebnisse an das Projektteam und die Stiftung weiterzugeben.



# Entwicklung einer Theory of Change

Tabea begann ihre Arbeit mit einem Workshop, der Vertreter\*innen der Stiftung und der beteiligten Hochschulen zusammenbrachte, um gemeinsam über die angestrebten Wirkungen des Projekts zu sprechen und eine **Theory of Change** zu entwickeln. Der Theory-of-Change-Ansatz sollte dabei helfen, explizit zu formulieren, **wie die Projektaktivitäten zu dem von ihnen angestrebten Wandel beitragen können** (→ [Schritte zur eigenen Theory of Change](#)). Dabei wurde diskutiert, welche Faktoren und Stakeholder den Wandel ermöglichen können, welche dieser Faktoren im Wirkungskreis des Projekts liegen und welche Rolle das Projekt somit in diesem Wandel spielen kann.

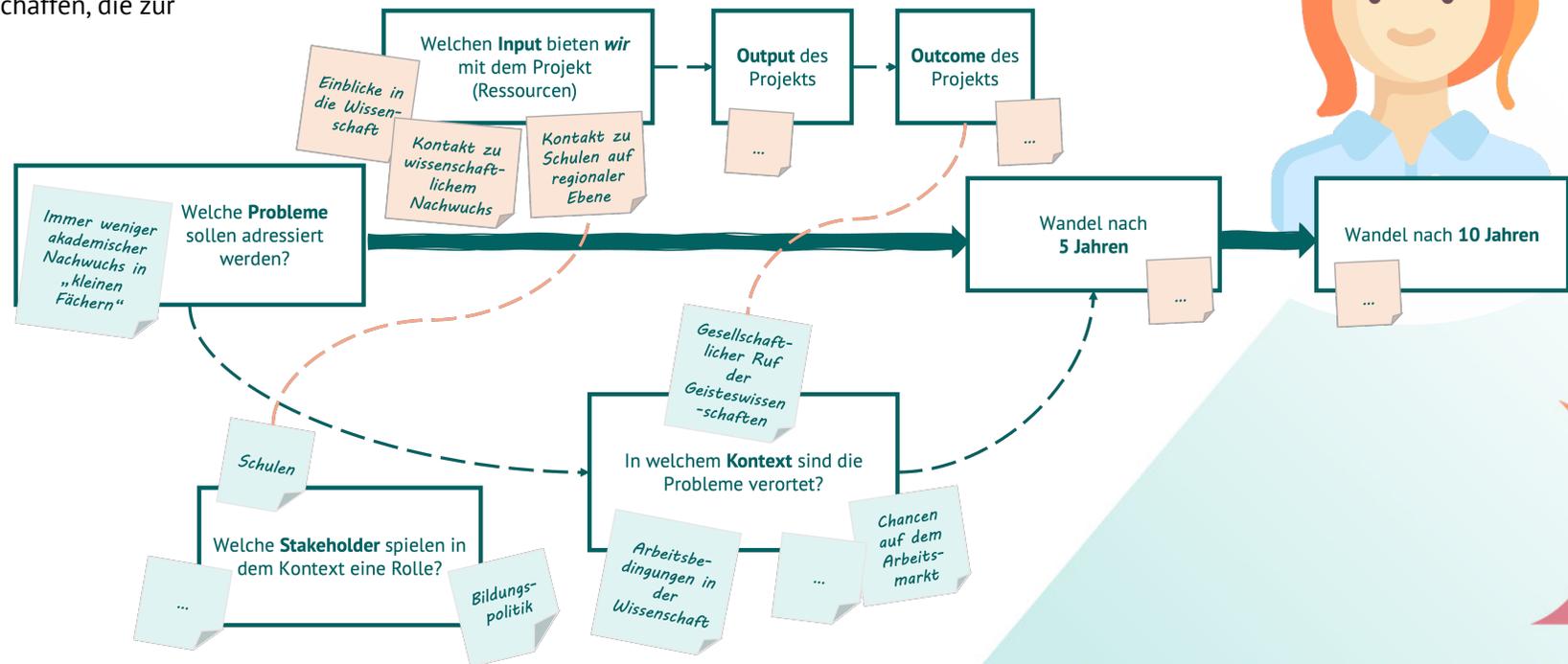
weiterlesen →



# Entwicklung einer Theory of Change

Das Team entwickelte im Workshop abschließend **Annahmen, wie das Projekt „Kleine Fächer stärken“ einen Beitrag zum Wandel leisten könnte**. Da die Projektumsetzenden im Hochschulsektor ansässig waren, konnten sie leicht Kontakt zu Studierenden und Studieninteressierten aufnehmen. So sahen sie ihre Rolle darin, diese gezielt zu informieren und Möglichkeiten zu schaffen, mit Forschenden aus den Fachbereichen zu diskutieren. Ihre Kommunikationsformate sollten die Wissensgrundlage und persönliche Erfahrungen mit den Fachbereichen schaffen, die zur Studienmotivation beitragen.

← zurück



# Prüfung bestehender Projektannahmen

Die Evaluation sollte überprüfen, ob das Projekt wirklich zur Stärkung kleiner Fächer beitragen konnte. Um das Projekt beurteilen zu können, musste Tabea **die zentralen Annahmen testen, die in der Theory of Change formuliert wurden:**

- *Die Social-Media-Kampagne informiert Studierende und Studieninteressierte über kleine Fächer an den beteiligten Hochschulen sowie die Inhalte dieser Fächer.*
- *Die Science-Watch-Partys zeigen dem Publikum aus Studierenden und Studieninteressierten, wie das Wissen der kleinen Fächer auf aktuelle gesellschaftliche Fragen und populäre Themen (dargestellt in Hollywood-Filmen) angewandt werden kann. So steigern sie das Interesse an den Fächern und die Motivation der Zielgruppe, diese Fächer zu studieren.*

**Das Projektteam hatte sich auf diese Annahmen als überprüfbare Projektziele geeinigt.** Dabei war ihnen aber auch bewusst, dass sich diese **im Verlauf des Projekts noch weiterentwickeln und verändern könnten.** Deshalb verständigten sie sich darauf, **regelmäßig im Austausch über den Stand der Evaluation zu bleiben.**



# Formative und summative Evaluation

Tabeas Vorhaben legt eine teils formative, teils summative Evaluation nahe. Für die **formative Evaluation, die Projektaktivitäten dokumentiert**, sollten laufend Daten gesammelt werden, besonders quantitative Daten zu Social-Media-Aktivitäten und Reaktionen von Followern sowie zur Teilnahme an den Veranstaltungen.

Mit der **summativen Evaluation, die den Fokus auf die Ergebnisse eines Projekts** richtet, sollten regelmäßig Zwischenbilanzen gezogen werden, inwieweit die Projektziele bisher erfüllt wurden. Neben der Analyse der oben genannten quantitativen Daten spielten für die summative Evaluation auch die Reaktion des Publikums auf die Veranstaltungen eine große Rolle.

Diese Zwischenbilanzen waren auch eine Gelegenheit für Tabea, formative und summative Erkenntnisse gemeinsam zu betrachten: Insbesondere, inwieweit die Gestaltung der Aktivitäten zu den erwarteten Projektergebnissen führt. (Mehr Informationen zu summativen und formativen Evaluationen liefert der [Entscheidungsbaum unter „Evaluationsdesigns“](#))



# Dokumentation und Resonanz

Der formative Teil der Evaluation beruhte auf Daten zur **Beschreibung der Projektaktivitäten**. So wertete Tabea aus, welche Inhalte auf den Social-Media-Kanälen verbreitet wurden, beispielsweise Werbung für die Science-Watch-Partys und Info-Posts über die verschiedenen Fächer. Sie interessierte sich auch für Social-Media-Kennwerte wie **Followerzahlen, Klickzahlen und Anzahl der Interaktionen**, um die Reichweite der Beiträge zu messen. Bei den Veranstaltungen sollten die Publikumszahlen und Merkmale der anschließenden Diskussion (z. B. die Dauer der Fragerunden) festgehalten werden. Neben solchen quantitativen Daten interessierte Tabea aber auch eine qualitative Beurteilung zum Ablauf der Veranstaltung. Deshalb wollte sie **Einschätzungen der Organisator\*innen und der wissenschaftlichen Gäste** einholen. Schließlich zählte zur formativen Evaluation noch eine Resonanzprüfung, die beispielsweise die Erfassung von Medienberichten über das Projekt einschloss.

Im summativen Teil der Evaluation sollte eine Bilanz zur **Zielerreichung bei der Zielgruppe** erfolgen. Hierbei sollte es darum gehen, wie wichtig und interessant diese Fächer nun wahrgenommen wurden und ob die Teilnehmenden sich vorstellen könnten, diese Fächer einmal zu studieren.



# Auszählungen, Interviews und Feedbackkarten

Die benötigten quantitativen Daten, von Social Media über Medienbeiträge bis hin zu Veranstaltungsinformationen, ließen sich leicht **auszählen**. Publikumszahlen und der Ablauf der Veranstaltung wurden **beobachtet**.

Mit den wissenschaftlichen Gästen und dem Organisationsteam führte Tabea **kurze Debriefing-Gespräche** zu ihrer Einschätzung der Veranstaltungen durch. Dabei ging es einerseits darum, ob sie insgesamt zufrieden waren, und andererseits um ihre Einschätzung der Qualität der Fragen, die vom Publikum in den Fragerunden gestellt wurden.

Die Einstellungen von Teilnehmenden der Science-Watch-Partys wurden über **Feedbackkarten zum Ankreuzen** erhoben. Diese fragten ab, wie bekannt die behandelten Fächer dem Publikum waren, sowie das Interesse und die Studienmotivation.



# Ein Monitoring-Plan zur standardisierten Erhebung

Die Datenerfassung der Social-Media-Analytics und der medialen Berichterstattung sollte in regelmäßigen Abständen erfolgen, um mögliche Entwicklungen im Verlauf zu untersuchen. Dazu entwickelte Tabea einen Monitoring-Plan, in dem die geplanten Zeitpunkte festgehalten wurden. Darin waren auch genaue Vorgaben zur Durchführung der Erhebung und zur Dokumentation von Veranstaltungen festgehalten, um die Datensammlung einheitlich zu gestalten.

Die Befragung der Teilnehmenden der Veranstaltungen über Feedbackkarten wurde ebenfalls standardisiert und fand wiederholt statt. Allerdings wusste Tabea, dass diese Erhebung dadurch eingeschränkt war, dass **jedes Mal unterschiedliche Personen befragt wurden**. Deshalb war in dem Fall **keine Erfassung einer Entwicklung**, sondern nur ein Vergleich zwischen verschiedenen Gruppen und Veranstaltungen möglich



# Von Social Media bis Studienberatung

Mit Organisator\*innen, Gästen, Studierenden und Medien kam für Tabea eine große Anzahl an Datenquellen zusammen, über die sie den Überblick behalten musste. Für die Debriefings mit den wissenschaftlichen Gästen und Organisator\*innen war es wichtig, frühzeitig und **idealerweise vor der Veranstaltung eine Anfrage für ein anschließendes Gespräch zu stellen**.

Um die Wahrscheinlichkeit der Teilnahme an der Evaluation nach den Science-Watch-Partys zu erhöhen, wurden die Feedbackkarten und Stifte bereits zu Beginn auf den Stühlen ausgelegt. **Doch Tabea war bewusst, dass die Ansichten und Verhaltensabsichten der Teilnehmenden schon einige Wochen später anders ausfallen könnten als direkt nach der Veranstaltung**. Deshalb wollte Tabea die Aussagekraft der Ergebnisse nicht überschätzen. Sie überlegte, welche Quelle weitere Hinweise auf die Verhaltensabsichten der Studierenden liefern könnte. Schließlich entschied sie sich, regelmäßig bei den Studienberatungen der Hochschulen anzufragen, ob sie einen Unterschied im Interesse an kleinen Fächern bemerkt hatten. Natürlich war ihr aber klar, dass mögliche Unterschiede zwar auf einen Wandel hinweisen könnten, aber nicht unbedingt direkte Effekte des Projektes wären.



# Excel-Sheets und Anekdoten

Ein Großteil der Daten, z. B. die Monitoring-Daten und die Ergebnisse der Feedbackkarten, war quantitativ auswertbar. Hierfür reichte **Excel zur Erfassung und Aufbereitung** für die regelmäßigen Berichte an das Projektteam sowie die fördernde Stiftung bereits aus. Die qualitativen Daten zu den Veranstaltungen, die sich aus Bewertung und Wahrnehmung der Gäste und Organisator\*innen ergaben, sammelte Tabea als Anekdoten und ergänzte sie in ihren Berichten neben den Zahlen. Die Anekdoten zog sie zur Einordnung der quantitativen Daten heran, um ihre Interpretation der Ergebnisse zu unterstützen.



# Uneinheitliche Ergebnisse

Die Rückmeldungen der Presse wie auch vonseiten der Veranstalter\*innen und Wissenschaftler\*innen waren positiv. Letztere berichteten, dass sie ein großes Interesse des Publikums an den Themen der Science-Watch-Partys wahrnahmen. Die quantitativen Daten dagegen waren weniger vielversprechend. Auf Social Media stagnierten die Kennwerte auf niedrigem Niveau. Die Studienberatungen berichteten von keinem bemerkenswerten Interesse an den Fächern. An den Ergebnissen der Feedbackkarten war abzulesen, dass die Bekanntheit der Fachbereiche unter den Besucher\*innen durch die Veranstaltungen gestiegen war, die Nützlichkeit dieses Wissens allerdings unklar blieb. Auch für die Studienwahl schienen sich keine Konsequenzen zu ergeben.

Insgesamt ließen sich also aus Tabeas Sicht die Ziele, die aus der Theory of Change abgeleitet wurden, bisher nicht als erfüllt ansehen. **Die Social-Media-Kampagne lieferte zwar die geplanten Outputs, erreichte aber nur wenige Menschen. Aus den bisher erhobenen Daten ließ sich nicht darauf schließen, ob die Veranstaltungen die Fächer wirklich attraktiver machen konnten. Tabea war allerdings nicht sicher, ob es an ihrer Evaluationsmethode lag oder an dem Programm selbst.** Sie stellte fest, dass sie nähere Informationen von der Zielgruppe brauchte, um zu verstehen, wie die Projektaktivitäten zielgerichteter gestaltet werden könnten.





# Das weit entfernte Ziel

Evaluationsphase 2

Weiter



# Eine neue Perspektive auf die Evaluationsfragen

In der zweiten Evaluationsphase interessierte sich Tabea weiterhin für die gleiche Frage: Wie könnte das Projekt zur Stärkung kleiner Fächer beitragen? Die Projektziele leiteten sich wieder aus den **Annahmen der Theory of Change\*** ab:

- *Die Social-Media-Kampagne informiert Studierende und Studieninteressierte über kleine Fächer an den beteiligten Hochschulen sowie die Inhalte dieser Fächer.*
- *Die Science-Watch-Partys zeigen dem Publikum aus Studierenden und Studieninteressierten, wie das Wissen der kleinen Fächer auf aktuelle gesellschaftliche Fragen und populäre Themen (dargestellt in Hollywood-Filmen) angewandt werden kann. So steigern sie das Interesse an den Fächern und die Motivation der Zielgruppe, diese Fächer zu studieren.*

**In dieser Evaluationsphase war es ihr aber wichtig, einen besseren Eindruck davon zu bekommen, wie die Zielgruppe die Projektaktivitäten wahrnahm.** Denn sie fand die Einschätzung der Projektbeteiligten zur Beurteilung der Zielerfüllung nicht ausreichend.

\* siehe „So kam es zur Einigung auf Projektziele“, Evaluationsphase 1 (S. 53-54)



# Ergänzungen im formativ-summativen Design

Tabea hielt an der **Kombination aus formativer und summativer Evaluation** fest. Ihr laufendes Monitoring der Daten zu Veranstaltungen, aus Social Media und Medienberichten behielt sie bei. Darüber hinaus entschied Tabea, mit Studierenden zu sprechen, um Feedback zum Social-Media-Content des Projekts zu bekommen. **Daraus wollte sie Informationen gewinnen, um die Kampagne noch im Projektverlauf verbessern zu können. Die Abfrage bei den Besucher\*innen nach den Veranstaltungen im Rahmen der summativen Evaluation sollte ebenfalls beibehalten werden, um eine Zwischenbilanz bezüglich der Zielerfüllung zu ziehen.** Hierbei sollte es aber methodische Änderungen geben.



## Tiefere Einblicke

Eine Neuerung, welche die zweite Evaluationsphase brachte: Tabea wollte mit Studierenden über den Social-Media-Auftritt des Projekts sprechen. **Konkret interessierte sie, wie sich die geringe Reichweite erklären ließe und wie man die Beiträge attraktiver für die Zielgruppe gestalten könnte.**

Bei der summativen Evaluation der Science-Watch-Partys über die Feedbackkarten hatte Tabea allgemeine Tendenzen der Besucher\*innen identifiziert, etwa als wie wichtig sie die Forschung in den kleinen Fächern einschätzten. Diesmal wollte Tabea im direkten Gespräch tiefergehende Fragen stellen und mehr über das **Vorwissen** der Teilnehmenden, über ihre möglichen **Berührungspunkte** mit und ihre (Vor-) **Einstellungen** gegenüber den behandelten Fächern erfahren.



# Gruppen- und Schnappschuss-Interviews

Tabea bereitete zwei Arten von Interviews vor. Für die Evaluation der Social-Media-Präsenz plante sie ein etwa 30 Minuten langes **Gruppeninterview** mit drei Freiwilligen, die nach einer der Science-Watch-Partys ihre Hilfe anboten.\* Um mehr über die Wahrnehmung der jeweiligen Studienfächer durch die Teilnehmenden zu lernen, entschied sie, die Feedbackkarten nach den Science-Watch-Partys durch kurze **Schnappschuss-Interviews** mit maximal vier Fragen zu ersetzen. (Für Hinweise zur Vorbereitung von Interview-Leitfäden: [How-To 4: Erhebungsinstrumente entwickeln.](#))

\* Auf die Auswahl der Freiwilligen wird im weiteren Verlauf eingegangen („Aus folgenden Quellen wurden die Daten erhoben“, Evaluationsphase 2, S. 69).



# Interview-Zeitpunkte für die optimale Verwertung

Der bestehende Monitoring-Plan sah weiterhin regelmäßige Erhebungen der laufenden Daten wie z. B. der Social-Media-Aktivitäten vor. Das **Gruppeninterview zur Social-Media-Kampagne sollte einmalig** während des Projektes durchgeführt werden. Es wurde für den **Beginn der Evaluationsphase eingeplant**, sodass daraus Lehren für die weitere Kampagne gezogen werden konnten. Die **Schnappschuss-Interviews wollte Tabea jeweils direkt im Anschluss an Veranstaltungen durchführen**. Deren Ergebnisse sollten in den Evaluationsbericht für das laufende Jahr einfließen.



# Freiwillige Unterstützung und zufällige Auswahl

**Die Idee für das Gruppeninterview war bei einer Science-Watch-Party entstanden**, als Tabea nach der Veranstaltung mit drei Besucher\*innen ins Gespräch kam. Als Tabea die Social-Media-Kanäle erwähnte und ihnen Beiträge zeigte, hatten die Studierenden viele, auch kritische Gedanken zu den Posts. Daher fragte Tabea spontan, ob sie bereit wären, ihr Feedback zu geben und dadurch dem Projekt zu helfen, eine für die Zielgruppe ansprechende Social-Media-Präsenz zu entwickeln. **Die Studierenden freuten sich, nach ihrer Expertise gefragt zu werden, und sagten zu.**

Für die Schnappschuss-Interviews bei den Veranstaltungen konnte Tabea ihre Interviewpartner\*innen nicht im Voraus wählen. Sie wollte eine **zufällige Auswahl** treffen. Hierfür legte sie **Umschläge mit Loszetteln** aus. Bei jeder Veranstaltung enthielten fünf der Umschläge Lose mit Gewinnen, die am Ende bei Tabea abgeholt werden konnten. Nach der Überreichung fragte Tabea die Gewinner\*innen, ob sie noch Zeit für ein paar kurze Fragen hätten.



# Software-Unterstützung

Das Gruppeninterview zeichnete Tabea auf und **verschriftlichte es mithilfe einer Transkriptionssoftware. Für die inhaltliche Auswertung nutzte sie ein Programm zur qualitativen Datenanalyse.** Solche Programme helfen bei der Datenauswertung und der Strukturierung der Ergebnisse, indem die Nutzer\*innen bestimmte Passagen oder Absätze markieren und bestimmten Themen (sogenannten Kategorien und Codes) zuordnen (hierzu mehr im [How-To 5: Daten auswerten](#)). Für ihr Auswertungsraster entwickelte sie ein Kodebuch (wie das funktioniert, erklärt das [How-To-Zusatzmaterial: Kodebuch-Vorlage](#)).

Für die Schnappschuss-Interviews nach den Veranstaltungen reichte ihr eine direkte, stichwortartige Mitschrift der Antworten. Die Inhalte verglich sie miteinander, um vorherrschende Meinungen zu identifizieren und um ihre Interpretation der Ergebnisse aus den vergangenen Feedbackkarten kritisch zu prüfen.



## In kleinen Schritten zum Ziel

Aus dem Gruppeninterview ergab sich, dass die Posts für die Zielgruppe zwar inhaltlich interessant waren, ihre Gestaltung aber nicht zum Teilen oder Liken motivierte. Die Interviewpartner\*innen erklärten Tabea, welche Bilder und Arten von Posts bei ihnen gut ankommen und mit welcher Art von Beiträgen sie gerne und viel interagieren. **Die Ergebnisse stellte Tabea den für die Kommunikation verantwortlichen Projektmitgliedern vor** und beriet sie für ihr weiteres Vorgehen. Tatsächlich konnte sie im weiteren Projektverlauf erkennen, dass sich die Reichweite etwas gesteigert hatte. Aus den Ergebnissen der Schnappschuss-Interviews schloss Tabea, dass die Veranstaltungen die Bekanntheit der Fächer erhöhten und Studierende erreichten, die sich bisher wenig dafür interessiert hatten. Viele konnten sich trotzdem nicht vorstellen, diese Fächer zu studieren; unter anderem, weil sie ihre Begabungen in anderen Bereichen sahen. Sie fanden die Forschungsfelder dennoch spannender als zuvor und fühlten sich motiviert, mehr über die Themen zu diskutieren.

Insgesamt erkannte Tabea Fortschritte, was die Zielerfüllung anging. **Verständlicherweise wirkten sich die Aktivitäten nicht unmittelbar auf die Studierendenzahlen aus, denn diese hingen von vielen weiteren Faktoren ab. Doch das Projekt nahm die Rolle im angestrebten Wandel ein, die sich das Team in der Theory of Change vorgenommen hatte:** Aufklärung und ein erster Kontakt zu Fachbereichen, die sonst unterrepräsentiert sind.



# Absprachen zur Ergebniskommunikation

Zum Ende der Projektlaufzeit schrieb Tabea einen Evaluationsbericht. **Die fördernde Stiftung und das Projektteam hatten ihr zu Beginn des Projekts zugesagt, dass sie die Evaluationserkenntnisse unter bestimmten Auflagen für Veröffentlichungen nutzen könne** – unabhängig davon, wie die Ergebnisse der Evaluation ausfallen würden. Die frühe Absprache über den Umgang mit den Ergebnissen war Tabea bei der ersten Entwicklung der Evaluation wichtig gewesen, um spätere Konflikte zu vermeiden (→ [Worauf kommen es bei einer guten Evaluation an?](#)).

Die **Kurzform des Berichtes wurde auf der Projektwebseite hochgeladen** und Tabea stellte ausgewählte Ergebnisse, die generell auf Hochschulkommunikation und Social-Media-Kampagnen übertragen werden konnten, auf **Konferenzen** vor.





# Die unsichtbare Evaluation



# Prüfung der Investition

Die interaktiven, spielerischen Exponate, die Rahel und ihr Team in Auftrag gaben, waren **kostenintensiv**. Umso mehr interessierte Rahel, ob diese aktivierenden Ausstellungsstücke die Vermittlung der zentralen Botschaften der Ausstellung wirklich **unterstützen oder ob eine spielerische Vermittlung von den eigentlichen Inhalten ablenkt**. Diese Erkenntnisse spielen nicht nur für diese Ausstellung eine Rolle, sondern auch für zukünftige Ausstellungen und Exponate.



# Rückbezug auf den Projektantrag

Die Klärung der Projektziele fiel dem Ausstellungsteam leicht, schließlich hatte es gemeinsam mit der Ausstellungsleitung den **Förderantrag** geschrieben und in diesem Prozess bereits klare Ziele festgelegt. Die folgende Tabelle veranschaulicht lediglich die Lernziele des Projekts, auf die sich die Evaluation konzentrierte:

<b>Output</b> (erhoffte Leistungen des Spiels)	<b>Outcome</b> (erhoffte Wirkung bei der Zielgruppe)	<b>Impact</b> (erhoffte gesellschaftliche Wirkung)
<ul style="list-style-type: none"> <li>• Die Ausstellung zeigt auf interaktive, anschauliche Weise, <i>wie Roboter lernen</i></li> <li>• Die Ausstellung zeigt die aktuellen und geplanten <i>Anwendungsfelder der Robotik</i>, ihre <i>Rahmenbedingungen</i> und <i>Regulierungsmaßnahmen</i></li> </ul>	<ul style="list-style-type: none"> <li>• Besucher*innen lernen, <i>was Roboter in Zukunft lernen könnten</i> (und was nicht), <i>wie sie eingesetzt werden</i> (und wie nicht)</li> </ul>	<ul style="list-style-type: none"> <li>• Die Ausstellung <i>baut Verunsicherung bezüglich des Einsatzes von Robotik in der Industrie ab</i>, die in der Bevölkerung besteht, indem sie klare Einsatzgebiete und ihre Regulierung aufzeigt. Indem zentrale Informationen vermittelt werden, ermöglicht die Ausstellung einen <i>informierten, gesellschaftlichen Diskurs</i> über Anwendungsfelder und Grenzen der Robotik in der Arbeitswelt.</li> </ul>



# Lerneffekte durch die Ausstellung

Die Evaluation sollte Aufschluss darüber geben, ob die Exponate bei der **Vermittlung der gesetzten Lernziele** helfen oder nicht. Auf die Evaluation weiterer Ziele für jüngere Zielgruppen (u. a. Begeisterung für das Thema entwickeln) wollte Rahel an dieser Stelle verzichten, um die Untersuchung der Lernziele sorgfältig durchführen zu können.

- *Verstehen Besucher\*innen die Inhalte der Ausstellung (Programmierung, Anwendungsfelder, Rahmenbedingungen und Regulierung der Robotik) nach der Auseinandersetzung mit den Exponaten besser als zuvor?*
- *Welche Erkenntnisse bleiben den Besucher\*innen nach der Ausstellung besonders im Gedächtnis?*

Mit der zweiten Frage wollte das Team nicht nur herausfinden, welche Inhalte besonders eindrücklich waren, sondern auch einen Überblick bekommen, welche Arten von Exponaten zu welchen Erkenntnissen führen.



# Summative Evaluation oder Impact Evaluation?

Rahel interessierte sich in der Evaluation für die Wirkung der Ausstellung in Bezug auf die Kenntnisse der Besucher\*innen über Robotik. **Damit stehen die Ergebnisse und Folgen des Projekts im Vordergrund, was eine summative Evaluation nahelegt.**

Manche Evaluator\*innen würden in solchen Fällen auch von einer Impact Evaluation sprechen. **Impact Evaluationen sind allerdings ausschließlich an Veränderungen interessiert, die bei den Beteiligten auftreten.** Eine summative Evaluation kann hingegen zu den Projektergebnissen auch Resultate wie die Reichweite oder die Bewertung einer Aktivität durch das Publikum einschließen. Rahel beschreibt ihre Evaluation als summativ, da dieser Begriff übergreifender ist und neben der allgemeinen Wissenssteigerung auch ihr Interesse an besonderen Erkenntnissen zum Ausstellungsende berücksichtigt.



# Fachwissen und neue Erkenntnisse

Mit der Evaluation sollte der Wissenserwerb durch die Ausstellung geprüft werden. Die Ausstellung enthielt konkrete Themenfelder, zu denen sich Besucher\*innen Wissen aneignen konnten. Darüber hinaus interessierte das Team auch, welche wohl die einprägsamsten Erkenntnisse aus der Ausstellung sein könnten und welche Exponate solche bieten würden.

Im ersten Schritt diskutierte das Team, **wie sich der „erfolgreiche Wissenserwerb“ zeigen würde**. Am wichtigsten war ihnen, dass Besucher\*innen danach falsche Aussagen zu Robotik, die ihnen im Alltag begegnen könnten, von richtigen unterscheiden konnten. Dazu mussten sie die Botschaften der Ausstellung auf konkrete Fakten herunterbrechen und einigten sich auf maximal drei wichtige Fakten pro Themenbereich. Um die einprägsamen Erkenntnisse zu erfassen, mussten sie zudem eine sehr konkrete Frage formulieren, die den Kern ihres Evaluationsinteresses trifft. Denn unter **„Erkenntnissen“ könnte nicht nur Wissen zu Robotik verstanden werden, sondern auch andere Erfahrungen während des Museumsbesuchs**, die sie an dieser Stelle weniger interessierten.



# Ein spielerisches Quiz und eine Pinnwand

Das Ausstellungskonzept sah viele interaktive Elemente vor. Rahel wollte die Evaluation daher mit in die Ausstellung integrieren. Wichtig war ihr dabei, dass **alle Daten im Sinne des Datenschutzes anonym gesammelt** werden und dass die **Evaluation stilistisch gut zur restlichen Ausstellung passt, sodass sie als solche kaum auffällt**. Eine kreative [Alternative zum klassischen Fragebogen](#) erschien Rahel die richtige Wahl.

Um zu prüfen, ob die Besucher\*innen nach der Ausstellung besser darin sind, falsche Aussagen von Fakten über Robotik zu unterscheiden, bot sich damit **ein Quiz mit verschiedenen Antwortmöglichkeiten** an. Dafür entwarf das Team eine Art **Murmelbahn**, bei der man die Antwort durch das Einwerfen eines Ping Pong Balls in die entsprechende Bahn auswählt. Um zum Ende der Ausstellung eindruckliche Erkenntnisse abzufragen, schlug Rahel eine **Pinnwand vor, auf der diese gesammelt würden und die sich mit genug Antworten zu einem durch Besucher\*innen kreierte Exponat entwickeln würde**.



# Vorher-Nachher-Erhebung und Abschlussfrage

Um mögliche **Wirkungen der Ausstellung** zu verstehen, wollte Rahel einen **Vorher-Nachher-Vergleich** mit den Besucher\*innen anstellen. Die gleichen Quizfragen sollten also zweimal gestellt werden, beim Eintritt in die Ausstellung und kurz vor Ende des Besuches, wenn alle Exponate durchlaufen sind. Die Auflösung des Quiz würde erst nach der zweiten Abfrage erfolgen.

Die **Erhebung der Erkenntnisse** würde dagegen nur **einmalig zum Ende** des Besuchs erfolgen.



# Eine befristete Vollerhebung

Für Rahel war klar, dass sie die Besucher\*innen zum Quiz und zur Angabe ihrer Erkenntnisse motivieren musste. Sie strebte für die Eröffnungswoche der Ausstellung eine **Vollerhebung** an, es sollten also alle Besucher\*innen teilnehmen. Dafür sollten die Besucher\*innen bereits beim Museumseintritt mit ihrem Ticket für die Ausstellung die Ping Pong Bälle bekommen, mit denen im Quiz abgestimmt wird. Zusätzlich würden Betreuer\*innen bereitstehen, um zur Teilnahme am Quiz zu motivieren und ungenutzte Bälle in der Ausstellung einzusammeln.

**Es war bereits beschlossen, dass nach der ersten Ausstellungswoche das Betreuungspersonal reduziert wird.** Zudem war die Ausgabe der Bälle an den Ticketschaltern dem Museumspersonal **über die Eröffnungswoche hinaus zu aufwendig.** Daher wusste Rahel, dass die Quizteilnahme nach der ersten Woche davon abhing, dass die Besucher\*innen selbst auf das Spiel und die Pinnwand aufmerksam werden. Sie hoffte jedoch, dass die **Spielmechanismen, das einladende Design und die bereits sehr interaktive Ausstellung** viele Menschen zur Teilnahme bewegen würden.



# Trefferquoten und Antwortsortierung

Jedes Mal, wenn die Behälter mit den Ping Pong Bällen geleert wurden, wurden die Stimmen für die jeweilige Antwortoption dokumentiert. So konnten am Ende **prozentuale Anteile der verschiedenen Antwortoptionen** berechnet werden. Entscheidend für die Evaluationsfrage war dabei natürlich der Anteil der korrekten Antwort im Vorher-Nachher-Vergleich.

Die Angaben an der Pinnwand wurden digitalisiert und zum Ende der Ausstellung betrachtet. **Rahel ordnete die Antworten gemäß den Inhalten den unterschiedlichen Exponaten zu.** So konnte sie erkennen, welche dieser Exponate die meisten Erkenntnisse lieferten, und in ihrem Team diskutieren, was sie daraus über die Gestaltung der Exponate lernen.



# Ergebnisse mit Einschränkungen

Zum Ende der Ausstellung begann Rahel ihre Auswertung und stellte die Ergebnisse vor. Zuvor hatte sie **mit den Betreuer\*innen über deren Beobachtungen und Eindrücke während der Evaluation gesprochen**. Der Trefferquote im Quiz zufolge steigerte sich das Wissen durch die Ausstellung nur geringfügig. Die Betreuer\*innen meldeten hierzu jedoch zurück, dass Erwachsene oft ihre Kinder die Bälle zufällig einwerfen ließen oder sich für die Antwortbehälter mit weniger Bällen entschieden. **Das Quiz wurde also mehr als Spielerei betrachtet und nicht von allen ernsthaft beantwortet. Auch die Pinnwand wurde vielfach für Quatschantworten genutzt**. Etwa ein Drittel der Antworten boten Rahel allerdings hilfreiche Hinweise, welche Exponate besonders im Kopf blieben.

Rahel war etwas enttäuscht, da sie nicht so viel über die Wirkung der Ausstellung gelernt hatte wie erhofft. **Aber das Team konnte aus den Ergebnissen viel über Evaluationen an sich lernen**: Ihre Evaluation konnte viele Menschen zur Teilnahme motivieren, war aber in ihrem Kontext schlecht für eine Wissensabfrage geeignet. Von dem niedrigschwelligen Ansatz waren sie überzeugt, doch für die nächste Ausstellung wollten sie kritisch hinterfragen, **ob die Evaluationsfragen sich dafür eignen oder eine andere Methode erfordern**.



# Eine Zusammenfassung für LinkedIn

Rahel war trotz der uneindeutigen Ergebnisse zufrieden damit, wie bereitwillig die Evaluation von Besucher\*innen aufgenommen und wie das Quiz in die Ausstellung integriert wurde. **Sie verfasste einen kurzen Rückblick** auf die Entwicklung der Murmelbahn und der Pinnwand sowie die gesammelten Erkenntnisse über den Einsatz von spielerischen Methoden zur Wissensabfrage. **Ihren Beitrag teilte sie auf LinkedIn** in einer Gruppe zum Thema Ausstellungsentwicklung. Sie hoffte, auch von den Erfahrungen der anderen in diesem Bereich zu profitieren.





# Die kritische Prüfung



# Misserfolge nachvollziehen

Lange Zeit waren die Pub-Science-Events das Flaggschiff-Kommunikationsprojekt des Clusters, mit stabilen Publikumszahlen von rund 60 Personen. **Doch in letzter Zeit waren die Besuchszahlen gesunken.** Mark wollte **die Gründe hierfür verstehen** und lernen, wie das Projekt wieder attraktiver für die Zielgruppe werden kann. Er sah in dieser erstmaligen Evaluation auch die Gelegenheit, nach einigen Jahren der erfolgreichen Durchführung mit dem restlichen Kommunikationsteam und der Forschungsleitung die Rolle des Projekts zu **rekapitulieren**: Inwieweit adressiert es die Ziele aus der heutigen Kommunikationsstrategie des Clusters?



# Unterschiedliche Zielvorstellungen

Mark und das Kommunikationsteam setzten sich mit der leitenden Wissenschaftlerin zusammen, um die Kommunikationsziele des Clusters zu besprechen und die Pub-Science-Reihe in dieser Strategie zu verorten. Dabei fiel auf, **dass dem Format in der Gruppe verschiedene Ziele zugeschrieben wurden**. Dem Kommunikationsteam war inzwischen der Unterhaltungsaspekt wichtiger, während die leitende Wissenschaftlerin immer noch die ursprünglichen Projektziele am höchsten priorisierte: die Vermittlung von Studienergebnissen und daraus folgenden Handlungsempfehlungen. Mark hatte **Zweifel, ob ein Kneipenabend all das würde leisten können**. Er notierte die gesammelten Ziele und nahm sich vor, diese in der Evaluation kritisch zu prüfen.



Output <i>(erhoffte Leistungen des Spiels)</i>	Outcome <i>(erhoffte Wirkung bei der Zielgruppe)</i>	Impact <i>(erhoffte gesellschaftliche Wirkung)</i>
<ul style="list-style-type: none"> <li>Das Publikum kennt Studien des Clusters sowie <i>interessante Fälle der Medizingeschichte</i></li> <li>Das Publikum bekommt <i>unterhaltsame Einblicke in den Arbeitsalltag</i> von Wissenschaftler*innen</li> </ul>	<ul style="list-style-type: none"> <li>Das Publikum ist über Studienergebnisse <i>informiert</i></li> <li>Das Publikum <i>versteh</i>t, wie sich <i>Krankheitsbilder</i> zwischen Menschen unterscheiden können</li> <li>Das Publikum versteht, wozu es <i>Gendermedizin</i> braucht</li> <li>Das Publikum fühlt sich <i>gut unterhalten</i></li> <li>Das Publikum hat eine <i>Vorstellung vom wissenschaftlichen Arbeitsalltag</i> und der Forschung des Clusters</li> </ul>	<ul style="list-style-type: none"> <li>die Öffentlichkeit ist für aktuelle <i>Wissenslücken</i> in der Medizin <i>sensibilisiert</i> und versteht die <i>Bedeutung geschlechtersensibler</i> Medizinforschung</li> </ul>

# Mehrwerte auf dem Prüfstand

Das Kommunikationsteam und die Forschungsleitung einigten sich darauf, **mit der Evaluation zu prüfen**, was das Projekt dem Publikum aktuell bietet, **welche der gesetzten Ziele erfüllt werden und welche nicht erfüllt werden:**

- *Welche Mehrwerte bietet die Pub-Science-Reihe für das Publikum?*
- *Wie bewertet das Publikum das Projekt?*
- *Schafft das Projekt die Voraussetzungen, um die Informationsziele zu erreichen?*
- *Schafft das Projekt die Voraussetzungen, um die Unterhaltungsziele zu erreichen?*



# Summative Evaluation

Mark interessierte sich dafür, welche „Leistungen“ (sogenannte Outputs) das Projekt dem Publikum bietet. Deshalb entschied er sich für eine **summative Evaluation**, die traditionell die Ergebnisse eines Projekts in den Fokus rückt und **nach der Veranstaltung** beginnt.



# Publikumsbeurteilung

Was die Zuschauer\*innen aus diesen Abenden mitnehmen, können sie selbst am besten einschätzen. Deshalb war Mark vor allem an den Meinungen des Publikums interessiert, an ihrer **Bewertung** der Veranstaltung und daran, ob sie sich (gut oder schlecht) **informiert und unterhalten fühlten**.



*Hat das Event Spaß gemacht?*

*War es informativ?*

*Erfüllte es die Erwartungen  
der Zuschauer\*innen?*

# Quantitativer Fragebogen

Mark wollte einen möglichst **umfassenden Eindruck der Publikumsmeinung** bekommen. **Er hatte Annahmen** darüber, welche Mehrwerte das Projekt dem Publikum bieten könnte (z. B. Unterhaltung). Mit der Evaluation wollte er herausfinden, inwieweit das Publikum diese Mehrwerte tatsächlich wahrnahm. Mark kam zu dem Schluss, dass eine **quantitative Befragung** Antworten liefern könnte, sofern ein Großteil des Publikums an der Evaluation teilnahm. Basierend auf den Erfahrungen der letzten Termine hoffte Mark auf etwa 40 Zuschauer\*innen bei der nächsten Veranstaltung. Er wünschte sich eine **minimale Fallzahl von 30 Personen**, um zumindest einen ersten Eindruck unterschiedlicher Meinungen zu bekommen. Ein **standardisierter Fragebogen**, der **schnell ausgefüllt** ist und **anonymes Feedback** ermöglicht, schien ihm die passende Lösung zu sein.



# Erhebung im Anschluss an die Maßnahme

Der Fragebogen sollte **einmalig, direkt nach der Veranstaltung** zum Einsatz kommen. Auch wenn es sich anbieten würde, zu Beginn bereits Erwartungen und aktuelle Wissensstände oder Einstellungen des Publikums abzufragen, verwarf Mark diese Idee aus pragmatischen Gründen wieder. Angesichts seiner begrenzten Ressourcen für die Evaluation erschien es ihm sinnvoller, **seine Energie in eine einmalige Erhebung zu investieren** und gute Bedingungen für eine hohe Teilnahmequote zu schaffen.



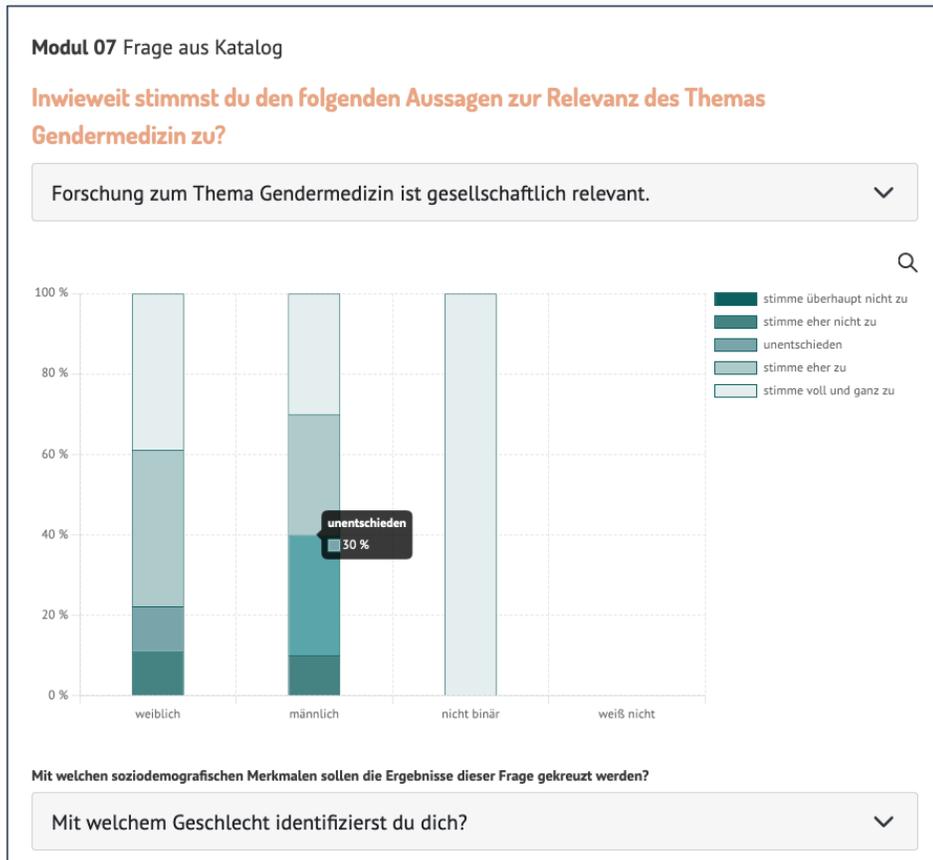
# Voraussetzungen für hohe Teilnahmequoten

Mark machte sich viele Gedanken darüber, wie er die Zuschauer\*innen dazu motivieren könnte, **nach der Veranstaltung noch einen Fragebogen auszufüllen**. Meistens gingen die Kneipenbesucher\*innen schnell nach Hause, sobald die Veranstaltung vorbei war. Glücklicherweise konnte er auf Forschungsergebnisse und Erfahrungswerte anderer Praktiker\*innen zurückgreifen, um die **Bereitschaft zur Evaluationsteilnahme** zu erhöhen (→ [Wegweiser: Teilnahmequoten erhöhen](#)).

Zum einen entschied er sich, den **Fragebogen sowohl vor Ort auf Papier als auch online** (über QR-Codes zum Mitnehmen) anzubieten. So würden Personen die Fragen beantworten können, während sie auf ihre Rechnung warten, aber auch noch später teilnehmen können, wenn sie es eilig haben. Zum anderen achtete er besonders darauf, **nur die wichtigsten Fragen zu stellen** – auf Daten, die zur Prüfung der Ziele nicht entscheidend waren, verzichtete er: z. B. das Alter der Befragten, bei dem er keine großen Unterschiede erwartete und das nicht ausschlaggebend für die Zielgruppendefinition war. So blieb der Fragebogen kurz. Nicht zuletzt **wies das Organisationsteam bereits während der Veranstaltung auf die Evaluation hin** und versprach **Süßigkeiten** für alle Teilnehmenden, die den Fragebogen vor Ort ausfüllen.



# Deskriptive Statistiken



Für die Fragebogenerstellung und -auswertung nutzte Mark → die [Evaluationsplattform der Impact Unit](#), die speziell für standardisierte Befragungen in der Wissenschaftskommunikation entwickelt wurde. Hier ließ er sich die **Antwortverteilungen** ausgeben, betrachtete die **Mittelwerte** und **Gruppenunterschiede** in den Angaben. Mark hatte die Vermutung, dass die Zuschauer\*innen abhängig von ihrem eigenen Geschlecht die Mehrwerte einer Veranstaltung zu Gendermedizin unterschiedlich wahrnehmen. Trotzdem interpretierte er diese Diagramme mit Vorsicht, denn für belastbare Ergebnisse waren die Fallzahlen pro Gruppe mit teilweise weniger als zehn Personen zu gering.



# Die Ermutigung zum Kurswechsel

Die **Antworten aus dem Publikum bestätigten Marks Vermutung**: Das letzte Science-Pub-Event wurde zwar als unterhaltsam bewertet und gab Einblicke in wissenschaftliche Prozesse, doch konkrete Informationen zu Studieninhalten und Krankheitsbildern nahmen die wenigsten mit. **In der weiteren Diskussion der Ergebnisse entwickelte das Team Theorien, woran das liegen könnte**: Nachdem die zentralen Studien und Forschungsprojekte im Laufe des Projekts vorgestellt waren, beleuchteten die Forschenden immer häufiger andere Themen, um Abwechslung zu bieten. Sie gaben den Forschungsmethoden mehr Raum als den Studienergebnissen, teilten persönliche Geschichten über ihren Berufsalltag und ihren Weg in die Wissenschaft. Mit der Zeit bot das Projekt nicht mehr das Programm, für das es bekannt geworden war und beworben wurde. Das könnte Erwartungen früherer Zuschauer\*innen enttäuscht haben.

Dass sich ihr Format mit der Zeit neu ausgerichtet hatte, wertete das Team nicht zwangsläufig als schlecht. **Die Evaluation verdeutlichte ihnen jedoch die Folgen der Veränderungen und stellte sie vor eine Entscheidung, die sie gemeinsam treffen wollten**: Entweder die Projektstrategie und Kommunikation müssten sich verändern oder das Team müsste sich auf die ursprüngliche Projektidee zurückbesinnen.



# Eine Stammtisch-Diskussion

In Gesprächen mit Kolleg\*innen an der Universität wurde Mark bewusst, dass viele Kommunikationsteams mit ähnlichen Herausforderungen konfrontiert sind und von den Erfahrungen seines Teams profitieren könnten. Auch wenn es ihm nicht direkt leicht fiel, entschied er sich, von den schlechter besuchten Veranstaltungen und den hierzu evaluierten Hintergründen auch außerhalb des Projekts zu erzählen.

**Er stellte die Geschichte der Pub-Science-Reihe bei einem Stammtisch für Kommunikator\*innen an seiner Universität vor.**

Ermutigt durch Marks Offenheit berichteten dort auch andere Kolleg\*innen von gescheiterten Projekten und ihren Lernerfahrungen.

