

12 Mehr sehen

Triangulation als Forschungsstrategie in der Ausstellungsevaluation

Claudia Gorr

Der vorliegende Band versammelt »Innovative Wege der Konzeption und Evaluation von Ausstellungen«. Manchmal ist es innovativ, sich auf etwas zu besinnen, das schon länger existiert, aber neue Relevanz gewinnen kann. Das gilt aus meiner Sicht für die Strategie der Triangulation.

In der Evaluation von Ausstellungen begegnen Evaluator*innen oft typischen An- und Herausforderungen. Z.B. geht es häufig weniger um Grundlagenforschung als mehr darum, die Wirkung eines speziellen Führungsangebots, eines Exponats oder einer bestimmten Ausstellung zu prüfen. Eine komplexe Umgebung wie eine Ausstellung, in der es im täglichen Verkehr kaum möglich ist, einzelne Einflüsse zu kontrollieren, macht hier allerdings ausschließlich standardisierte Verfahren schwierig. Zudem ist das Ziel oftmals, herauszufinden, welche Gründe oder Prozesse hinter einem Erfolg oder Misserfolg stecken. Rein quantitative Herangehensweisen gelangen hier schnell an ihre Grenzen, und so liegt es nahe, auch qualitative Methoden einzusetzen. Bei qualitativen Studien bestehen jedoch typischerweise Probleme hinsichtlich der Wiederhol- und Übertragbarkeit von Ergebnissen, was es im Zweifelsfall erschwert, Entscheidungsträger*innen von der Tragfähigkeit der Evaluationsergebnisse zu überzeugen oder die Ergebnisse in einschlägigen wissenschaftlichen Journalen zu veröffentlichen. Eine Debatte über angemessene und verbindliche Qualitätskriterien für die qualitative Forschung wird zudem seit Jahrzehnten kontrovers geführt (vgl. Flick 2005; Döring/Bortz 2016).

Ein pragmatischer Ausweg ist eine sinnvolle Verbindung qualitativer und quantitativer Methoden. Laut Döring und Bortz lassen sich durchaus

»auch im Zuge einer qualitativen Studie und im Verständnis des Sozialkonstruktivismus [...] quantitative Datenerhebungen einbeziehen – typischerweise, indem deskriptiv-statistische Ergebnisse zur Häufigkeit oder Ausprägung bestimmter Merkmale oder Phänomene in der qualitativen Gesamtinterpretation Berücksichtigung finden« (ebd.: 72).

Diese Nutzung unterschiedlicher Datentypen kann als spezielle Form der sogenannten Triangulation verstanden werden (vgl. Flick 2008). Und das Plädoyer dieses Beitrags lautet: Triangulation kann ein Schlüssel dazu sein, verschiedene typische Herausforderungen, vor die uns Evaluationen stellen, miteinander in Einklang zu bringen.

Die ursprünglich metaphorische Bezeichnung Triangulation wurde ab den 1950er Jahren aus der Geodäsie in die Sozialwissenschaften importiert. Um sich diesem sozialwissenschaftlichen Begriff der Triangulation zu nähern, lohnt ein Blick auf die Arbeit des Soziologen Uwe Flick. Er versteht darunter eine »Kombination unterschiedlicher Theorien, Methoden, Daten und/oder Forschender zur wechselseitigen Absicherung und Ergänzung im qualitativen Forschungsprozess« (Döring/Bortz 2016: 111, dort zit.n. Flick 2004). Es handelt sich in seinem Verständnis um eine Strategie zur wissenschaftlichen Gütesicherung, um höhere Plausibilität und Glaubwürdigkeit für seine Forschungsergebnisse zu gewährleisten. Der Soziologe Norman Kent Denzin entwickelte 1970 eine Typologisierung von Triangulation (vgl. Denzin 1970), wonach bis heute vier grundlegende Formen der Triangulation unterschieden werden: die Nutzung verschiedener Methoden bei der Erhebung und der Datenanalyse (methodologische Triangulation), die Einbinden verschiedener Datensätze/-quellen (Datentriangulation), die Einbindung verschiedener Personen im Forschungsprozess, um einen zu starken Bias eines einzelnen Forschenden zu vermeiden (Forschertriangulation), und schließlich eine Kombination verschiedener theoretischer Perspektiven (Theorientriangulation).

Evaluationsprojekt in der neuen experimenta in Heilbronn

Die experimenta in Heilbronn existiert bereits seit 2009, ursprünglich im ehemaligen Hagebuchener Speicher. 2013 wurde von der Dieter-Schwarz-Stiftung eine umfassende Erweiterung beschlossen und im März 2019 öffnete die neue experimenta ihre Türen. Die neue Ausstellung bewog die Exponatebaufirma

Hüttinger GmbH und die experimenta dazu, ihre Wirkung zu prüfen und Ausstellungsstockwerke unterschiedlichen Charakters in ihrer Wirkungsweise gegenüberzustellen.¹ Im Folgenden konzentriere ich mich auf die Ausstellungssektion »Forscherland«, in der Kinder zwischen vier und zehn Jahren mit klassischen naturwissenschaftlich-technischen Phänomenen spielen und experimentieren können.

Der Wirkungsaspekt, der uns im Zuge unseres Evaluationsprojekts besonders interessierte, war das Lernen. Unsere Gesamtforschungsfrage lautete: *Wie lernen Besucher*innen in unserer neuen Dauerausstellung?* Wie aber misst man Lernen? Verhaltensaspekte des Lernens lassen sich von außen beobachten, aber Erlebensaspekte und kognitive Veränderungen nur von Besucher*innen selbst beschreiben. Beide Herangehensweisen haben ihre Vorteile und Grenzen. Wenn wir beobachten, sehen wir möglicherweise Dinge, die den Handelnden selbst nicht bewusst werden oder die sie bei Selbstauskunft verschweigen. Auf der anderen Seite unterliegt jegliche Beobachtung einer Interpretation; wir werden nie vollständig sicher sein können, welche tieferen Beweggründe hinter einem äußerlich sichtbaren Verhalten stecken.

Wir entschieden deshalb, die Frage nach dem Lernen aufzuschlüsseln, sowohl in direkt beobachtbare als auch nicht beobachtbare Aspekte, was in verschiedene Unterfragen mündete. Diese ermöglichten uns im selben Zuge, die übergreifende Generalfrage mess- und damit beantwortbar zu machen. Konkret wollten wir erfahren, welche Stationen unsere Besucher*innen nutzen und wie lange sie sich jeweils dort aufhalten. Allgemeiner ging es uns darum, welche spezifischen Indikatoren für Lernen sich identifizieren lassen und wie stark diese an den Stationen ausgeprägt sind. Schließlich wollten wir verstehen, ob Besucher*innen die Hauptthemen der Stationen erfassen, ob sie an bekannte Dinge anknüpfen können und woran sie sich kurz nach dem Besuch erinnern.

Es drängte sich hier eine methodologische Triangulation auf, denn wir planten bewusst unterschiedliche Methoden ein, um gleichermaßen beobachtbaren und nicht beobachtbaren Aspekten gerecht zu werden. Quasi als Nebenwirkung ergab sich daraus eine sogenannte Datentriangulation.

1 Die Studie wurde weitgehend finanziert von der Firma Kurt Hüttinger GmbH & Co. KG. Durchgeführt wurde sie in Kooperation zwischen der experimenta und der Evaluatorin Dr. Elsa Bailey. Die Firma Hüttinger wollte u.a. Aufschluss darüber gewinnen, wie die klassischen Hands-on-Exponate in Relation zu eher digitalen Exponaten bei den Besucher*innen ankommen und Lernprozesse anstoßen.

Denn dadurch, dass wir unterschiedliche Stichproben in unterschiedlichen Situationen zu unterschiedlichen Zeitpunkten untersuchten, gewannen wir unterschiedliche Datensätze aus unterschiedlichen Datenquellen.

Im Ganzen führte das nun zu einer Aufteilung in vier Teilstudien (siehe Abb. 1), die ich im Folgenden näher beschreibe.

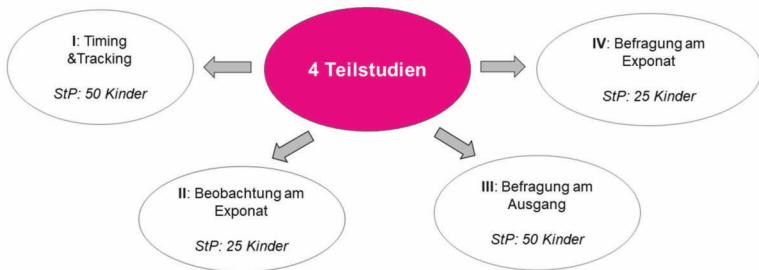


Abb. 1: Die vier Teilstudien des Evaluationsprojekts (StP = Stichprobe), eigene Darstellung

Teilstudie I: Beobachtung (Timing & Tracking)

Hierfür wurden Besucher*innen im »Forscherland« beobachtet und ihre Wege durch die Ausstellung dokumentiert. Die Stichprobe umfasste 50 Besucher*innen zwischen fünf und zehn Jahren. 27 Personen wurden auf ein Alter zwischen fünf und sieben Jahren geschätzt, 23 Personen auf ein Alter zwischen acht und zehn Jahren. 48 % wurden als weiblich, 52 % als männlich identifiziert. 90 % kamen mit der Familie, 10 % mit ihrer Klasse bzw. einer Gruppe.

Notiert wurden nicht nur Nutzungsreihenfolgen und Verweilzeiten einer Person an den Exponaten, sondern auch, ob Texte gelesen wurden. Hieraus ließ sich nun einerseits die »attraction power« der einzelnen Stationen bestimmen, also die Anzahl der Nutzer*innen, andererseits die Verweildauer an den jeweiligen Stationen unter Angabe der mittleren Dauer, der Minimal- und der Maximaldauer (vgl. Tab. 1).

Exponate	Verweildauer Mittel	Verweildauer Min	Verweildauer Max	Anzahl Nutzer*innen
Pustespiele	02:12	00:13	07:12	38
Wasserwerk	02:00	00:10	09:09	33
Wasserleitung	01:41	00:08	12:51	29
Luftpost	01:37	00:12	06:34	29
Kurbelwerk	00:29	00:05	01:34	29
Luftstoß	00:51	00:08	02:13	27
Windspiele	01:15	00:08	03:41	26
Luftfontäne	00:49	00:10	02:15	26
Großes Baumhaus	01:28	00:14	04:53	25
Luftbremse	02:41	00:19	09:17	24
Luftaufzug	02:00	00:18	05:50	24
Flaschenzug	00:22	00:06	00:52	22
Luftkringel	00:39	00:08	02:00	21
Magnetbilder	00:52	00:11	01:48	21
Monsterhebel	00:22	00:06	00:55	20
Transportkette	05:02	00:19	42:25	19
Stromprüfung	00:56	00:18	03:02	19
Nebelzeichen	00:24	00:06	01:20	19
Luftrakete	01:41	00:10	05:42	18
Rennscheiben	01:00	00:06	02:45	18
Kurvenfahrt	01:30	00:15	07:49	17
Schiffstreppe	02:00	00:13	07:24	17

Tab. 1: Übersicht über Verweildauern (»holding power«) und Zahl der Nutzer*innen (»attraction power«) an den evaluierten Exponaten (eigene Darstellung). Die Farbskala markiert Exponate, die hinsichtlich »holding power« und »attraction power« über dem Durchschnitt bzw. Mittelwert (grün), eher im Bereich des Durchschnitts (gelb) oder unter dem Durchschnitt (rot) liegen.

Teilstudie II: Beobachtung am Exponat

Die zweite Teilstudie ging mit einer tiefergehenden Beobachtung an fünf ausgewählten Stationen einher. Es handelte sich dabei um eine Kombination aus a) Beobachtung spezifischer deduktiver Indikatoren für Lernen, die aus vorhandenen Theorien abgeleitet wurden, und b) vollständig offener Beobachtung mit Protokollierung aller Handlungen und Verhaltensweisen von Besucher*innen an einer Station, woraus induktiv weitere Indikatoren abgeleitet wurden.

Die Auswertung von Literatur über die kognitive und soziale Entwicklung von Kindern (vgl. Galinsky 2010, Vygotsky 1978) ergab neun deduktive Kategorien bzw. Indikatoren (vgl. Abb. 2), die mit Lernen in Verbindung gebracht werden. Fünf Indikatoren stehen mit der Wahrnehmung, der physischen

Interaktion und Kognition eines Individuums in Verbindung: »beobachten«, »konzentriert sein«, »trial and error«, »wiederholen« und »laut denken«. Zwei Indikatoren sind mit sozialer Interaktion, aber auch Kognition verbunden: »Entdeckungen mit anderen teilen« und »anderen die Handhabung erklären«. Zwei Indikatoren beziehen sich auf Emotionen: »frustriert sein« und »zufrieden sein«.

Die Balkendiagramme (vgl. Abb. 2) veranschaulichen, in welcher Ausprägung sich jeder der genannten Indikatoren an den fünf Exponaten zeigte. Zwei interessante Aspekte können aus den Diagrammen abgeleitet werden. Es wird erstens spezifisch sichtbar, für welche Exponate eine bestimmte Kategorie des Lernens vorherrschte. Zweitens wird generalisierend erkennbar, welche der Lernkategorien im Allgemeinen häufiger beobachtet werden konnten als andere. Im Detail ist u. a. erkennbar, dass Kinder am Exponat Pustespiele im Vergleich zu den anderen Exponaten zwar am meisten beobachteten, also ihre Aufmerksamkeit bewusst auf Teile der Station lenkten, um Informationen direkt über die Sinne aufzunehmen, aber in Relation zu den anderen Exponaten auch das geringste Wiederholungs- oder Trial-and-Error-Verhalten zeigten. Trial-and-Error-Verhalten involvierte, dass jemand eine Handlung ausprobiert, die vielversprechend erscheint, sich aber auch als Misserfolg entpuppen kann.

Generell fällt Trial-and-Error-Verhalten, das am ehesten auf einen Lernfortschritt schließen lässt, bei allen evaluierten Exponaten etwas weniger stark aus als bloßes Beobachten oder das reine Wiederholen von bereits erprobten Handlungen. Zudem sind die sozialen Interaktionsindikatoren »anderen die Handhabung erklären« und »mit anderen Entdeckungen teilen« weniger prominent ausgeprägt als jene Indikatoren, die die Interaktion mit den Exponaten betreffen. Hinsichtlich der emotionalen Zustände »frustriert sein« ließen Mimik, Gestik und Sprache der Kinder eher auf Zufriedenheit schließen, außerdem wurde an keinem der evaluierten Exponate mehr als »ein wenig« Frustration beobachtet.

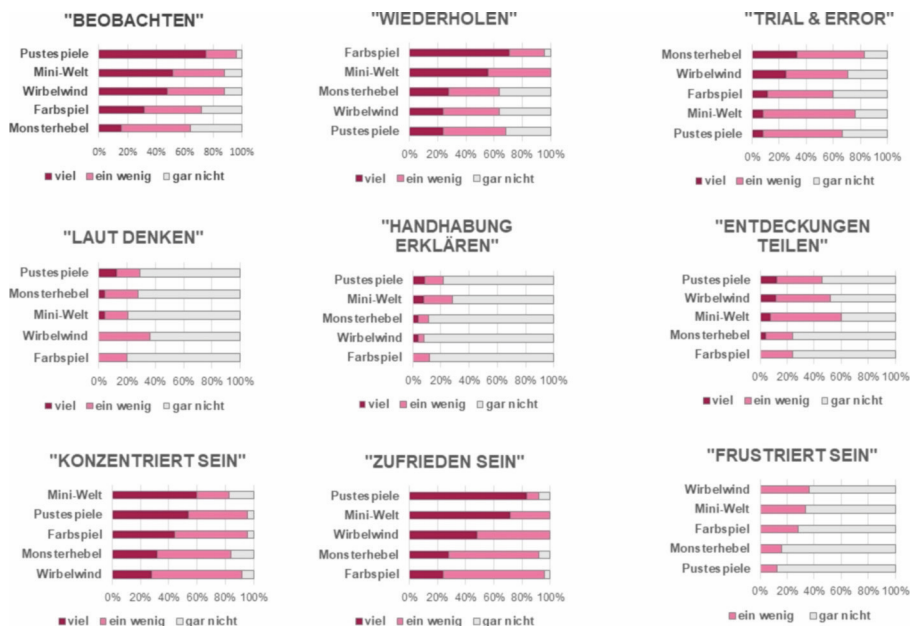


Abb. 2: Ausprägung der beobachteten neun Indikatoren an den Exponaten Farbspiel, Mini-Welt, Monsterhebel, Pustespiele und Wirbelwind des »experimenta-Forscherlands«, eigene Darstellung

Ungeachtet dessen, dass wir bereits deduktive Kategorien aus den Theorien des Lernens abgeleitet hatten, gingen wir nun noch einmal offen an die Erfassung der Verhaltensweisen heran, einerseits, um die quantitativen Daten zu unterfüttern, und andererseits, um weitere Kategorien aus dem Material zu entwickeln. Dies wird im Folgenden beispielhaft anhand der Station Wirbelwind (vgl. Abb. 3) deutlich.



Abb. 3: Die interaktive Station Wirbelwind im »Forscherland« der experimenta, 2019 (Foto: experimenta gGmbH)

An dieser Station können Besucher*innen erleben, wie ein Tornado entsteht. Auf Knopfdruck steigt feuchtwarme Luft auf und beginnt zu rotieren, wie es bei einem echten Tornado durch die Corioliskraft geschieht. Um die Rotation des Nebels besser beobachten zu können, ist die Station mit einem beweglichen Lasergerät ausgestattet, das Besucher*innen in den Wirbelwind richten können. Wir protokollierten zunächst in einer offenen Herangehensweise alle Verhaltensweisen, die eine Person von Beginn bis Ende während der Interaktion an dieser Station zeigte, so z.B. bei einem ca. siebenjährigen Jungen:

»M7? tritt an die Station, nimmt sofort das Lasergerät in die Hand und drückt die Taste. Er richtet den Laser nach unten auf die Luken, aus denen der Nebel strömt. Sein Blick richtet sich allerdings nach oben, dort, wo der Nebel nun aufsteigt und verwirbelt wird. Er dreht sich langsam um und richtet den Laser nach rechts weg vom Exponat in den Raum hinein. Er dreht sich dann wieder zum Exponat und richtet den Laser nun gezielter in den Wirbelsturm, schwenkt dabei langsam auf und ab. Er interagiert weiter mit dem Wirbelsturm und legt das Lasergerät ab. Er guckt in den Wirbel und nimmt das Gerät erneut in die Hand, um es wieder auf den Wirbelsturm zu richten. Der

Wirbelsturm dreht sich immer schneller. Der Junge legt das Lasergerät wieder hin und verlässt die Station.«²

Mittels qualitativer Inhaltsanalyse nach Kuckartz (vgl. 2018) gelang es uns zum einen, hieraus unterfütternde Informationen für unsere bestehenden deduktiven Kategorien zu gewinnen. Beispielsweise half uns die Beobachtung, tiefere Einblicke in das Kommunikationsverhalten an der Station Wirbelwind zu erlangen. Wir sahen, dass diese Station die Kommunikation zwischen Besucher*innen nur mäßig unterstützte. Die am häufigsten dokumentierten Kommunikationsformen waren das Erklären von Entdeckungen (20 %) und das Hinweisen auf Entdeckungen (12 %). Diese Arten der Kommunikation stehen in einem engen Zusammenhang mit dem Lernprozess. Dass begleitende Personen zuhörten oder zuschauten, schien das Interesse an der Station aufrechtzuerhalten. Wenn eine begleitende Person wegging, schien dies in einigen Fällen die Aufmerksamkeit des oder der Beobachteten zu beeinträchtigen.

Zum anderen legte eine Analyse des Textmaterials nun aber auch neue induktive Kategorien nahe, die nicht bereits in unseren Ausgangskategorien enthalten waren. Hierzu zählte die Kategorie »Aspekte der Handhabung«. An der Station Wirbelwind wurden zwei wesentliche Probleme in der Handhabung identifiziert: Ein Problem schien die Latenz zwischen dem Drücken des Knopfes und dem Aufsteigen und Drehen des Nebels zu sein. Dies kam vor allem zum Tragen, wenn die Ausstellung leer war und die Station zwischendurch immer wieder »herunterfuhr«. Das zweite Problem betraf das Lasergerät. Sein Zweck blieb vielen Besucher*innen unklar, sichtbar daran, dass der Laserstrahl nicht immer auf den Tornado gerichtet wurde. Zudem verstanden viele Nutzer*innen nicht, dass man das Gerät in die Hände nehmen kann, sondern gingen davon aus, dass es fest verankert sei.

Erst durch eine konsequente qualitative Inhaltsanalyse gelang es uns also, ein umfassenderes Bild zur Beantwortung der beiden Forschungsfragen: Welche spezifischen Indikatoren für Lernen lassen sich identifizieren?, und: Wie stark sind diese Indikatoren an den Stationen ausgeprägt?, zu zeichnen.

2 Beobachtungsprotokoll vom 9. Oktober 2019, Team Besucherforschung und Evaluation der experimenta.

wert erschienen. Diese Exponate wurden aus Gründen der Übersichtlichkeit aus der Grafik ausgespart.

gelernt?		Anzahl der Nennungen	Beispielaussagen
nein		8	Nein. Eher nicht.
nicht sicher		6	Ja, aber ich kann mich an nichts Spezielles erinnern. Neu war der Staubsaugerlift.
ja	konzeptuelles Wissen	34	Magnete: dass sich Dinge anziehen und abstoßen beim Strom – was leitet und was nicht dass ein Boot schwer untergeht
	prozedurales Wissen	5	ja, arbeiten wie in der Werkstatt Mit einem Flaschenzug geht es einfacher. Wir haben Sachen hochfliegen lassen!
	NoS (Tools + Methoden)	3	Infrarotlicht, dass man Wärme sichtbar machen kann Beim Mikroskop, das war seltsam, weil es plötzlich ganz anders aussah. das UV-Licht, womit man die Steine anschaute

Tab. 2: Analyse der qualitativen Daten zur Interviewfrage an die Eltern: Hat Ihr Kind hier heute etwas Neues entdecken oder etwas Bekanntes vertiefen können? (Interviewleitfaden, Team Besucherforschung und Evaluation der experimenta 2019)

Befragt danach, ob die Eltern den Eindruck gewonnen haben, ihr Kind habe in der Ausstellung »Forscherland« etwas gelernt, antworteten zwar acht Personen mit einem Nein und weitere sechs Eltern waren sich unsicher, aber 42 von den befragten 56 Eltern konnten explizit benennen, worin das Lernen aus ihrer Sicht bestanden hat. Aus den Antworten wurden drei Arten des Lernens bestimmt: a) eine Erweiterung konzeptuellen Wissens, b) ein Zugewinn proze-

duralen Wissens und c) Erkenntnisse hinsichtlich der Nature of Science (NoS), also der Methoden und Tools, die in den Naturwissenschaften zum Erkenntnisgewinn eingesetzt werden (vgl. Tab. 2).

Teilstudie IV: Befragung am Exponat

Mit der letzten Teilstudie wurden Aspekte erfasst, die ganz spezifisch mit einzelnen Stationen zusammenhingen. Es standen die Fragen im Mittelpunkt, ob die Besucher*innen an bekannte Dinge anknüpfen können und ob sie die Hauptthemen der Exponate erfassen.

Das Interview erfolgte kurz nach der Beobachtung an einer Station u.a. mit folgenden Fragen:

- »- Erzähl mir mal, was du hier gemacht hast! (evtl. nachfragen mit a und b)
- a) Ich habe gesehen, dass du ... Kannst du mir mehr dazu sagen?
- b) Hat dich etwas an dieser Station überrascht?
- Was würdest du einem Freund über die Station sagen, wenn er oder sie fragen würde, worum es geht?
- Hast du so etwas Ähnliches schon mal gemacht – oder war das für dich etwas ganz Neues?³

Durch eine qualitative Inhaltsanalyse der Antworten gelang es uns, ein umfassenderes Bild zur Beantwortung der beiden Fragen zu zeichnen. Das lässt sich am Beispiel des Exponates Wirbelwind erläutern. Nicht alle Interviewten verstanden das Konzept und konnten sagen, worum es bei diesem Exponat genau geht. Das Lasergerät spielte bei vielen Beobachtungen eine tragende Rolle, was die Frage aufwarf: Ist »Wirbelwind« ein Exponat über ein Tornadophänomen, ein Laserphänomen oder beides? Die Ergebnisse zeigten, dass nicht für alle Besucher*innen deutlich wurde, wie das Lasergerät mit dem Tornado zusammenhängt. Für uns stellte sich angesichts der Ergebnisse die Frage, ob die spezielle Gestaltung des Exponats die Besucher*innen hinsichtlich der Ursache eines Tornados verwirrt.

3 Auszug aus dem Interviewleitfaden, Team Besucherforschung und Evaluation der experimenta 2019.

Fazit: Was war der Mehrwert von Triangulation in unserem Projekt?

Flick verspricht sich von der Triangulation eine »Verbreiterung der Erkenntnismöglichkeiten« (Flick 2008: 19), eine Aussage, die unsere Studie stützt. Der Einsatz der Triangulation ermöglichte uns einen sowohl umfassenderen als auch detaillierteren Blick auf Aspekte und Prozesse, die charakteristisch für das Lernen unserer Besucher*innen sind, vor allem, was für sie attraktiv ist (anzieht, hält, in Erinnerung bleibt ...), welche unterschiedlichen Indikatoren sie an verschiedenen Stationen in Hinsicht auf Lernen zeigen, welche Arten des Lernens in der Ausstellung generell stattfinden und welche Faktoren Orientierung, Handhabung und damit das Lernen erschweren.

Nicht immer bestätigten sich die Daten der Teilstudien gegenseitig, es ergab sich aus ihnen eher ein komplexes, vielschichtiges Bild der Besucher*innerfahrung bzw. des Lernens. Besonders wertvoll für uns erwies sich der Wechsel von der Makroebene, dem »großen Überblick« (Wege, Nutzungsfrequenzen und -dauern, erinnerte Exponate, Verhaltensausrprägungen an Exponaten), zur Mikroebene (spezielle Verhaltensmuster, Lernformen und Inhalte).

Hieraus gewannen wir z. B. die Erkenntnis, dass, obwohl laut Beobachtung die Texte in 98 % der Fälle nicht gelesen wurden, ein hoher Prozentsatz der Besucher*innen (84 %) im Interview dennoch angab, mindestens einen der angebotenen Texte genutzt und ihn als hilfreich empfunden zu haben. Viele berichteten zudem, dass sie interaktive Exponate zunächst lieber selbst testeten und Texte grundsätzlich erst dann hinzuzögen, wenn sie nicht zurechtkämen.

Ein anderes Beispiel: Durch die Verschränkung von Timing- und Trackingdaten mit detaillierten Beobachtungen an Einzelexponaten wurde offensichtlich, dass einige Exponate zwar eine geringere erste »attraction power« auswiesen, aber dennoch viele mit Lernen verbundene Verhaltensweisen förderten und Kinder bzw. Jugendliche, sofern sie sich hieran niederließen, ausdauernd und wiederkehrend experimentierten. Auf der anderen Seite wurden auch Exponate identifiziert, die zwar eine höhere anfängliche Attraktion ausübten, Besucher*innen aber nur kurz an sich zu binden vermochten und Schwierigkeiten in der Handhabung bereiteten.

Nicht zuletzt legten die Interviews offen, was den bloßen Beobachtungen teils entgegenstand: So erwiesen sich im direkten Gespräch teilweise konzeptuelle Missverständnisse, obwohl die Befragten ausdauernd an einer Station gespielt hatten. Andere Interviewte überraschten mit einem klaren Verständnis für ein Phänomen, obwohl sie bei der vorhergehenden Interaktion frustriert gewirkt hatten. Die durch die nachfolgenden Interviews erfolgte Neube-

wertung der Beobachtungen überzeugte uns demnach u.a. davon, dass Frustration ein (wenn auch vorübergehend) erwünschter Zustand beim Interagieren an Exponaten sein kann, regte er doch in einigen Fällen offensichtlich zu weiterem Trial-and-Error-Verhalten an und führte damit potenziell zu einem erhöhten Verständnis. Zudem kamen wir zu der deutlichen Einsicht, dass nur eine Kombination von Beobachtungs- mit anschließenden Interviewdaten eine Lernsituation annähernd in ihrer Komplexität erfassen kann. Weder die reine Beobachtung, die für sich gesehen zu vorschnellen Schlüssen führen kann, noch das alleinstehende Interview, in dem eine Person im Sinne der Erfüllung angenommener Erwartungen antworten könnte, hätte dies ermöglicht.

Transfer: Triangulation für die Besucher*innenforschung?

»Randomized trials may systematically fail to account for phenomena that violate this method's basic assumptions – that is, phenomena that are contextually dependent or those that result from the interaction of dozens, if not hundreds, of factors. Indeed, such phenomena are precisely what educational research most needs to account for in order to have application to educational practice.« (Baumgartner et al. 2003: 6)

Wenn Studienergebnisse auf alltagsnahe Situationen übertragen werden sollen (für die Evaluation besonders relevant), sollte auch die Untersuchungssituation so gestaltet sein, dass sie mit der Alltagssituation in möglichst vielen Merkmalen übereinstimmt (vgl. Sedlmeier/Renkewitz 2008). Das allerdings steht den Ansprüchen eines quantitativen Paradigmas nach Kontrollierbarkeit und Wiederholbarkeit entgegen, wie das obige Zitat von Baumgartner et al. unterstreicht. Die Herausforderung besteht also in der Ausstellungsevaluation immer wieder darin, eine Balance zwischen einem naturalistischen, schwer kontrollierbaren, komplexen Setting und evidenzbasiertem Wissen herzustellen. Der Artikel hat gezeigt, dass Triangulation hierzu ein Schlüssel sein kann.

Aber: Dies bedeutet nicht automatisch mehr wissenschaftliche Güte. Wird eine Methode gewählt, die per se ungeeignet ist, um eine Forschungsfrage zu beantworten, kann sie nicht durch Triangulation mit anderen Methoden bessere Daten erbringen, möglicherweise erweitert sich das Problem dadurch eher noch. Wenn z.B. zwei verwendete Methoden eine niedrige Validität haben, addieren sich die dadurch entstehenden Erhebungsfehler bei Kombination der Methoden. Daher ist es bei der Triangulation besonders wichtig,

dass im Vorhinein adäquate Methoden gewählt werden, mit denen sich die Gegenstände, die untersucht werden wollen, tatsächlich erfassen lassen. Wollte man beispielsweise Besucher*innenhandlungen in ihren Einzelheiten und Reihenfolgen erheben, sollte man eher keine retrospektive Befragung einsetzen, denn dann würde man vielmehr das Erinnerungsvermögen oder auch soziale Erwünschtheit messen. Eine validere Methode wäre hier die verdeckte Beobachtung.

Wer Triangulation als Forschungsstrategie für seine Evaluation wählt, sollte zudem nicht unbedingt auf der Suche nach Vereinigung oder Bestätigung zwischen den Datensätzen sein. Auch wenn eine gegenseitige Bestätigung der verschiedenen Teilergebnisse in den Anfängen der Theorie der Triangulation oft noch als Ziel und Zweck verbalisiert wurde, so verfolgt ein moderneres Verständnis von Triangulation diesen Ansatz nicht mehr (vgl. Moran-Ellis et al. 2006). Wurde in den Sozialwissenschaften ursprünglich die Annahme verfolgt, dass ähnliche Ergebnisse die Genauigkeit der eingesetzten Forschungsmethoden bezeugen, voneinander abweichende Ergebnisse hingegen auf ein oder mehrere »fehlerhafte Messungen«⁴ verweisen (ebd.: 48), erkennt ein moderneres Verständnis an, dass jede Methode mit unvermeidlichen Schwächen behaftet ist, die sich gegenseitig ausgleichen können. Im Mittelpunkt des hier vorgestellten Triangulationsmodells steht also der Gewinn eines komplexeren Bildes vom Forschungsgegenstand und ein damit einhergehendes erhöhtes Vertrauen in die qualitative Forschung: Was mit einer Methode oberflächlich betrachtet wird oder ihr sogar ganz verborgen bleibt, wird durch andere Methoden vertieft oder sogar erst aufgedeckt, was insgesamt robustere soziologische Erklärungen dessen ermöglicht, was in unseren Ausstellungen geschieht.

Da Evaluation in vielen Häusern noch ein Luxusgut ist, sei zum Schluss der recht hohe Aufwand erwähnt, den eine Triangulation gemeinhin mit sich bringt. Unsere Erfahrung zeigte, dass es besser ist, weder zu viele noch zu große Forschungsfragen zu wählen, da man bereits beim sorgfältigen Aufschlüsseln einer scheinbar einfachen Frage wie: Was lernen unsere Besucher*innen?, oft unweigerlich zu dem Bedarf gelangt, verschiedene Methoden und Daten miteinander zu kombinieren.

4 »flawed measurement« (Übers. d. Verf.).

Literatur

- Baumgartner, Eric/Bell, Philip/Brophy, Sean/Hoadley, Christopher/Hsi, Sherry/Joseph, Diana (2003): »Design-Based Research. An Emerging Paradigm for Educational Inquiry«, in: *Educational Researcher* 32, H. 1, S. 5–8.
- Denzin, Norman K. (1970): *The Research Act. A Theoretical Introduction to Sociological Methods*, Chicago.
- Döring, Nicola/Bortz, Jürgen (2016): *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften*, Heidelberg.
- Flick, Uwe (2008): *Triangulation. Eine Einführung*, 2. Aufl., Wiesbaden.
- Ders. (2005): Standards, Kriterien, Strategien: Zur Diskussion über Qualität qualitativer Sozialforschung, in: *Zeitschrift für qualitative Bildungs-, Beratungs- und Sozialforschung* 6, H. 2, S. 191–210.
- Ders. (2004): *Triangulation. Eine Einführung*, Wiesbaden.
- Galinsky, Ellen (2010): *Mind in the Making: The Seven Essential Life Skills Every Child Needs*, New York.
- Kuckartz, Uwe (2018): *Qualitative Inhaltsanalyse. Methoden, Praxis, Computerunterstützung*, Weinheim.
- Moran-Ellis, Jo/Alexander, Victoria D./Cronin, Ann/Dickinson, Mary/Fielding, Jane/Sleney, Judith/Thomas, Hilary (2006): »Triangulation and Integration: Processes, Claims and Implications«, in: *Qualitative Research* 6, H. 1, S. 45–59.
- Sedlmeier, Peter/Renkewitz, Frank (2008): *Forschungsmethoden und Statistik in der Psychologie*, München.
- Vygotsky, Lew S. (1978): *Mind in Society: The Development of Higher Psychological Processes*, Cambridge.