Grundwissen Politik

LEHRBUCH

Markus Tausendpfund

Quantitative Methoden in der Politikwissenschaft

Eine Einführung

Springer VS

Grundwissen Politik

Begründet von
Ulrich von Alemann

Reihe herausgegeben von
Lars Holtkamp, Hagen, Deutschland
Viktoria Kaina, Hagen, Deutschland
Susanne Lütz, Hagen, Deutschland
Michael Stoiber, Hagen, Deutschland
Annette Elisabeth Töller, Hagen, Deutschland

Reihe herausgegeben von
Lars Holtkamp
Viktoria Kaina
Susanne Lütz
Michael Stoiber
Annette Elisabeth Töller

FernUniversität Hagen, Deutschland

Weitere Bände in der Reihe http://www.springer.com/series/12703

Markus Tausendpfund

Quantitative Methoden in der Politikwissenschaft

Eine Einführung

Springer VS

Markus Tausendpfund
Fakultät für Kultur- und Sozialwissenschaften
FernUniversität in Hagen
Hagen, Deutschland

Grundwissen Politik
ISBN 978-3-658-20697-0 ISBN 978-3-658-20698-7 (eBook)
https://doi.org/10.1007/978-3-658-20698-7

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über http://dnb.d-nb.de abrufbar.

Springer VS
© Springer Fachmedien Wiesbaden GmbH, ein Teil von Springer Nature 2018
Das Werk einschließlich aller seiner Teile ist urheberrechtlich geschützt. Jede Verwertung, die nicht ausdrücklich vom Urheberrechtsgesetz zugelassen ist, bedarf der vorherigen Zustimmung des Verlags. Das gilt insbesondere für Vervielfältigungen, Bearbeitungen, Übersetzungen, Mikroverfilmungen und die Einspeicherung und Verarbeitung in elektronischen Systemen.
Die Wiedergabe von Gebrauchsnamen, Handelsnamen, Warenbezeichnungen usw. in diesem Werk berechtigt auch ohne besondere Kennzeichnung nicht zu der Annahme, dass solche Namen im Sinne der Warenzeichen- und Markenschutz-Gesetzgebung als frei zu betrachten wären und daher von jedermann benutzt werden dürften.
Der Verlag, die Autoren und die Herausgeber gehen davon aus, dass die Angaben und Informationen in diesem Werk zum Zeitpunkt der Veröffentlichung vollständig und korrekt sind. Weder der Verlag noch die Autoren oder die Herausgeber übernehmen, ausdrücklich oder implizit, Gewähr für den Inhalt des Werkes, etwaige Fehler oder Äußerungen. Der Verlag bleibt im Hinblick auf geografische Zuordnungen und Gebietsbezeichnungen in veröffentlichten Karten und Institutionsadressen neutral.

Gedruckt auf säurefreiem und chlorfrei gebleichtem Papier

Springer VS ist ein Imprint der eingetragenen Gesellschaft
Springer Fachmedien Wiesbaden GmbH und ist Teil von Springer Nature
Die Anschrift der Gesellschaft ist: Abraham-Lincoln-Str. 46, 65189 Wiesbaden, Germany

Vorwort

Das vorliegende Lehrbuch bietet eine Einführung in die sozialwissenschaftliche Forschungslogik und den quantitativen Forschungsprozess. Dabei werden die Herausforderungen der Konzeptspezifikation und Operationalisierung vermittelt, die Formulierung von Hypothesen erläutert sowie Vor- und Nachteile verschiedener Untersuchungsformen, Auswahlverfahren und Varianten der Datenerhebung vorgestellt. Diese Informationen sind erforderlich, um sich einerseits gehaltvoll mit empirischen Studien auseinandersetzen zu können und andererseits die Stärken und Schwächen sozialwissenschaftlicher Publikationen zu identifizieren.

Das Buch ist weitgehend identisch mit dem Kurs „Quantitative Sozialforschung. Eine Einführung", der seit einigen Semestern im Modul M1 „Quantitative Methoden der Sozialwissenschaften" im BA-Studiengang „Politikwissenschaft, Verwaltungswissenschaft und Soziologie" an der FernUniversität in Hagen eingesetzt wird.

Freundlicherweise haben Lena Dieckmann, Marie-Sophie Heinelt, Nikolaus Hollermeier, Toni Alexander Ihme, Tuuli-Marja Kleiner, Raphael Kruse, Sebastian Kuhn, Kathrin Loer, Lisa Schäfer und Daniel Schubbe-Åkerlund einzelne Kapitel gelesen und kommentiert. Ein besonderer Dank geht an Christian Cleve, der sich sehr intensiv mit dem Text beschäftigt und mich auf Ungenauigkeiten sowie Tippfehler aufmerksam gemacht hat.

Profitiert hat dieses Buch auch durch die Rückmeldungen der Studierenden der FernUniversität in Hagen. Deshalb möchte ich mich herzlich bei allen Personen bedanken, die mich auf Fehler und Verbesserungsmöglichkeiten hingewiesen haben. Mein Dank gilt zudem allen Studierenden, die sich bei der Evaluation „Studienmaterial im Fokus" der FernUniversität in Hagen beteiligt haben.

Über Hinweise auf Fehler, Ungenauigkeiten und Verbesserungsvorschläge freue ich mich. Sie erreichen mich unter der E-Mail-Adresse Markus.Tausendpfund@fernuni-hagen.de.

Hagen, im Dezember 2017
Markus Tausendpfund

Inhalt

Vorwort V

Abbildungsverzeichnis XIII
Tabellenverzeichnis XV

1 Einführung 1
1.1 Methoden als Grundlage der Sozialwissenschaften 1
1.2 Kriterien sozialwissenschaftlicher Forschung 3
1.3 Ausblick auf die weiteren Kapitel 4

2 Grundbegriffe 7
2.1 Definitionen 7
2.2 Variable und Konstante 10
2.3 Abhängige und unabhängige Variable 11
2.4 Analyseebenen: Mikro und Makro 14
2.5 Qualitative und Quantitative Methoden 17
2.6 Erklärungen 26
2.7 Kritischer Rationalismus 29

3 Sozialwissenschaftlicher Forschungsprozess 33
3.1 Arten von Forschungsprojekten 33
3.2 Projektphasen im Überblick 35
3.2.1 Forschungsthema und Entwicklung einer Forschungsfrage 41
3.2.2 Konzeptspezifikation 42
3.2.3 Hypothesenbildung 43
3.2.4 Operationalisierung 44
3.2.5 Primär- und Sekundäranalyse 45
3.2.6 Forschungsdesign 46
3.2.7 Auswahlverfahren 48
3.2.8 Datenerhebung 49
3.2.9 Datenaufbereitung und -analyse 50
3.2.10 Publikation 51

4 Forschungsthema 53
4.1 Vom Forschungsthema zur Forschungsfrage 53
4.2 Literatur und Literaturrecherche 58
4.2.1 Was ist wissenschaftliche Fachliteratur? 59
4.2.2 Literaturrecherche 63
4.3 Strategien zur Formulierung einer Forschungsfrage 68
4.4 Relevanz der Forschungsfrage 73

5 Konzeptspezifikation 79
5.1 Über was sprechen wir? 80
5.2 Was ist ein Konzept? 83
5.3 Herausforderungen der Konzeptspezifikation 86
5.4 Hinweise für eine erfolgreiche Konzeptspezifikation 89

6 Hypothesen 91
6.1 Was sind Hypothesen? 91
6.2 Individualhypothesen 99
6.3 Kollektivhypothesen 100
6.4 Kontexthypothesen 103

7 Operationalisierung 107
7.1 Konzepte und Indikatoren 107
7.2 Messen in der empirischen Sozialforschung 115
7.3 Skalen 119
7.4 Gütekriterien der Operationalisierung 125
7.4.1 Objektivität 125
7.4.2 Reliabilität 126
7.4.3 Validität 129
7.4.4 Zusammenfassung 131
7.5 Indexbildung 132

8 Primär- und Sekundäranalyse 139
8.1 Was ist eine Primäranalyse? Was ist eine Sekundäranalyse? 139
8.2 Vor- und Nachteile von Primär- und Sekundäranalysen 141
8.3 Sozialwissenschaftliche Datensätze 145
8.3.1 Mikroebene 146
8.3.2 Makroebene 160
8.4 Recherche nach sozialwissenschaftlichen Datensätzen 164

9 Forschungsdesign 167
9.1 Untersuchungsebene 167
9.1.1 Individualdatenanalyse 168
9.1.2 Aggregatdatenanalyse 170
9.1.3 Kontextanalyse 171
9.1.4 Zusammenfassung 173
9.2 Untersuchungsform 173
9.2.1 Korrelation und Kausalität 174
9.2.2 Experiment 176
9.2.3 Ex-post-facto-Design 188
9.2.4 Zusammenfassung 193
9.3 Häufigkeit der Datenerhebung 194
9.3.1 Querschnittdesign 194
9.3.2 Längsschnittdesign 195
9.3.3 Zusammenfassung 204

10 Auswahlverfahren 207
10.1 Grundgesamtheit 207
10.2 Vollerhebung 209
10.3 Teilerhebung 210
10.3.1 Zufällige Auswahlverfahren 213
10.3.2 Nichtzufällige Auswahlverfahren 218
10.4 Auswahlverfahren in der Forschungspraxis 225
10.4.1 ADM-Design für persönliche Befragungen 225
10.4.2 Stichprobendesigns für telefonische Befragungen 230
10.4.3 Registerstichproben 234

11 Datenerhebung 237
11.1 Befragung 237
11.1.1 Befragungsarten 238
11.1.2 Fragearten 247
11.1.3 Frageformulierung 255
11.1.4 Antwortvorgaben 261
11.1.5 Gestaltung des Fragebogens 265
11.1.6 Pretest 267
11.1.7 Nonresponse 270
11.2 Beobachtung 274
11.2.1 Beobachtung in den Sozialwissenschaften 276
11.2.2 Varianten der Beobachtung 278
11.2.3 Phasen einer Beobachtung 280
11.3 Inhaltsanalyse 282
11.3.1 Inhaltsanalyse in den Sozialwissenschaften 284
11.3.2 Varianten der Inhaltsanalyse 285
11.3.3 Phasen einer Inhaltsanalyse 286

12 Datenaufbereitung und -analyse 291
12.1 Datenaufbereitung 291
12.2 Datenanalyse 297

13 Publikation 305
13.1 Publikationsformen 305
13.2 Sozialwissenschaftliche Fachzeitschriften 307

14 Weiterführende Literaturhinweise 315
14.1 Monographien und Sammelbände 315
14.2 Fachzeitschriften 319

Literatur 323

Abbildungsverzeichnis

Abbildung 1 Unabhängige und abhängige Variable 12
Abbildung 2 Mikro- und Makroebene als klassische Analyseebenen 15
Abbildung 3 Verhältnis von Wissenschaftstheorie und Sozialforschung 30
Abbildung 4 Phasen eines Forschungsprojekts 36
Abbildung 5 Recherchestrategien im Überblick 64
Abbildung 6 Mögliche Beziehungen zwischen zwei Merkmalen 97
Abbildung 7 Kontextanalyse 104
Abbildung 8 Konzepte und Indikatoren 108
Abbildung 9 Bildung und Einkommen 110
Abbildung 10 Kognitive Mobilisierung und Zustimmung zur EU 112
Abbildung 11 Morphismus 116
Abbildung 12 Morphismus am Beispiel formaler Bildungsabschlüsse 117
Abbildung 13 Konzept, Indikator und Messung 118
Abbildung 14 Übersicht der Gütekriterien 131
Abbildung 15 Konzept mit mehreren Indikatoren 133
Abbildung 16 Indexbildung des Konzepts „Kognitive Mobilisierung“ 135
Abbildung 17 GLES-Design im Überblick 154
Abbildung 18 Wahlbeteiligung bei der Europawahl 2014 170
Abbildung 19 Kontextanalyse am Beispiel der Wahlbeteiligung 172
Abbildung 20 Korrelationen 174
Abbildung 21 Nachher-Untersuchung mit Kontrollgruppe 179
Abbildung 22 Vorher-Nachher-Untersuchung mit Kontrollgruppe . . . 180
Abbildung 23 Solomon-Vier-Gruppen-Design 181

Abbildung 24 Trenddesign mit fiktiven Daten 196
Abbildung 25 Grund- und Auswahlgesamtheit sowie Stichprobe 212
Abbildung 26 Einfache Zufallsstichprobe 215
Abbildung 27 ADM-Design für persönliche Befragungen 226
Abbildung 28 Dual-Frame-Ansatz 233
Abbildung 29 Interviewanteile nach Befragungsart in Prozent 240
Abbildung 30 Klassifikation von Fragearten 249
Abbildung 31 Nonresponse 271
Abbildung 32 Entwicklung der Beteiligungsrate in Deutschland 273
Abbildung 33 SPSS-Datenansicht der ALLBUS 2014 294
Abbildung 34 SPSS-Variablenansicht der ALLBUS 2014 294
Abbildung 35 Errata-Angaben zur ALLBUS 2014 297
Abbildung 36 Auswahl von regressionsanalytischen Verfahren 302

Tabellenverzeichnis

Tabelle 1 Beispiele für Nominaldefinitionen 8
Tabelle 2 Beispiele für mögliche Ursache-Wirkungs-Beziehungen . . . 12
Tabelle 3 Politisches Interesse als aV und als uV 14
Tabelle 4 Qualitative und Quantitative Methoden im Vergleich 18
Tabelle 5 Struktur und Bestandteile einer D-N-Erklärung 27
Tabelle 6 Struktur und Bestandteile einer I-S-Erklärung 28
Tabelle 7 Anteile von Arbeitslosen und NSDAP-Wählern 102
Tabelle 8 Abstimmungsverhalten auf Individualebene 103
Tabelle 9 Skalenniveaus im Vergleich 124
Tabelle 10 Beispiele für additive Indexbildung 134
Tabelle 11 Beispiele für multiplikative Indexbildung 136
Tabelle 12 Vergleich verschiedener Arten von Experimenten 186
Tabelle 13 Vergleich von Querschnitt-, Trend- und Paneldesign 205
Tabelle 14 Beispiel für einen Schwedenschlüssel 228
Tabelle 15 Vergleich der Vor- und Nachteile von Befragungsarten . . . 241
Tabelle 16 Datenmatrix 292
Tabelle 17 Taschengeld von Stadt- und Landkindern 299
Tabelle 18 Zusammenhangsmaße in Abhängigkeit des Skalenniveaus 300
Tabelle 19 Ausgewählte deutschsprachige Fachzeitschriften 309
Tabelle 20 Politikwissenschaftliche Fachzeitschriften im SSCI 311
Tabelle 21 Soziologische Fachzeitschriften im SSCI 312
Tabelle 22 Verwaltungswissenschaftliche Fachzeitschriften im SSCI . . . 312

Einführung[1]

1

Die Methoden der empirischen Sozialforschung stellen einen wichtigen Bestandteil des sozialwissenschaftlichen Studiums dar. Die Bedeutung des Wissens über die sozialwissenschaftliche Forschungslogik, ihre Strategien und Techniken – kurz: ihre Methoden – wird in diesem Kapitel erläutert. Daran anknüpfend werden Merkmale sozialwissenschaftlicher Forschung dargestellt. Die Einführung schließt mit einem Ausblick auf die weiteren Kapitel dieses Buchs.

1.1 Methoden als Grundlage der Sozialwissenschaften

Warum unterscheidet sich die Wahlbeteiligung in westlichen Demokratien (z. B. Hadjar und Beck 2010)? Fördert das Internet die Beteiligung der Bürgerinnen und Bürger (z. B. Voss 2014)? Was sind die Ursachen kommunaler Haushaltsdefizite (z. B. Bogumil et al. 2014)? Warum unterscheidet sich das politische Wissen bei Kindern (z. B. Abendschön und Tausendpfund 2017)? Was ist ein Politikfeld und warum entsteht es (z. B. Loer et al. 2015)? Warum gefährdet mehr Demokratie die politische Gleichheit der Bürger (z. B. Schäfer und Schoen 2013)?

Diese sechs – exemplarischen – Fragestellungen illustrieren das große Themenspektrum der Sozialwissenschaften. Sozialwissenschaftler beschäftigen sich mit Fragen der politischen Beteiligung, untersuchen soziale Ungleichheit und erforschen die Auswirkungen institutioneller Strukturen auf individuelle Einstel-

1 Ausschließlich aus Gründen der besseren Lesbarkeit wird in diesem Buch nicht durchgängig eine geschlechterneutrale Sprache verwendet. Männliche, weibliche und genderneutrale Formen wechseln sich in diesem Buch zufallsverteilt ab. Mit den Bezeichnungen sind jeweils alle Geschlechter gemeint.

© Springer Fachmedien Wiesbaden GmbH, ein Teil von Springer Nature 2018

M. Tausendpfund, *Quantitative Methoden in der Politikwissenschaft*,
Grundwissen Politik, https://doi.org/10.1007/978-3-658-20698-7_1

lungen und Verhaltensweisen. Zusammengefasst: Sozialwissenschaftlerinnen beschreiben und erklären soziale Sachverhalte.

Für die Beschreibung und Erklärung sozialer Sachverhalte steht eine große Anzahl an Strategien und Techniken zur Verfügung: Experimente, Querschnitts- und Längsschnittstudien, Befragungen, Beobachtungen oder auch Textanalysen. Die Gesamtheit dieser Strategien und Techniken wird als „Methoden" bezeichnet. Diese Methoden – also die Strategien und Techniken zur Bearbeitung einer sozialwissenschaftlichen Forschungsfrage – stehen im Mittelpunkt des vorliegenden Buchs.

Metapher „Werkzeugkiste" Zur besseren Illustration der Methoden der Sozialwissenschaften nutzt Diekmann (2011, S. 17) die nützliche Metapher der „Werkzeugkiste". Bei den Strategien und Techniken der Sozialwissenschaften handelt es sich um Werkzeuge, die zur Bearbeitung einer konkreten Forschungsfrage zur Verfügung stehen – vergleichbar mit Schraubenzieher, Hammer und Zollstock in einer Werkzeugkiste. Forscherinnen müssen die verschiedenen Werkzeuge kennen, um das für die konkrete Forschungsfrage geeignete Werkzeug auswählen zu können. Im übertragenen Sinn: Ein Hammer eignet sich beispielsweise dafür Nägel in die Wand zu schlagen, für das Anschrauben einer Tafel wäre ein Schraubenzieher einem Hammer allerdings vorzuziehen. Das Beispiel illustriert: Für jedes Problem muss das geeignete Werkzeug ausgewählt werden. Eine Sozialwissenschaftlerin sollte allerdings nicht nur die verschiedenen Werkzeuge kennen, auch die sachgemäße Handhabung der zahlreichen Techniken muss gelernt werden. Wer eine Kreissäge nicht ordnungsgemäß bedient, der gefährdet sein Leben. Wer nicht die Prinzipien der Fragebogenkonstruktion kennt, der produziert bei einer Befragung unsinnige Ergebnisse.

Methoden werden immer wichtiger Aber nicht nur der Anwender – die Person, die ein konkretes Forschungsprojekt durchführt –, sondern auch der passive Beobachter, der das spezifische „Produkt" in Augenschein nimmt, muss über eine angemessene sozialwissenschaftliche Methodenkompetenz verfügen. Wie soll die Hauseigentümerin die wärmetechnische Sanierung einer Fachfirma bewerten, wenn sie nicht über entsprechende Grundlagenkenntnisse verfügt? Formuliert in der Sprache der empirischen Sozialforschung: Wie wollen Sie sich mit sozialwissenschaftlichen Studien auseinandersetzen, theoretische Argumente reflektieren, empirische Befunde bewerten und Schlussfolgerungen kritisieren, wenn Sie nicht die entsprechenden Strategien und Techniken der empirischen Sozialforschung kennen?

Aus diesem Grund bilden die Methoden der empirischen Sozialforschung einen wichtigen Bestandteil des sozialwissenschaftlichen Studiums. Die sozial-

wissenschaftliche „Methodenausbildung ist wichtig und wird immer wichtiger“ (Schnapp et al. 2004, S. 160; siehe auch Kohler 2016). Zum einen sind grundlegende Methodenkenntnisse für viele Berufsfelder erforderlich, zum anderen haben sich die Sozialwissenschaften weiterentwickelt. Ohne grundlegende Methodenkompetenz sind zahlreiche Themenfelder der Sozialwissenschaften (z. B. Wahlforschung) nicht mehr zugänglich (siehe dazu die Beiträge in Engel 2002). Die Methoden bilden die Grundlage der Sozialwissenschaften. Die Methoden machen die Sozialwissenschaften überhaupt erst zu einer Wissenschaft (King et al. 1994, S. 9).

Zielsetzung des Lehrbuchs Im vorliegenden Buch werden die sozialwissenschaftliche Forschungslogik und grundlegende Forschungsmethoden vorgestellt. Das Ziel ist die „methodische Alphabetisierung“ (Literacy). Studierende sollen durch die Lektüre ein besseres Verständnis sozialwissenschaftlicher Forschung erhalten, um sich mit empirischen Studien angemessen auseinandersetzen zu können.

1.2 Kriterien sozialwissenschaftlicher Forschung

In dem einflussreichen Werk „Designing Social Inquiry“ entwickeln King et al. (1994, S. 7–9) vier Kriterien zur Charakterisierung sozialwissenschaftlicher Forschung:

„The goal is inference.“ 1. Das Ziel wissenschaftlicher Forschung ist Inferenz. Sozialwissenschaftliche Forschung beschreibt und erklärt auf Basis empirischer Informationen soziale Sachverhalte. Sorgfältige Beschreibungen des konkreten Sachverhaltes sind häufig unverzichtbar, aber die reine Sammlung von Faktenwissen ist nicht ausreichend für Wissenschaft. Die Wissenschaft geht über die reine Beschreibung der empirischen Beobachtungen hinaus. Das Ziel ist „descriptive inference“ oder „causal inference“. Bei deskriptiver Inferenz werden die empirischen Beobachtungen genutzt, um über nicht beobachtete Fakten etwas zu lernen. Bei kausaler Inferenz werden Erklärungen über die empirischen Beobachtungen angestrebt.

„The procedures are public.“ 2. Die wissenschaftliche Vorgehensweise ist öffentlich. Sozialwissenschaftliche Forschung basiert auf einer transparenten und nachvollziehbaren Vorgehensweise. Diese Vorgehensweise ist öffentlich zugänglich. Andere Personen (z. B. Forscher, Interessierte) können sich mit den empirischen Beobachtungen, der theoretischen Argumentation, der methodischen Vorgehensweise und den Schlussfolgerungen auseinandersetzen und diese kritisch begleiten.

> *„If the method and logic of a researcher's observations and inferences are left implicit, the scholarly community has no way of judging the validity of what was done. We cannot evaluate the principles of selection that were used to record observations, the ways in which observations were processed, and the logic by which conclusions were drawn. We cannot learn from their methods or replicate their results. Such research is not a public act." (King et al. 1994, S. 8)*

„The conclusions are uncertain." 3. Die Schlussfolgerungen sind unsicher. Bei der Erhebung, der Analyse und der Interpretation von empirischen Informationen (Daten) sind Fehler möglich. Deshalb sind die Schlussfolgerungen der wissenschaftlichen Vorgehensweise grundsätzlich unsicher, da sie auf unsicheren Daten beruhen. Das Ausmaß dieser Unsicherheit lässt sich allerdings abschätzen.

„The content is the method." 4. Die Gemeinsamkeit ist die Methode. Das Themenspektrum der Sozialwissenschaften ist reichhaltig. Sozialwissenschaftliche Forschung lässt sich daher nicht über Inhalte definieren, sondern nur über die sozialwissenschaftliche Methode. Die sozialwissenschaftliche Forschungslogik, ihre Strategien und Techniken bilden die Grundlage der Wissenschaft.

Auf Basis dieser vier Merkmale leiten King et al. (1994, S. 9) eine weitere Implikation ab:

> *„Science at its best is a social enterprise. Every researcher or team of researchers labors under limitations of knowledge and insight, and mistakes are unavoidable, yet such errors will likely be pointed out by others. Understanding the social character of science can be liberating since it means that our work need not to be beyond criticism to make an important contribution – whether to the description of a problem or its conceptualization, to theory or to the evaluation of theory."*

Die Auseinandersetzung mit der sozialwissenschaftlichen Forschung – mit ihren Theorien, ihren Analysen und ihren Befunden – setzt die Kenntnis der sozialwissenschaftlichen Methoden voraus. Nur wer die sozialwissenschaftliche Forschungslogik, ihre Strategien und Techniken kennt, kann sich am sozialen Unternehmen „Wissenschaft" angemessen beteiligen.

1.3 Ausblick auf die weiteren Kapitel

Das vorliegende Buch bietet eine Einführung in die empirische Sozialforschung und stellt grundlegende Forschungsmethoden vor. Die Darstellung orientiert sich dabei am sozialwissenschaftlichen Forschungsprozess. Die einzelnen Phasen des

Forschungsprozesses – z. B. Forschungsfrage, Konzeptspezifikation, Hypothesen, Operationalisierung, Forschungsdesign, Auswahlverfahren und Datenerhebung – werden dargestellt und erläutert.

Grundbegriffe Das anschließende zweite Kapitel stellt Grundbegriffe der empirischen Sozialforschung vor. Die Kenntnis dieser Begriffe bildet die Basis für die weiteren Ausführungen in diesem Buch und dient dazu, sich mit dem spezifischen sozialwissenschaftlichen Vokabular vertraut zu machen.

Forschungsprozess Das dritte Kapitel bietet eine Übersicht über den sozialwissenschaftlichen Forschungsprozess. Dabei werden typische Phasen vorgestellt und erläutert. Dieses erste Kennenlernen des (kompletten) Forschungsprozesses soll helfen, die einzelnen Phasen sozialwissenschaftlicher Forschung in den Gesamtzusammenhang einordnen zu können.

Die Kapitel 4 bis 13 vertiefen die einzelnen Phasen des sozialwissenschaftlichen Forschungsprozesses. Angefangen von der Auswahl eines Forschungsthemas und der Entwicklung einer Forschungsfrage (Kapitel 4), der Bedeutung der Konzeptspezifikation (Kapitel 5), der Formulierung von Hypothesen (Kapitel 6), der Operationalisierung (Kapitel 7), der Differenzierung von Primär- und Sekundäranalyse (Kapitel 8) über die Entwicklung eines Forschungsdesigns (Kapitel 9), der Festlegung des Auswahlverfahrens (Kapitel 10) und der Durchführung der Datenerhebung (Kapitel 11) bis hin zur Datenaufbereitung und -analyse (Kapitel 12) sowie der Publikation (Kapitel 13).

Weiterführende Literatur Dieses Buch bietet eine Einführung in die sozialwissenschaftliche Forschungslogik und stellt grundlegende Forschungsmethoden vor. Mittlerweile existiert umfangreiche Literatur zu den Grundlagen und fortgeschrittenen Methoden der empirischen Sozialforschung. Deshalb finden sich in Kapitel 14 weiterführende Literaturhinweise. Zum einen werden ausgewählte Bücher vorgestellt, die einzelne Themen dieser Einführung vertiefen und/oder einen allgemeinen Überblick geben. Zum anderen werden Fachzeitschriften dargestellt, die sich schwerpunktmäßig mit den Methoden der empirischen Sozialforschung beschäftigen.

Grundbegriffe

2

In der empirischen Sozialforschung wird zur präzisen Kommunikation ein spezifisches Vokabular genutzt. In diesem Kapitel werden zentrale Grundbegriffe vorgestellt, die bei der Lektüre empirischer Studien häufig vorausgesetzt werden.

2.1 Definitionen

Eine Voraussetzung für die Beschreibung und Erklärung sozialer Sachverhalte ist die Präzisierung der verwendeten Begriffe. Wenn Personen Begriffe wie „Demokratie", „Legitimität", „Macht" oder auch „Familie", „Parteiidentifikation" und „Kognitive Mobilisierung" unterschiedlich verstehen, dann ist ein gehaltvoller Austausch über empirische Sachverhalte praktisch ausgeschlossen. Damit Aussagen über soziale Sachverhalte nicht nur von einer einzigen oder wenigen Personen verstanden werden, sondern möglichst allen interessierten Personen zugänglich sind, muss die gemeinte Bedeutung der verwendeten Begriffe von allen Personen erfasst und geteilt werden können.

> *„Man muss wissen, worüber geredet wird. Die am Kommunikationsprozess Beteiligten sollen den verwendeten Begriffen weitestgehend die gleichen Bedeutungen, die gleichen Vorstellungsinhalte zuschreiben." (Schnell et al. 2013, S. 46)*

Opp (2014, S. 119) versteht unter einer Definition die Einführung einer Konvention über die Verwendung bestimmter Begriffe. Sie soll zu mehr Klarheit und Deutlichkeit beitragen, indem bestimmte Dinge, Sachverhalte oder Ereignisse eindeutig abgegrenzt werden können. In den Sozialwissenschaften werden drei verschiedene Definitionsarten unterschieden: Nominaldefinition, Realdefinition und operationale Definition.

© Springer Fachmedien Wiesbaden GmbH, ein Teil von Springer Nature 2018

M. Tausendpfund, *Quantitative Methoden in der Politikwissenschaft*,
Grundwissen Politik, https://doi.org/10.1007/978-3-658-20698-7_2

Nominaldefinition Bei einer Nominaldefinition handelt es sich um eine reine Bedeutungsübertragung. Ein bestimmter Ausdruck A_1 wird mit einem anderen Ausdruck A_2 gleichgesetzt. Dabei wird die Bekanntheit des Ausdrucks A_1 vorausgesetzt (Opp 2014, S. 119). Entsprechend besteht eine Nominaldefinition aus zwei Bestandteilen: Definiens (A_1) und Definiendum (A_2). Das Definiendum (übersetzt: „das zu Definierende") ist der Ausdruck, für den die Definition erstellt werden soll. Das Definiens (übersetzt: „das Definierende") ist der Ausdruck, der das Definiendum bestimmt (bzw. definiert).

Eine Nominaldefinition legt fest, wie ein (neuer) Begriff verwendet wird, indem sie auf bereits bekannte bzw. definierte Begriffe Bezug nimmt. In Tabelle 1 finden sich einige Beispiele für Nominaldefinitionen. Die Nominaldefinition für Schimmel ist „Pferd mit weißem Fell", ein Rappe ist ein „Pferd mit schwarzem Fell" und die Bezeichnung Fohlen wird für ein „Junges Pferd" verwendet. Eine Nominaldefinition ist eine kurze und handliche Bezeichnung für einen (mehr oder weniger) komplexen Begriff.

Tabelle 1 Beispiele für Nominaldefinitionen

Definiens (A_1)	Definiendum (A_2)
Pferd mit weißem Fell	Schimmel
Pferd mit schwarzem Fell	Rappe
Junges Pferd	Fohlen

Bei einer Nominaldefinition ist es möglich, dass die im Definiens verwendeten Begriffe selbst nicht eindeutig sind. Sollte beispielsweise unklar sein, was mit „Pferd" oder „Fell" gemeint ist, dann müssten beide Begriffe selbst zum Definiendum erhoben werden. Dies führt dann zu weiteren Definitionen. Um einen unendlichen Definitionsprozess zu vermeiden („definitorischer Regress"), muss der Definitionsprozess an einer Stelle abgebrochen werden. Bestimmte Begriffe müssen in ihrer Bedeutung als allgemein verständlich akzeptiert werden (Schnell et al. 2013, S. 47).

! Bei einer Nominaldefinition sind definitorische Zirkel zu vermeiden. Dabei handelt es sich um Definitionen, „in denen die zu definierenden Begriffe selbst zur Definition verwendet werden" (Opp 2014, S. 123). „Politische Beteiligung ist, wenn sich ein Bürger politisch beteiligt" ist ein einfaches Beispiel für einen definitorischen Zirkel.

Definitorischer Zirkel: Was sind Werte?

Eine klassische und vielfach beachtete Definition von Werten hat Clyde Kluckhohn (1967, S. 395) vorgelegt: „A value is a conception, explicit or implicit, distinctive of an individual or characteristic of a group, of the desirable which influences the selection from available modes, means, and ends of action." In einer kritischen Auseinandersetzung bewerten Jan W. van Deth und Elinor Scarbrough (1995, S. 27) diese Nominaldefinition als zirkulär. Die englische Sprache verdeckt die Probleme mit den Begriffen „value" und „desirable", die aber in der deutschen Übersetzung mit „Wert" und „wünschenswert" offensichtlich werden. „Ist etwas ein Wert, weil es wünschenswert ist, oder ist etwas wünschenswert, weil es ein Wert ist", fragt Simone Abendschön (2010, S. 46). Nach van Deth und Scarbrough (1995, S. 27) sind in der vorgelegten Definition von Kluckhohn Teile des Definiens auch Bestandteile des Definiendums, sodass die Definition zirkulär ist.

Nominaldefinitionen erfüllen eine wichtige sprachökonomische Funktion. Mit Hilfe eines „kurzen" Begriffs (z. B. Hypothese) können längere Formulierungen (z. B. Aussage über einen Zusammenhang zwischen mindestens zwei Merkmalen) vermieden werden. Mit Nominaldefinitionen wird einerseits Platz in Publikationen gespart und andererseits die Kommunikation zwischen Wissenschaftlern beschleunigt (Opp 2014, S. 121). Deshalb bieten sich Nominaldefinitionen an, wenn ein Definiens relativ häufig benötigt wird.

Die Nominaldefinition ist ein wichtiger Bestandteil der Konzeptspezifikation (siehe Kapitel 5). Schnell et al. (2013, S. 118) betrachten die Konzeptspezifikation als „eine spezielle Art der Nominaldefinition". Nominaldefinition und Konzeptspezifikation zielen jeweils darauf ab, einen unklaren Begriff eindeutig zu definieren und von anderen Begriffen abzugrenzen, damit die gemeinte Bedeutung der verwendeten Begriffe von allen Personen erfasst werden kann.

Realdefinition Von Nominaldefinitionen sind Realdefinitionen abzugrenzen, die darauf abzielen, das Wesen eines Sachverhalts zu erfassen (Opp 2014, S. 124–125; Häder 2015, S. 31; Kromrey et al. 2016, S. 157–160). Bei einer Realdefinition wird nicht geklärt, wie beispielsweise Begriffe wie „Mensch" oder „Staat" verwendet werden oder sollten, sondern was „Mensch" und „Staat" ihrem Wesen nach sind.

Damit ist eine Realdefinition eine Behauptung über die Beschaffenheit bzw. das Wesen eines Sachverhalts. Während eine Nominaldefinition mehr oder weniger zweckmäßig ist, kann eine Realdefinition richtig oder falsch sein. Eine Real-

definition legt die Eigenschaften eines Sachverhalts, die für den Sachverhalt „wesentlich“ – nicht was vereinbarungsgemäß als wesentlich betrachtet wird – sind, verbindlich fest (Kromrey et al. 2016, S. 159).

! Allerdings gibt es keine objektiven Kriterien, ob durch eine vorgeschlagene Realdefinition das Wesen eines Sachverhalts auch angemessen erfasst wird. Mit anderen Worten: Es kann nicht geklärt werden, ob eine Realdefinition richtig oder falsch ist (Häder 2015, S. 31; siehe für verschiedene Bedeutungen der Wesensbestimmungen auch Opp 2014, S. 124–125). Nach Schnell et al. (2013, S. 47–48) eignen sich deshalb Realdefinitionen nicht für wissenschaftliche Arbeiten, da keine Kriterien existieren, ob das Wesen eines Sachverhalts durch die Definition auch angemessen erfasst wird.

Operationale Definition Operationale Definitionen schaffen eine Verbindung zwischen Begriffen und der Realität. Eine operationale Definition legt fest, wie ein theoretischer Begriff (z. B. Intelligenz) mit Indikatoren zu verknüpfen ist, die sich empirisch beobachten lassen (Häder 2015, S. 32; Opp 2014, S. 133–141). Die Übersetzung eines theoretischen Begriffs in empirische Indikatoren wird allgemein „Operationalisierung“ genannt und in Kapitel 7 ausführlich behandelt.

! Wichtig ist an dieser Stelle der Hinweis, dass Begriffen wie „Bildung“ oder „Demokratie“ in empirischen Studien durchaus unterschiedliche Indikatoren zugeordnet werden. Das heißt, die Indikatoren eines Begriffs unterscheiden sich zwischen Studien. Dies kann erhebliche Konsequenzen für die empirischen Ergebnisse haben. Bei der kritischen Auseinandersetzung mit empirischen Studien müssen deshalb stets die verwendeten Indikatoren auf ihre Angemessenheit überprüft werden.

2.2 Variable und Konstante

Politikwissenschaftlerinnen beschäftigen sich mit den unterschiedlichsten Untersuchungsobjekten: Menschen, Gemeinden, Organisationen oder auch Staaten. Diese Objekte weisen eine unüberschaubare Anzahl an Merkmalen auf. Bei Menschen sind das beispielsweise Geschlecht, Alter, Einkommen, aber auch politisches Interesse, soziales Vertrauen oder Intelligenz. Bei Gemeinden sind die Einwohnerzahl, bei Organisationen die Anzahl der Hierarchieebenen und bei Staaten die Regierungsform mögliche interessante Merkmale. Solche Merkmale werden in der Sprache der empirischen Sozialforschung als Variablen bezeichnet. Der Begriff Variable deutet auf die Variabilität eines Merkmals hin. Bei der Varia-

ble Geschlecht lassen sich beispielsweise Männer und Frauen unterscheiden, bei Gemeinden kleine und große Gemeinden und bei (demokratischen) Staaten parlamentarische, präsidentielle und semipräsidentielle Regierungsformen.

- **Definition: Variable und Konstante**
  Eine Variable ist ein sozialwissenschaftliches Merkmal mit mindestens zwei Ausprägungen. Ein Merkmal mit nur einer Ausprägung wird Konstante bezeichnet.

Die Bezeichnung Variable ist eine allgemeine Beschreibung – ein sogenannter Platzhalter – für ein sozialwissenschaftliches Merkmal. Der Inhalt einer Variable (z. B. Geschlecht, Gemeindegröße oder Staatsform) ist jeweils von der konkreten Forschungsfrage abhängig.

Manifeste und latente Variablen In den Sozialwissenschaften werden manifeste und latente Variablen unterschieden. Bei manifesten Variablen handelt es sich um sozialwissenschaftliche Merkmale, die direkt beobachtbar sind. Eine manifeste Variable ist beispielsweise das Geschlecht, die Größe oder auch die Haarfarbe einer Person. Dagegen sind latente Variablen Merkmale, die sich der direkten Beobachtung entziehen. Latente Variablen sind beispielsweise Intelligenz, Einstellungen wie Ausländerfeindlichkeit oder auch soziales Vertrauen. Für eine empirische Untersuchung müssen latente Variablen erst „beobachtbar" gemacht werden. Dieser Vorgang wird als Operationalisierung bezeichnet (siehe Kapitel 7).

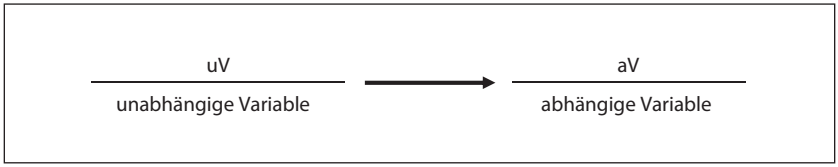
2.3 Abhängige und unabhängige Variable

Das Ziel sozialwissenschaftlicher Forschung ist die Beschreibung und die Erklärung sozialer Sachverhalte. Nach King et al. (1994, S. 7–8) ist die sorgfältige Beschreibung zwar unentbehrlich für wissenschaftliche Forschung, aber die Sammlung von Fakten und Informationen ist allein nicht ausreichend für wissenschaftliche Forschung. Sozialwissenschaftliche Forschung zielt darauf ab, soziale Sachverhalte zu erklären. Warum unterscheiden sich die Leistungen von Schülern bei der PISA-Studie? Welche Merkmale beeinflussen die Wahlbeteiligung? Warum unterscheidet sich das Gehalt von Männern und Frauen? Sozialwissenschaftliche Forschung ist insbesondere an Ursache-Wirkungs-Beziehungen interessiert.

aV und uV Zur allgemeinen Bezeichnung von (vermuteten) Zusammenhängen zwischen zwei Merkmalen haben sich in der Wissenschaftssprache die Abkür-

zungen aV und uV durchgesetzt.[2] Die Abkürzung aV steht für abhängige Variable (auch als bedingte Variable, Zielvariable oder Kriteriumsvariable bezeichnet), die Bezeichnung uV für unabhängige Variable. Die abhängige Variable bezeichnet den Sachverhalt, der erklärt werden soll. Die unabhängige Variable ist die Bezeichnung für das Merkmal, das einen (vermuteten) Einfluss auf die abhängige Variable ausübt. Die uV wird manchmal auch erklärende Variable oder Prädiktorvariable genannt. Abbildung 1 illustriert die Beziehung zwischen unabhängiger und abhängiger Variable. Das Modell unterstellt, dass die uV einen (kausalen) Einfluss auf die aV ausübt.

Abbildung 1 Unabhängige und abhängige Variable



Quelle: Eigene Darstellung

Auf Grundlage theoretischer Überlegungen wird zwischen uV und aV ein Zusammenhang begründet. In Tabelle 2 sind die Ursache-Wirkungs-Beziehungen für die formulierten Beispielfragen dokumentiert. Die Motivation hat vermutlich einen Einfluss auf das PISA-Testergebnis. In diesem Beispiel stellt die Motivation die unabhängige Variable, das PISA-Testergebnis die abhängige Variable dar. Im zweiten

Tabelle 2 Beispiele für mögliche Ursache-Wirkungs-Beziehungen

unabhängige Variable		abhängige Variable
Motivation	→	PISA-Testergebnis
politisches Interesse	→	Wahlteilnahme
Geschlecht	→	Gehalt

2 In der englischen Fachliteratur finden sich die analogen Bezeichnungen wie „dependent variable“ (dv) und „independent variable“ (iv). „Dependent variable“ steht dabei für „abhängige Variable“ und „independent variable“ für „unabhängige Variable“.

Beispiel ist das politische Interesse die unabhängige Variable, die Wahlbeteiligung die abhängige Variable. Schließlich stellt das Geschlecht einen Prädiktor (uV) für das Gehalt (aV) dar.

In einem wissenschaftlichen Text sind die theoretischen Überlegungen für den vermuteten Zusammenhang zwischen einer uV und einer aV selbstverständlich auszuführen und mit Forschungsliteratur zu verknüpfen. An dieser Stelle steht die Einführung dieser zentralen Begriffe im Vordergrund. Zu Beginn des Studiums wirken die Begriffe aV und uV meist noch etwas abstrakt. Die Verwendung der Begriffe fördert allerdings die Klarheit der Argumentation. Die Darstellung eines möglichen Ursache-Wirkungs-Zusammenhangs erfordert einerseits eine aV klar und präzise festzulegen und andererseits über mögliche Einflussfaktoren (die uVs) nachzudenken.

Häufig mehrere uVs Die gewählten Beispiele machen implizit auch deutlich, dass in der Regel nicht nur eine uV, sondern mehrere unabhängige Variablen existieren.

Eine aV, aber mehrere uVs

In einer wissenschaftlichen Arbeit steht meist eine abhängige Variable im Mittelpunkt. Es werden aber mehrere unabhängige Variablen betrachtet, die die abhängige Variable „erklären" sollen. Auf Basis des Forschungsstands werden die möglichen unabhängigen Variablen identifiziert, die einen Einfluss auf die abhängige Variable haben können. Die möglichen Zusammenhänge zwischen uV und aV sind theoretisch zu begründen und empirisch zu überprüfen.

Das PISA-Testergebnis wird sicherlich nicht nur von der individuellen Motivation des Schülers abhängen, sondern auch von anderen Faktoren (z. B. Intelligenz). Die Wahlteilnahme wird nicht nur vom politischen Interesse abhängig sein, sondern auch von der internalisierten Wahlnorm. Schließlich wird sich nicht nur das Geschlecht auf das Einkommen auswirken, sondern auch die Beschäftigungsart (z. B. Voll- versus Teilzeit).

Festlegung von aV und uV ist kontextabhängig Die Festlegung eines Merkmals als abhängige oder unabhängige Variable basiert auf theoretischen Überlegungen und ist kontextspezifisch. Je nach Fragestellung kann zum Beispiel das politische Interesse eine abhängige oder eine unabhängige Variable sein. Bei der oberen For-

schungsfrage in Tabelle 3 wird ein Zusammenhang zwischen Bildung und politischem Interesse postuliert. Bildung (uV) hat einen vermuteten Einfluss auf das

Tabelle 3 Politisches Interesse als aV und als uV

unabhängige Variable		abhängige Variable
Bildung	→	politisches Interesse
politisches Interesse	→	Wahlteilnahme

politische Interesse (aV). Bei der unteren Forschungsfrage in Tabelle 3 steht die Wahlteilnahme (aV) im Vordergrund. Es werden Prädiktoren (uVs) gesucht, die einen Einfluss auf die aV ausüben. Eine mögliche unabhängige Variable könnte das politische Interesse sein.

Abhängige Variable im Blick behalten

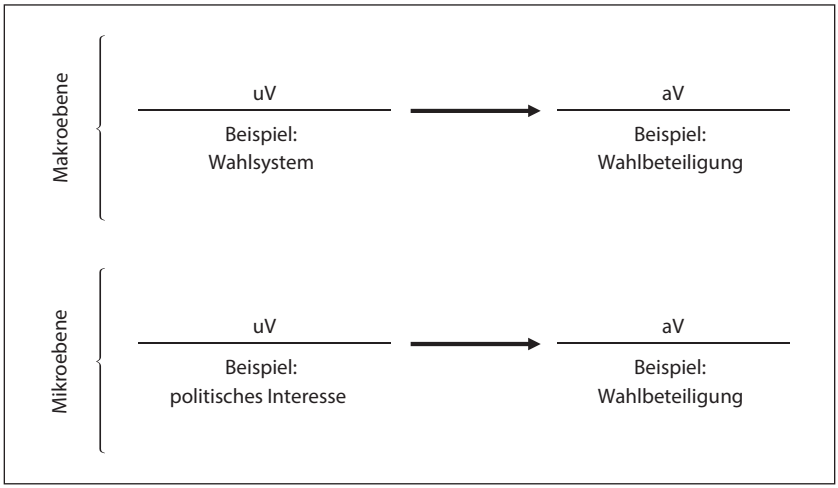
In einer wissenschaftlichen Arbeit ist es wichtig, die abhängige Variable im Blick zu behalten. Welcher Sachverhalt bzw. welche aV soll erklärt werden? Ein häufiger Fehler in Haus- und Abschlussarbeiten ist es, dass die aV plötzlich ausgetauscht wird. Zunächst wird die aV überzeugend dargestellt (z. B. Wahlbeteiligung), dann werden plötzlich Determinanten der Parteiwahl (Wahlverhalten) betrachtet.

2.4 Analyseebenen: Mikro und Makro

Der Unterscheidung zwischen aV und uV liegt eine vermutete Ursache-Wirkungs-Beziehung zugrunde, die Differenzierung zwischen Mikro- und Makroebene basiert auf einer Trennung der Analyseebenen. Bei der Mikroebene handelt es sich um die Individualebene, bei der Makroebene um die gesellschaftliche Ebene. Die Makroebene wird häufig auch als Aggregatebene bezeichnet.

An einem Beispiel lässt sich der Unterschied zwischen Mikro- und Makroebene illustrieren: In Abbildung 2 stellt die Wahlbeteiligung jeweils die abhängige Variable dar. Bei der Analyse der Wahlbeteiligung kann zwischen der individuellen

Abbildung 2 Mikro- und Makroebene als klassische Analyseebenen



Quelle: Eigene Darstellung

Wahlbeteiligung und der gesellschaftlichen Wahlbeteiligung unterschieden werden. Übersetzt in eine Forschungsfrage: Warum unterscheidet sich die Wahlbeteiligung zwischen den Bürgerinnen und Bürgern? Warum unterscheidet sich die Wahlbeteiligung zwischen Nationalstaaten? Bei der Betrachtung auf der Mikroebene (erste Forschungsfrage) werden Beziehungen zwischen Individualmerkmalen, bei der Analyse auf der Makroebene (zweite Forschungsfrage) werden Zusammenhänge zwischen Systemeigenschaften betrachtet.

Auf der Mikroebene werden individuelle Merkmale (uVs) gesucht, die einen Einfluss auf die individuelle Wahlbeteiligung (aV) haben. Einen solchen Faktor stellt beispielsweise das politische Interesse dar (uV). Bisherige Forschung konnte wiederholt einen positiven Zusammenhang zwischen politischem Interesse und Wahlbeteiligung nachweisen: Je stärker das politische Interesse, desto höher die Wahrscheinlichkeit der Beteiligung an Wahlen (z. B. Steinbrecher und Rattinger 2011; Westle et al. 2013; Caballero 2014; Lamers und Roßteutscher 2014).

Bei der Makroebene stehen Systemmerkmale im Mittelpunkt, die sich auf das Niveau der Wahlbeteiligung auswirken. Auf der Makroebene werden also nicht individuelle Merkmale betrachtet, sondern vom Individuum losgelöste Merkmale. Ein klassischer gesellschaftlicher Einflussfaktor der Wahlbeteiligung ist beispielsweise das (nationale) Wahlsystem. Geys (2006, S. 651) kann einen förderlichen Effekt des Verhältniswahlsystems (im Vergleich zum Mehrheitswahlsystem) auf die

Wahlbeteiligung belegen (siehe auch Blais 2006; Blais und Aarts 2006). In Ländern mit einem Verhältniswahlsystem ist die Wahlbeteiligung in der Regel höher als in Ländern mit einem Mehrheitswahlsystem.

Keine willkürliche Festlegung Die Differenzierung zwischen Mikro- und Makroebene dient der analytischen Präzision. Dadurch wird einerseits die Fragestellung präzisiert, andererseits kann die Darstellung des Forschungsstands systematischer aufbereitet werden. In einer Hausarbeit ist häufig die Betrachtung einer analytischen Ebene ausreichend, aber die Festlegung darf nicht willkürlich erfolgen. Die Analyse der Zusammenhänge zwischen Merkmalen auf der Individualebene bzw. auf der gesellschaftlichen Ebene ist jeweils mit Vor- und Nachteilen verbunden, die der Forscher reflektieren sollte (für die Wahlbeteiligung siehe z. B. Keil 2009, S. 424–425). Die Festlegung einer Analyseebene ist deshalb stets zu begründen (siehe auch Kapitel 9).

Die Makroebene bzw. die gesellschaftliche Ebene wird häufig mit der nationalstaatlichen Ebene gleichgesetzt. Bei der Wahlbeteiligung wird beispielsweise die unterschiedliche Wahlbeteiligung zwischen Nationalstaaten betrachtet. Entsprechend stellt der Nationalstaat die Analyseeinheit dar. Makroanalysen sind aber nicht nur auf den Nationalstaat begrenzt. Mit Blick auf die Bundesrepublik Deutschland stellen die 16 Bundesländer ebenfalls eine gesellschaftliche Ebene dar und selbstverständlich können auch die Unterschiede der Wahlbeteiligung zwischen Bundesländern analysiert werden (z. B. Vetter 2008; Völkl et al. 2008). Denkbar ist auch die Analyse von lokalen Unterschieden. Schließlich unterscheidet sich die Wahlbeteiligung nicht nur zwischen Staaten und Bundesländern, sondern auch zwischen Gemeinden (z. B. Faas 2013) und innerhalb von Gemeinden (z. B. Schmitt-Beck et al. 2008; Schäfer 2012; Roßteutscher und Schäfer 2016). Die Makroebene sollte deshalb nicht auf die nationalstaatliche Ebene reduziert werden, da Makroanalysen auch auf regionaler und lokaler Ebene eine interessante Forschungsperspektive bieten (z. B. Putnam 1993; van Deth und Tausendpfund 2013; Heinsohn und Freitag 2016). Die Festlegung für eine spezifische Ebene ist inhaltlich zu begründen.

Mesoebene In der Literatur wird neben der Makro- und Mikroebene häufig noch die Mesoebene unterschieden. Dabei handelt es sich um eine analytische Ebene zwischen Mikro- und Makroebene. Zwischen der (gesamt-)gesellschaftlichen Ebene und dem sozialen Handeln des Individuums stehen beispielsweise Verbände, Parteien oder etwas allgemeiner: Organisationen (ausführlicher siehe Esser 2000a, S. 59–62).

2.5 Qualitative und Quantitative Methoden

Wer sich mit den Methoden der empirischen Sozialforschung beschäftigt, der wird schnell mit der Unterscheidung in „Qualitative Methoden“ und „Quantitative Methoden“ konfrontiert. Diese Differenzierung beinhaltet kein Werturteil, deutet aber auf unterschiedliche Forschungstraditionen hin. Die beiden methodischen Zugänge werden gelegentlich als „Methoden-Schisma“ (Schnapp et al. 2006, S. 11) oder als „Gegensatz“ (Baur und Blasius 2014b, S. 42) gedeutet. Einen Überblicksartikel zu qualitativen und quantitativen Methoden beginnen John Gerring und Craig W. Thomas (2011, S. 2189) mit folgender Feststellung:

> *„Perhaps no division in the social sciences is so persistent, nettlesome, and poorly understood as the division between quantitative and qualitative ways of knowing.“*

Auch Schnapp et al. (2006, S. 11) verweisen auf emotional geführte Diskussionen über die „wirklich“ angemessenen Methoden der empirischen Sozialforschung:

> *„Erscheinen die Motive, aus denen diese Debatten geführt werden, nicht immer ausschließlich der Sorge um den wissenschaftlichen Fortschritt geschuldet und lässt sich an der Ergiebigkeit dieser Debatten oft mit guten Gründen zweifeln, so ist es sicherlich noch bedauerlicher, wenn diese Debatten erst gar nicht stattfinden und jede Seite für sich bleibt, ohne die andere zur Kenntnis zu nehmen. Denn fernab dieses leider oft ideologisch gefärbten Richtungsstreits erscheint es als Selbstverständlichkeit, dass die komplexe Beschreibung und Analyse eines beliebigen Gegenstandes die Einnahme mehrerer Beobachtungsperspektiven nötig macht.“ (Schnapp et al. 2006, S. 11)*

In diesem Verständnis sind „Qualitative Methoden“ und „Quantitative Methoden“ kein Gegensatz, sondern ergänzen sich. In den Sozialwissenschaften sollte es darum gehen, „Fragen zu beantworten – mit den für die jeweilige Fragestellung und den Gegenstandsbereich bestmöglich geeigneten Methoden. Und das können je nach Kontext entweder qualitative oder quantitative oder eine Kombination aus beiden Methoden sein“ (Baur und Blasius 2014b, S. 42). Deshalb sind grundlegende Kenntnisse beider methodischen Zugänge erforderlich, um in Abhängigkeit der konkreten Fragestellung die geeignete Methode auswählen zu können. Methoden sind immer nur ein Mittel zum Zweck und müssen stets in Abhängigkeit von der konkreten Forschungsfrage ausgewählt werden (Schnapp et al. 2006, S. 21).

Für ein besseres Verständnis der beiden Forschungstraditionen ist die Kenntnis zentraler Merkmale der qualitativen und quantitativen Methoden notwendig. In Tabelle 4 sind mit Forschungsziel, Forschungsprozess, Fallzahl, Forschungsdaten, Hypothesen, Forschungslogik, Auswertung und Generalisierung acht Ver-

Tabelle 4 Qualitative und Quantitative Methoden im Vergleich

Vergleichsdimension	Qualitative Methoden	Quantitative Methoden
Forschungsziel	„verstehen“	„erklären“
Forschungsprozess	zirkulär	linear
Fallzahl	wenige	viele
Forschungsdaten	Worte	Zahlen
Hypothesen	generierend	prüfend
Forschungslogik	Induktion	Deduktion
Auswertung	offen	statistische Verfahren
Generalisierung	gering	hoch

gleichsdimensionen aufgeführt. Auf Basis dieser Dimensionen werden die zentralen Unterschiede der qualitativen und quantitativen Methoden herausgearbeitet. Dabei werden die „Extrempositionen“ der beiden Forschungstraditionen betont. In der Forschungspraxis verschwinden gelegentlich die Grenzen zwischen den einzelnen Dimensionen (für Vergleiche der qualitativen und quantitativen Methode siehe auch Schnapp et al. 2006; Behnke et al. 2006, S. 42–47; Lamnek 2010, S. 215–244; Blatter et al. 2018).

Forschungsziel Das Ziel qualitativer Forschung ist das Verstehen spezifischer Handlungen und Praktiken der Menschen. In dieser Perspektive wird der subjektive Sinn, den Individuen mit ihren Handlungen und Praktiken verbinden, zum empirischen Ansatzpunkt (Flick 2010, S. 34). Qualitativ arbeitende Wissenschaftler fragen, wie dieser Sinn für die Menschen zustande gekommen ist, wie er ihr Handeln und ihre Entscheidungen prägt und sich verändert (Krumm 2009, S. 96–97). Dazu werden Situationen detailliert nachgezeichnet, soziale Interaktionen rekonstruiert und Regeln gesellschaftlichen Handelns offengelegt. Durch die Interpretation einer Handlung oder Situation – in der Regel auf Basis von Texten – wird auf die spezifische Sichtweise und Deutungsmuster der Akteure geschlossen. Im Mittelpunkt qualitativer Forschung steht das Verstehen des subjektiven Sinns der Akteure bei Handlungen.

Das Ziel quantitativer Forschung ist die Erklärung sozialwissenschaftlicher Phänomene. Quantitativ arbeitende Sozialforscherinnen orientieren sich dabei an den Naturwissenschaften, in denen allgemeingültige Gesetzmäßigkeiten gesucht werden, die Erklärungen und Prognosen ermöglichen. Entsprechend suchen Sozialwissenschaftler nach sozialen Regelmäßigkeiten, die soziale Phänomene erklä-

ren können. Dazu werden Annahmen („Hypothesen") über mögliche Zusammenhänge von Merkmalen formuliert, die anschließend empirisch überprüft werden. In Abgrenzung zu den Naturwissenschaften formulieren quantitativ arbeitende Sozialforscher allerdings keine deterministischen, sondern probabilistische Hypothesen (siehe Kapitel 6). Quantitative Forschung konzentriert sich auch nicht auf die Erklärung einer Beobachtung (z. B. die Wahlbeteiligung einer konkreten Person), sondern stellt den durchschnittlichen Effekt einer oder mehrerer unabhängigen Variablen auf die abhängige Variable (z. B. politisches Interesse auf die Wahlbeteiligung) bei vielen Beobachtungen (hier: Personen) in den Mittelpunkt.

Forschungsprozess In der qualitativen Sozialforschung ist der Forschungsprozess zirkulär angelegt. Die verschiedenen Phasen einer Untersuchung – Forschungsfrage, Datenerhebung und Datenauswertung – sind einerseits nicht klar voneinander abgegrenzt und andererseits existiert auch keine bestimmte Reihenfolge dieser Phasen (Przyborski und Wohlrab-Sahr 2014a, S. 118). Qualitative Forscher betonen die Offenheit des qualitativen Forschungsprozesses gegenüber dem Untersuchungsgegenstand, die die theoretische und methodische Ebene betrifft.

> *„Der Forschungsprozess muss so offen dem Gegenstand gegenüber gehalten werden, dass Neufassungen, Ergänzungen und Revisionen sowohl der theoretischen Strukturierungen und Hypothesen als auch der Methoden möglich sind, wenn der Gegenstand dies erfordert." (Mayring 2002, S. 28)*

Bei der Auseinandersetzung mit dem Forschungsfeld werden Fragestellung, Konzepte und Instrumente immer wieder überprüft und angepasst. Je offener das Verfahren, desto genauer muss allerdings jeder einzelne Verfahrensschritt expliziert und dokumentiert werden (Mayring 2002, S. 28).

In der quantitativen Sozialforschung ist der Forschungsprozess stärker linear organisiert. Ausgehend von einer Forschungsfrage werden zunächst Konzepte spezifiziert und Hypothesen formuliert, bevor Daten erhoben und ausgewertet werden. Auf Basis empirischer Daten werden schließlich Hypothesen überprüft. Wenn die Daten die Hypothesen stützen, dann gelten die Hypothesen vorläufig als bestätigt. Falls die Daten nicht mit den Hypothesen übereinstimmen, dann gelten die Hypothesen als widerlegt (falsifiziert). In der Forschungspraxis sind auch beim quantitativen Forschungsprozess „Überschneidungen, Sprünge und Rückkopplungen" (Kromrey et al. 2016, S. 69) zu beobachten, aber quantitative Forschungsdesigns sind in ihrer Grundstruktur linear-chronologisch aufgebaut (siehe Kapitel 3).

Fallzahl Qualitative Forscherinnen setzen sich mit dem Datenmaterial intensiv auseinander, um die Einzigartigkeit des Menschen und die Kontextgebundenheit sozialer Interaktionen zu verdeutlichen. Entsprechend werden möglichst viele empirische Informationen pro Fall (z. B. Personen) gesammelt, um soziale Interaktionen detailliert zu rekonstruieren. Die Sammlung, Aufbereitung und Analyse der Daten eines Falls ist sehr aufwändig, sodass innerhalb einer qualitativen Untersuchung häufig nur wenige oder auch nur ein Fall betrachtet wird.

Schwerpunktsetzung dieses Lehrbuchs: Quantitative Methoden

Das vorliegende Buch bietet eine Einführung in die quantitativen Methoden der empirischen Sozialforschung. Diese Festlegung hat pragmatische Gründe. Innerhalb der qualitativen und quantitativen Methoden existieren mittlerweile unzählige Ansätze, Strategien und Verfahren zur Entwicklung von Forschungsfragen, zur Sammlung von empirischen Informationen sowie zur Datenanalyse. Entsprechende Einführungen in die qualitativen bzw. quantitativen Methoden umfassen selten weniger als 500 Seiten (siehe für die qualitativen Methoden z. B. Flick 2010; Lamnek 2010; Heiser 2018; siehe für die quantitativen Methoden z. B. Diekmann 2011; Schnell et al. 2013). Eine angemessene Darstellung quantitativer und qualitativer Methoden innerhalb dieses Buchs „sprengt“ einerseits den Umfang dieser Einführung und birgt andererseits die Gefahr, die spezifischen Besonderheiten und Herausforderungen der beiden Forschungstraditionen verkürzt darzustellen.

Bei einer quantitativen Untersuchung werden die empirischen Informationen standardisiert erfasst, z. B. durch eine Befragung mit vorgegebenen Antwortmöglichkeiten. Dabei werden individuelle Besonderheiten bei der Datensammlung systematisch ausgeblendet. Die Befragte kann nicht mit „eigenen Worten“ antworten, sondern wählt eine vorgegebene Antwort aus. Dadurch ist die Informationsdichte für den einzelnen Fall bei einer quantitativen Datenerhebung zwar geringer als bei einer qualitativen Datenerhebung, dafür werden empirische Informationen von vielen Fällen (z. B. Personen) gesammelt. Die konkrete Anzahl der Fälle variiert in Abhängigkeit von der Forschungsfrage. Bei einer quantitativen Analyse auf Makroebene (z. B. Staaten) liegt die Fallzahl meist im mittleren zweistelligen Bereich, bei einer Analyse auf Individualebene (z. B. Personen) werden häufig mehrere Tausend Fälle untersucht.

Forschungsdaten Qualitative Sozialforscher erheben ihre empirischen Daten in der Regel offen – mit Hilfe nicht standardisierter Interviews, wenig strukturierter Beobachtungen oder Erzählungen. Die empirischen Informationen liegen als Texte vor, etwa in Form von Beschreibungen oder Protokollen. Konzepte und Merkmale werden in Worten erfasst; die verbale Form der Datensammlung ermöglicht eine vielfältige und detaillierte Datenerfassung (siehe für einen Überblick z. B. Flick et al. 2013).

In der quantitativen Sozialforschung dominieren standardisierte Verfahren der Datenerhebung (z. B. Befragung). Die gesammelten empirischen Informationen werden in Zahlen „übersetzt". Bei einigen Merkmalen, beispielsweise Alter oder Einkommen der Befragten, ist das relativ einfach. Quantitative Forschung erfasst jedoch auch komplexere soziale Merkmale (z. B. das politische Interesse oder das soziale Vertrauen) in Zahlen. Die Zuordnung von Zahlen zu Objekten wird in der empirischen Sozialforschung als Messung bezeichnet (Stevens 1946). Diese Zahlen bilden die Grundlage für die spätere (statistische) Datenanalyse.

Hypothesen Bei Hypothesen handelt es sich – vereinfacht formuliert – um (plausible) Vermutungen über den Zusammenhang zwischen zwei Merkmalen (siehe Kapitel 6). Ein Kennzeichen qualitativer Forschung ist es, Hypothesen zu Beginn des Forschungsprozesses (wenn überhaupt) noch sehr allgemein zu formulieren, um neue Aspekte, die bei der Auseinandersetzung mit dem Untersuchungsgegenstand sichtbar werden, noch berücksichtigen zu können. Hypothesen werden deshalb in Auseinandersetzung mit dem Untersuchungsgegenstand formuliert und überprüft.

> *„Im Unterschied zur quantitativen Auffassung ist die Hypothesenentwicklung in der qualitativen Sozialforschung ein konstitutives Moment des Forschungsprozesses. Kennzeichnend ist der offene Charakter der theoretischen Konzepte, d. h. der ständige Austausch zwischen den qualitativ erhobenen Daten und dem zunächst noch vagen theoretischen Vorverständnis, sodass es zu einer fortwährenden Präzisierung, Modifizierung und Revision von Theorien und Hypothesen kommt." (Lamnek 2010, S. 80)*

Bei einem quantitativen Untersuchungsdesign werden Hypothesen explizit vor der Datenerhebung und der Datenanalyse formuliert. Auf Basis theoretischer Überlegungen und existierender empirischer Befunde (Forschungsstand) werden allgemeine Zusammenhänge zwischen Merkmalen (Hypothesen) formuliert. Die Hypothesen werden anschließend überprüft. Auf Basis empirischer Daten wird beispielsweise untersucht, ob sich ein Zusammenhang zwischen zwei Merkmalen (z. B. politisches Interesse und Wahlbeteiligung) nachweisen lässt oder nicht. Die empirische Prüfung führt zu einer (vorläufigen) Bestätigung oder Widerlegung

einer Hypothese. Quantitative Forschung geht damit hypothesenprüfend, qualitative Forschung hypothesengenerierend vor.

Forschungslogik Qualitative Sozialforscher folgen einer induktiven Forschungslogik. Bei der Induktion wird von Einzelfällen (konkrete Beobachtung) auf eine allgemeine Regel („Gesetz") geschlossen. Auf Basis des Einzelfalls bzw. weniger Fälle formulieren qualitativ arbeitende Sozialforscher allgemeine Aussagen. Die wiederholte Beobachtung des gleichen Falls oder verschiedener Fälle erhöht zwar die Wahrscheinlichkeit einer allgemeinen Regel, aber die induktive Methode kann eine allgemeine Regel nicht abschließend belegen (Krumm 2009, S. 102). Dazu müssten „alle" Fälle untersucht werden.

Quantitative Sozialforscherinnen orientieren sich an der deduktiven Forschungslogik. Deduktion meint den Schluss vom Allgemeinen auf das Besondere. Aus allgemeinen Theorien werden empirisch überprüfbare Hypothesen abgeleitet. Diese Hypothesen werden dann am „Einzelfall" überprüft. Ein Beispiel: Die allgemeine Hypothese „Je stärker das politische Interesse, desto wahrscheinlicher die Beteiligung an einer Wahl" bezieht sich nicht auf ein konkretes Land oder eine konkrete Wahl. Entsprechend sollte der Zusammenhang zwischen politischem Interesse und Wahlbeteiligung sowohl bei den Wahlen zum amerikanischen Kongress als auch bei der Oberbürgermeisterwahl in Hagen nachweisbar sein. In einem quantitativen Forschungsprojekt wird die allgemeine Hypothese überprüft (z. B. Zusammenhang zwischen politischem Interesse und Wahlbeteiligung bei der Bundestagswahl).[3]

Auswertung Qualitative Sozialforscher nutzen Auswertungsverfahren, die der jeweiligen Forschungsfrage sowie der offenen Struktur der Daten angemessen sind. Przyborski und Wohlrab-Sahr (2014b, S. 189) unterscheiden vier verschiedene Auswertungsverfahren, die in der qualitativen Sozialforschung eine wichtige Rolle spielen: die Methodologie der Grounded Theory, das Verfahren der Narrationsanalyse, die objektive Hermeneutik und die dokumentarische Methode. Im Handbuch von Flick et al. (2013) werden weitere Auswertungsverfahren der qualitativen Forschung dargestellt (siehe auch Blatter et al. 2018).

Bei einer quantitativen Datenerhebung werden den empirischen Informationen Zahlen zugeordnet. Ein Beispiel: Bei einer Befragung werden die Personen nach ihrem politischen Interesse gefragt. Als Antwort können die Befragten zwischen den Alternativen „sehr interessiert", „ziemlich interessiert", „wenig interes-

3 Neben Induktion und Deduktion existiert mit der Abduktion ein drittes Schlussverfahren, das zu einer gegebenen Beobachtung eine mögliche allgemeine Regel („Gesetz") sucht, die den beobachteten Fall erklären könnte (Sturm 2006).

siert“ und „überhaupt nicht interessiert“ wählen. Die einzelnen Antwortmöglichkeiten sind mit Zahlen verknüpft, z. B. 0 für „überhaupt nicht interessiert“ und 3 für „sehr interessiert“. Die numerische Kodierung der verbalen Antworten erlaubt quantitativen Sozialforschern die empirischen Informationen mit Hilfe statistischer Verfahren auszuwerten. Die Regression bzw. die Regressionsanalyse ist dabei das zentrale Analyseverfahren in den Sozialwissenschaften (z. B. Urban und Mayerl 2011). Eine Regression zielt darauf ab, die Werte eines oder mehrerer Merkmale (unabhängige Variablen) zur Erklärung eines anderen Merkmals (abhängige Variable) heranzuziehen. Beispiel: Es werden die Merkmale Alter, Bildung und Berufserfahrung (uVs) genutzt, um das Einkommen (aV) zu „erklären“. Einen Überblick der wichtigsten quantitativen Auswertungsverfahren bietet das Handbuch der sozialwissenschaftlichen Datenanalyse von Wolf und Best (2010a).

Generalisierung Wissenschaftliche Forschung ist an Generalisierung interessiert (Verallgemeinerung). Generalisierung setzt die Abstraktion vom Einzelfall voraus.

> *„Nur durch empirische Arbeit, die über den einzelnen Fall hinaus zu theoretischer Abstraktion kommt, ist wissenschaftlicher Fortschritt möglich. Nur so platziert man sich mit der eigenen Forschung in einem wissenschaftlichen Feld und gibt anderen die Möglichkeit, daran anzuschließen oder sich kritisch damit auseinanderzusetzen.“ (Przyborski und Wohlrab-Sahr 2014a, S. 127)*

Bei der Generalisierung sind qualitative Sozialforscher mit zwei Herausforderungen konfrontiert: Zum einen arbeiten qualitative Forscherinnen in der Regel mit kleinen Fallzahlen – manchmal sogar nur mit Einzelfällen. Diese Einzelfälle werden zwar sehr intensiv analysiert, erlauben allerdings nicht unmittelbar den Schluss vom Besonderen auf das Allgemeine (Induktion). Zum anderen wählen qualitative Sozialforscher ihre Untersuchungsfälle meistens gezielt aus („Theoretical Sampling“). Eine gezielte Auswahl kann immer auch willkürlich sein, sodass möglicherweise eine selektive, das heißt eine verzerrte Auswahl der Untersuchungsfälle vorliegt (Lamnek 2010, S. 236). Deshalb hat diese Forschungstradition spezifische Probleme bei der Generalisierung der Ergebnisse (Mayring 2002, S. 24). Eine Übersicht verschiedener Lösungsstrategien bieten Przyborski und Wohlrab-Sahr (2014a).

Quantitative Sozialforscherinnen orientieren sich bei der Fallauswahl häufig an der Zufallsstichprobe (Kapitel 10). Auf Grundlage der Inferenzstatistik[4] kön-

4 Die Inferenzstatistik (häufig auch als schließende oder induktive Statistik bezeichnet) beschäftigt sich mit der Übertragbarkeit von empirischen Befunden von Zufallsstichproben auf die zugehörige Grundgesamtheit (siehe Kapitel 10).

nen die Befunde von Zufallsstichproben innerhalb bestimmter Fehlergrenzen auf die Grundgesamtheit übertragen werden.

Gemeinsamkeiten Bei dem Vergleich der qualitativen und quantitativen Vorgehensweise wurden insbesondere die Unterschiede der beiden Forschungstraditionen dargestellt. Bei allen Unterschieden machen Schnapp et al. (2006, S. 17) explizit auf zwei Gemeinsamkeiten der beiden Forschungstraditionen aufmerksam:

- Beide Forschungstraditionen arbeiten mit empirischen Informationen.
- Qualitative und quantitative Analysen haben das Ziel, Erkenntnisse über gesellschaftliche Strukturen und Handlungen zu erlangen, die über den konkreten Einzelfall hinausweisen.

Darüber hinaus stellen Schnapp et al. (2006, S. 18) auch einen Konsens über Qualitätsanforderungen fest, denen „gute" qualitative und quantitative Forschung genügen muss:

- Forschungsergebnisse müssen intersubjektiv nachvollziehbar sein. Aus den Ergebnissen und der Dokumentation des Forschungsprozesses muss für Dritte erkennbar sein, wie die Forschenden zu ihren Resultaten gekommen sind.
- Generalisierungen sollten auf einer tiefgehenden Kenntnis des Forschungsgegenstands beruhen.
- Forschende müssen die Ergebnisse der eigenen Forschung kritisch reflektieren.

Mixed Methods Wie oben angedeutet, ist bei vielen Forschungsfragen sowohl eine qualitative als auch eine quantitative Vorgehensweise denkbar. In jüngerer Zeit ist eine verstärkte Verknüpfung qualitativer und quantitativer Methoden zu beobachten. Die Kombination beider Forschungsmethoden in einem Forschungsprojekt wird als „Mixed Methods" bezeichnet (z. B. Kelle 2014, S. 153; siehe auch Tashakkori und Teddlie 2010; Kuckartz 2014). Die Kombination qualitativer und quantitativer Methoden zur Beantwortung einer Forschungsfrage ist allerdings keine völlig neue „Erfindung". Bereits die klassische Studie „Die Arbeitslosen von Marienthal" (Jahoda et al. 1975 [1933]) verknüpft gewinnbringend qualitative und quantitative Methoden, um die Konsequenzen von Arbeitslosigkeit empirisch zu untersuchen.

In der Forschungspraxis ist der Übergang zwischen qualitativer und quantitativer Sozialforschung fließend. Beide methodischen Zugänge haben ihre spezifischen Vor- und Nachteile. Es ist daher nicht zielführend, pauschal eine Forschungstradition als die „bessere" zu bewerten.

> *„Vielmehr sollte man sich abhängig von der jeweiligen Forschungsfrage entscheiden, ob man eine rein qualitative bzw. eine rein quantitative Strategie verfolgt oder ob man qualitative und quantitative Forschung kombiniert." (Behnke et al. 2010, S. 44)*

Bei der Kombination von qualitativen und quantitativen Methoden lassen sich zwei allgemeine Strategien unterscheiden: Zum einen können qualitative und quantitative Forschung aufeinander folgen, und zum anderen können beide Strategien auch komplementär zueinander sein. Zwar ist eine Zunahme von „Mixed Methods"-Forschungsstrategien zu beobachten, der Großteil der Forschung lässt sich allerdings eindeutig als qualitative oder quantitative Forschung charakterisieren.

Klassiker der Sozialwissenschaften: „Die Arbeitslosen von Marienthal"

In der Studie „Die Arbeitslosen von Marienthal" (Jahoda et al. 1975 [1933]) untersuchten Marie Jahoda, Paul F. Lazarsfeld und Hans Zeisel die Auswirkungen langandauernder Arbeitslosigkeit. Die Studie wurde 1931/32 im österreichischen Arbeiterdorf Marienthal durchgeführt. Dort schloss 1929 eine Fabrik, die bis dahin fasst die Gesamtheit der erwerbsfähigen Bevölkerung beschäftigt hatte. Von den 478 Familien des Orts waren plötzlich 367 arbeitslos (kein Familienmitglied war in Arbeit). Etwa 75 Prozent der Familien waren auf Arbeitslosenunterstützung angewiesen (Jahoda et al. 1975 [1933], S. 38–39). Der zentrale Befund der Marienthal-Studie ist, dass lange Arbeitslosigkeit zur „Reduktion des Anspruchs- und Aktivitätsbereichs" und damit zur Resignation führt (Jahoda et al. 1975 [1933], S. 25). Die Dorfbewohner bildeten eine „müde Gemeinschaft" (S. 55). Bei der Untersuchung wurde die „Lücke zwischen den nackten Ziffern der Statistik und den zufälligen Eindrücken der sozialen Reportage" (Jahoda et al. 1975 [1933], S. 15) mit einem großen Methodeninstrumentarium gefüllt. Die Forscher nutzten unter anderem standardisierte Befragungen und Beobachtungen, analysierten Schulaufsätze, führten informelle Gespräche und werteten vorhandenes Datenmaterial (z. B. Bibliothekstatistiken) aus, um die Forschungsfrage zu beantworten (Jahoda et al. 1975 [1933], S. 26–27).

2.6 Erklärungen

Ein zentrales Anliegen der Sozialwissenschaften ist die Erklärung sozialer Sachverhalte. Zwei Grundmodelle allgemeiner Erklärungen werden an dieser Stelle unterschieden: Das deduktiv-nomologische Modell von Hempel und Oppenheim (1948) und das induktiv-statistische Modell (Hempel 1977, S. 55–123; siehe auch Schurz 2009).

Deduktiv-Nomologisches Modell Als Grundmodell einer Erklärung gilt in der quantitativen Sozialforschung das deduktiv-nomologische Erklärungsmodell, das Hempel und Oppenheim (1948) im Anschluss an die Arbeiten von Popper (1982 [1934]) entwickelt haben. Das D-N-Modell (nach deduktiv-nomologisch) oder auch H-O-Modell (nach Hempel-Oppenheim) besteht aus zwei Elementen: Explanandum und Explanans.

Das Explanandum (übersetzt: „das zu Erklärende") ist das zu erklärende Phänomen. Nach Esser (1999b, S. 40) beginnt jede Erklärung mit einer Aussage, die den zu erklärenden Sachverhalt korrekt beschreibt. Eine solche Aussage enthält meist zeitlich-räumliche Angaben über das Auftreten des Sachverhalts (z. B. in Deutschland ist die Beteiligung bei Bundestagswahlen von 1998 bis 2009 gesunken). Eine korrekte Aussage über einen Sachverhalt, dessen Auftreten erklärt werden soll, wird Explanandum genannt.

! Die Erklärung des Explanandums erfolgt über allgemeine Gesetze (auch Allaussagen genannt) und Randbedingungen (auch Antecedensbedingungen genannt). Gesetz und Randbedingung werden Explanans (übersetzt: „das Erklärende") genannt.

Das allgemeine Gesetz benennt eine Beziehung zwischen zwei Sachverhalten und ist als Allaussage formuliert. Das heißt, die Beziehung ist an allen Orten und zu allen Zeiten gültig. Die Randbedingungen beziehen sich dagegen auf eine kon-

Deduktiv-Nomologische Erklärung

Eine D-N-Erklärung besteht aus zwei Arten von Sätzen, die erst gemeinsam eine Erklärung liefern: Allgemeine Sätze (Gesetze) und besondere Sätze (Randbedingungen), die nur für die konkrete Situation gelten. Aus den allgemeinen Sätzen wird mit Hilfe der Randbedingungen das Explanandum deduziert (bzw. abgeleitet).

krete räumlich-zeitliche Situation. Nach dem D-N-Modell erfolgt eine korrekte Erklärung über eine logische Deduktion (Ableitung) aus einem allgemeingültigen Gesetz und der Kontrolle des Vorliegens der Randbedingung (Schnell et al. 2013, S. 54).

In Tabelle 5 wird die Vorgehensweise bei einer D-N-Erklärung mit einem klassischen Beispiel nach Popper (1982 [1934], S. 31–32) illustriert. Erklärt wird, warum ein konkreter Faden reißt (Explanandum). Als Explanans werden ein allgemeines Gesetz und Randbedingungen formuliert. Das Gesetz beinhaltet eine Wenn-Dann-Komponente („Wenn ein Faden mit einer Last von einer gewissen Mindestgröße belastet wird, dann reißt er") und ist allgemein formuliert. Das Gesetz ist zeitlich-räumlich nicht eingeschränkt. Das allgemeine Gesetz wird durch zwei Randbedingungen ergänzt. Erstens beträgt die maximale Belastungsgröße für den vorliegenden Faden ein Kilogramm und zweitens beträgt das angehängte Gewicht am Faden zwei Kilogramm. Falls die beiden Randbedingungen empirisch korrekt sind, dann reißt der Faden.

Tabelle 5 Struktur und Bestandteile einer D-N-Erklärung

Allgemeine Struktur einer D-N-Erklärung		**Beispiel nach Popper (1982, S. 31–32)**
Explanans	G: Gesetz (Allaussage)	G: Wenn ein Faden mit einer Last von einer gewissen Mindestgröße belastet wird, dann reißt er.
	R: Randbedingung	R1: Für diesen Faden beträgt die maximale Belastungsgröße ein Kilogramm. R2: Das an diesem Faden angehängte Gewicht beträgt zwei Kilogramm.
Explanandum	Phänomen	Dieser Faden reißt.

Während das Gesetz zeitlich-räumlich unbeschränkten Geltungsanspruch erfordert, beziehen sich die Randbedingungen auf eine konkrete Situation. Die Randbedingungen bezeichnen einzelne Faktoren, die sich auf die Wenn-Komponente beziehen. Sind die Bedingungen erfüllt und das Gesetz korrekt, dann ist das Explanandum zu beobachten.

Hempel und Oppenheim (1948, S. 137–140) nennen vier Voraussetzungen (Adäquatheitsbedingungen), die bei einer angemessenen Erklärung erfüllt sein müssen: Erstens muss das Explanandum logisch korrekt aus dem Explanans abgeleitet werden. Zweitens muss das Explanans ein Gesetz enthalten. Drittens muss das Explanans wahr sein. Viertens muss das Explanans empirischen Gehalt besit-

zen, das heißt Gesetz und Randbedingungen müssen empirisch überprüfbar sein. Esser (1999b, S. 43) erweitert diese Liste um eine weitere Voraussetzung: Das Explanandum muss empirisch wahr sein.

Induktiv-Statistisches Modell Eine deduktiv-nomologische Erklärung setzt ein allgemeingültiges Gesetz voraus, das zeitlich und räumlich nicht eingeschränkt ist (Allaussage). Allerdings sind in den Sozialwissenschaften solche Gesetze (bisher) nicht bekannt (Schnell et al. 2013, S. 61; Kromrey et al. 2016, S. 40). Deshalb wird in den Sozialwissenschaften eine modifizierte Version der klassischen D-N-Erklärung genutzt, in der nicht von strikten Gesetzen, sondern nur von Wahrscheinlichkeiten gesprochen wird (Hempel 1977, S. 55–123). Dieses Modell wird induktiv-statistische (I-S-Modell) oder probabilistische Erklärung genannt.

! Die Elemente einer I-S-Erklärung sind mit den Komponenten einer D-N-Erklärung vergleichbar, jedoch wird statt eines allgemeingültigen Gesetzes ein probabilistisches Gesetz verwendet. Der Unterschied zum D-N-Modell besteht nun darin, dass das Explanandum nicht deduktiv abgeleitet werden kann, sondern nur mit annähernder oder mit großer Wahrscheinlichkeit vorhergesagt wird, da im Gegensatz zu deduktiven Ableitungen im Fall einer I-S-Erklärung das Explanandum auch bei wahrem Explanans falsch sein kann.

In Tabelle 6 wird die allgemeine Struktur einer I-S-Erklärung dargestellt und an einem Beispiel aus der Wahlforschung illustriert. Im Beispiel ist das probabilistische Gesetz die Aussage, dass Bürger, die sich sehr für Politik interessieren, sehr wahrscheinlich auch wählen gehen. Aus diesem probabilistischen Gesetz kann das Verhalten von Max Mustermann, der sich sehr für Politik interessiert, aber nicht

Tabelle 6 Struktur und Bestandteile einer I-S-Erklärung

Allgemeine Struktur einer I-S-Erklärung		Beispiel
Explanans	PG: Probabilistisches Gesetz	PG: Die Wahrscheinlichkeit, dass sich Bürger, die sich sehr für Politik interessieren, an Wahlen beteiligen, ist hoch.
	R: Randbedingung	R: Max Mustermann interessiert sich sehr für Politik.
Explanandum	Phänomen	Max Mustermann beteiligt sich wahrscheinlich an Wahlen.

deduziert werden. Man kann zwar vermuten, dass Max Mustermann sich an Wahlen beteiligt, aber das statistische Gesetz lässt auch die Möglichkeit zu, dass er sich nicht an Wahlen beteiligt. Je höher die Wahrscheinlichkeit des probabilistischen Gesetzes ist, desto wahrscheinlicher ist zwar das Explanandum, aber es ist weder logisch ableitbar noch mit Sicherheit prognostizierbar (Hempel 1977, S. 70).

Im Vergleich zu D-N-Erklärungen weisen I-S-Erklärungen deutliche Mängel auf, da das Explanandum nicht aus dem (allgemeingültigen) Gesetz und den Randbedingungen abgeleitet werden kann. Statt von I-S-Erklärungen wird in der Literatur daher häufig von I-S-Begründungen gesprochen (Hempel 1977, S. 98–123; Schnell et al. 2013, S. 63).

2.7 Kritischer Rationalismus

Wie funktioniert Wissenschaft? Welche Aussagen sind zulässig? Welche Vorgehensweise ist angebracht? Diese Fragen beantwortet die Wissenschaftstheorie (zur Einführung siehe z. B. Chalmers 2006; Schülein und Reitze 2012; Schurz 2014). Gehring und Weins (2009, S. 2) beschreiben Wissenschaftstheorien als „*Aussagenbündel darüber, was* Wissenschaft ist und wie diese vorzugehen hat“ (Hervorhebung im Original). Eine Wissenschaftstheorie ist keine Theorie über einen Ausschnitt der Realität (z. B. Wahlverhalten), sondern sie gibt allgemeine Prinzipien vor, wie die inhaltliche Theorie (z. B. Erklärung des Wahlverhaltens) zu formulieren ist und die methodische Vorgehensweise bei der Untersuchung aussehen sollte. Deshalb nennen Gehring und Weins (2009, S. 2) Wissenschaftstheorien auch Theorien über Theorien, die auch als Metatheorien bezeichnet werden (siehe auch Esser et al. 1977, S. 13). Dieser Abschnitt möchte mit wichtigen wissenschaftstheoretischen Grundlagen der quantitativen Sozialforschung vertraut machen. Eine anschauliche Beschreibung des Anliegens einer Wissenschaftstheorie findet sich bei Häder (2015, S. 19):

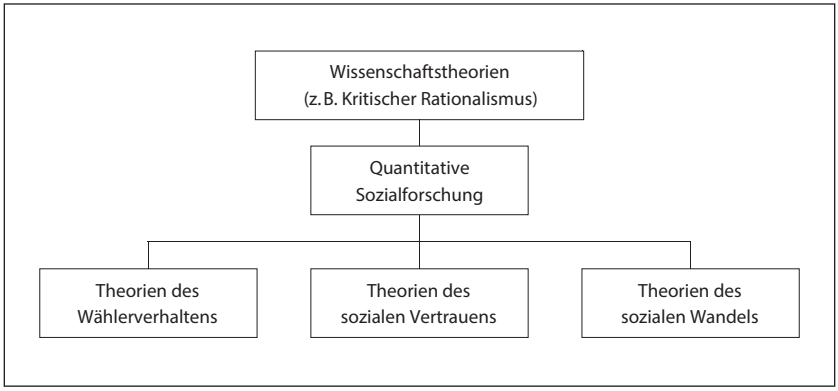
> *„Die Wissenschaftstheorie beschäftigt sich […] mit der Logik des Forschens, sie arbeitet die Spielregeln aus, an welchen sich Personen zu orientieren haben, die wissenschaftlich tätig sind. Die Wissenschaftstheorie geht damit über das einfache Festlegen von Rezepten für das Handeln hinaus: sie begründet diese Regeln auch. Sie sagt nicht nur, auf welche Art und Weise wissenschaftlich geforscht werden soll, sondern auch, weshalb dies so am besten zu erfolgen hat. Damit ist Wissenschaftstheorie teilweise mit einem guten Kochbuch vergleichbar. Hier wird nicht nur gesagt, wie man ein Essen zubereitet sondern auch, weshalb bestimmte Prozeduren so zu befolgen sind, wie sie dort beschrieben wurden, etwa um die Vitamine zu schonen oder um den natürlichen Geschmack aller Zutaten zu unterstreichen.“*

Es gibt nicht eine – allgemein akzeptierte – Wissenschaftstheorie, sondern viele Wissenschaftstheorien (z. B. Induktivismus, Positivismus, Konstruktivismus). Die (quantitative) Sozialforschung orientiert sich am kritischen Rationalismus, eine von Karl Popper (1969, 1982 [1934]) begründete Wissenschaftstheorie. Viele methodische Prinzipien und Vorgehensweisen der quantitativen Sozialforschung sind mit der wissenschaftstheoretischen Position von Popper verknüpft. In der Forschungspraxis werden diese allgemein formulierten Prinzipien häufig nicht (mehr) hinterfragt bzw. thematisiert; sie fungieren vielmehr als Postulat für eine bestimmte methodische Denk- und Vorgehensweise.

In Abbildung 3 ist das Verhältnis von Wissenschaftstheorie (hier: Kritischer Rationalismus) und empirischer Sozialforschung dargestellt. Die Wissenschaftstheorie formuliert allgemeine Regeln, wie das wissenschaftliche „Spiel" der Beschreibung, Erklärung und Prognose empirischer Sachverhalte funktioniert. Eine bestimmte Wissenschaftstheorie entspricht etwa der Verfassung eines Landes, die grundlegende Prinzipien der Staatsorganisation festschreibt. Sie steckt den Rahmen ab, in dem die Akteure handeln können. Die wissenschaftstheoretischen Prinzipien bilden die Grundlage für die methodische Vorgehensweise bei der Untersuchung der inhaltlichen Fragestellungen. Auf Basis dieser Prinzipien werden Hypothesen und Theorien entwickelt, die empirisch überprüft werden.

Mit Blick auf die quantitative Sozialforschung formuliert der kritische Rationalismus zwei zentrale Prinzipien: das Falsifikationsprinzip und das Basissatzproblem.

Abbildung 3 Verhältnis von Wissenschaftstheorie und Sozialforschung



Quelle: Eigene Darstellung

Falsifikationsprinzip Die Beschreibung und die Erklärung sozialer Sachverhalte ist ein zentrales Ziel empirischer Sozialforschung. Ein zentrales Prinzip des kritischen Rationalismus fordert, dass solche Beschreibungen und Erklärungen (im Folgenden: Aussagen) grundsätzlich an der Realität überprüfbar sein müssen und prinzipiell auch scheitern können (also evtl. nicht korrekt sind).

Aussagen wie „Wenn der Hahn kräht auf dem Mist, ändert sich das Wetter oder es bleibt, wie es ist" lehnt ein Anhänger des kritischen Rationalismus ab, da es sich um eine sogenannte analytische Aussage handelt (Tautologie). Der Satz ist immer wahr, da er alle Möglichkeiten einschließt, die jemals eintreten können. Deshalb kann ein solcher Satz nicht an der Realität scheitern. Auch müssen die Begriffe, die in Aussagen verwendet werden, einen empirischen Bezug haben. Aussagen wie „Der Fußballgott muss ein Münchner sein" oder „Gute Menschen kommen in den Himmel" haben keinen empirischen Gehalt, da die verwendeten Begriffe „Fußballgott" und „Himmel" sich nicht auf die erfahrbare Realität beziehen (für weitere Beispiele siehe z. B. Chalmers 2006, S. 51–62).

! Popper (1982 [1934], S. 3–6) vertritt im kritischen Rationalismus zudem den Standpunkt, dass die endgültige Bestätigung von Aussagen (Verifikation) nicht möglich ist. Ein Beispiel: Die Aussage „Alle Raben sind schwarz" kann nicht endgültig bestätigt werden, da niemals die Federnfarben aller Raben der Vergangenheit, Gegenwart und Zukunft untersucht werden können. Allerdings kann die Aussage falsifiziert (widerlegt) werden. Dazu muss „nur" ein weißer (oder auch gelber) Rabe beobachtet werden. Gemäß dieser Argumentation haben auch bestätigte Aussagen nur einen vorläufigen Charakter. Aussagen, die auch wiederholten und verschärften empirischen Prüfungen standgehalten haben, gelten als bewährte Aussagen (Gesetze). Da allerdings die endgültige Verifikation von Aussagen nicht möglich ist, steht im kritischen Rationalismus die Falsifikation von Aussagen im Vordergrund (Kromrey et al. 2016, S. 35–39).

Basissatzproblem Die in einem Forschungsprojekt formulierten Aussagen werden überprüft und müssen grundsätzlich an der Realität scheitern können. Eine Aussage, die der Überprüfung nicht standhält, ist widerlegt (falsifiziert). Die Falsifikation (Widerlegung) einer Aussage ist allerdings mit einem Problem konfrontiert, das Popper als Basissatzproblem bezeichnet hat (Popper 1982 [1934], S. 60–76). Danach kann eine Aussage nicht direkt durch Beobachtungen geprüft werden, sondern nur anhand einer Aussage über eine Beobachtung. Diese Beobachtungsaussage kann allerdings fehlerhaft sein (Gehring und Weins 2009, S. 8). Zur Lösung des Basissatzproblems schlägt Popper vor, die Beobachtungssätze durch Beschluss anzuerkennen und von ihrer vorläufigen Gültigkeit auszugehen:

> *„Die Basissätze werden durch Beschluss, durch Konvention anerkannt, sie sind Festsetzungen." (Popper 1982 [1934], S. 71)*

Diese Festsetzungen sind allerdings keine individuelle oder gar willkürliche Angelegenheit. Basissätze müssen sich nach Esser et al. (1977, S. 149) auf intersubjektiv beobachtbare Sachverhalte beziehen, dürfen nicht mit bereits anerkannten Basissätzen in Widerspruch stehen und müssen unter Berücksichtigung des gegenwärtig höchstmöglichen methodischen Wissens einer wissenschaftlichen Disziplin formuliert werden (Schnell et al. 2013, S. 75–80; siehe auch Kromrey et al. 2016, S. 42–46).

Für die Forschungspraxis leiten sich daraus zwei zentrale Prinzipien ab: Erstens muss der komplette Forschungsprozess intersubjektiv überprüfbar und nachvollziehbar sein. Die Vorgehensweise in einem Forschungsprozess muss dokumentiert sein sowie jede Entscheidung explizit gemacht und begründet werden, damit andere Forscher das Vorgehen zumindest gedanklich nachvollziehen können, um mögliche Fehler entdecken zu können. Zweitens sind bei der Durchführung des Forschungsvorhabens (z. B. bei einer Befragung) die höchstmöglichen methodischen Standards einzuhalten.

Fazit Aus dem kritischen Rationalismus leiten sich allgemeine Prinzipien für die Formulierung von Theorien und die Vorgehensweise bei der Untersuchung empirischer Fragestellungen ab. Zwei zentrale Grundsätze sind die intersubjektive Nachvollziehbarkeit und das Falsifikationsprinzip. Diese bilden das Fundament für die quantitative Sozialforschung.

3 Sozialwissenschaftlicher Forschungsprozess

In diesem Kapitel wird der (quantitative) sozialwissenschaftliche Forschungsprozess vorgestellt. Nach der Darstellung verschiedener Arten von Forschungsprojekten werden die einzelnen Phasen des Forschungsprozesses skizziert. Dieses erste Kennenlernen des (kompletten) Forschungsprozesses soll helfen, die einzelnen Phasen des Forschungsprozesses, die in den weiteren Kapiteln des Buchs vorgestellt werden, immer wieder in den Gesamtzusammenhang einzuordnen. In jeder Phase eines Forschungsprojekts werden Entscheidungen getroffen, die Konsequenzen für nachgelagerte Forschungsphasen haben.

3.1 Arten von Forschungsprojekten

Zur allgemeinen Charakterisierung von Forschungsprojekten finden sich in der Literatur zwei gängige Unterscheidungen: Zum einen findet sich die Kategorisierung zwischen Auftragsforschung und selbst initiierten Forschungsprojekten (z. B. Schnell et al. 2013, S. 3), zum anderen findet sich die Differenzierung zwischen Grundlagenforschung und anwendungsorientierter Forschung (Häder 2015, S. 71; Kromrey et al. 2016, S. 19–21).

Auftragsforschung Bei Auftragsforschung sind Thema und ggf. auch die Forschungsfrage mehr oder weniger vom Auftraggeber vorgegeben. Auftraggeber sozialwissenschaftlicher Forschungsprojekte können Ministerien, Behörden, Gemeinden, Verbände, Stiftungen und Parteien oder auch die Privatwirtschaft sein. Ein Auftraggeber wendet sich mit einem mehr oder weniger klar umrissenen Forschungsthema direkt an einen Forscher oder eine Institution. Gelegentlich werden solche Projekte auch ausgeschrieben und Forscher können sich direkt bei der Institution bewerben. Der Auftraggeber wird den inhaltlichen, zeitlichen und vor

© Springer Fachmedien Wiesbaden GmbH, ein Teil von Springer Nature 2018

M. Tausendpfund, *Quantitative Methoden in der Politikwissenschaft*,
Grundwissen Politik, https://doi.org/10.1007/978-3-658-20698-7_3

allem finanziellen Rahmen des Forschungsprojekts vorgeben, sodass der Gestaltungsspielraum des Forschers relativ begrenzt ist. Gegen Auftragsforschung ist nichts einzuwenden, sofern der Auftraggeber keinen Einfluss auf die Ergebnisse des Forschungsprojekts nimmt und die Auftragsforschung für Dritte auch als Auftragsforschung erkennbar ist.

Selbst initiiertes Forschungsprojekt Bei einem selbst initiierten Forschungsprojekt wählt der Forscher ein eigenes Thema und entwickelt einen Forschungsplan. Zur Finanzierung des eigenen Forschungsprojekts muss die Wissenschaftlerin in der Regel einen Forschungsantrag verfassen und bei einer Förderinstitution (z. B. Deutsche Forschungsgemeinschaft) einreichen.[5] Im Forschungsantrag sind Forschungsthema, Forschungsfragen, Stand der Forschung, das Arbeitsprogramm sowie die Kosten des Forschungsprojekts darzulegen und zu begründen. Der Forschungsantrag wird begutachtet und ein Fachgremium entscheidet auf Basis der Gutachten über Bewilligung und Ablehnung eines Forschungsprojekts. Das klassische Förderinstrument der Deutschen Forschungsgemeinschaft (DFG) ist die sogenannte Sachbeihilfe. Die Sachbeihilfe ermöglicht allen Personen mit abgeschlossener wissenschaftlicher Ausbildung (in der Regel Promotion) die Durchführung eines einzelnen, thematisch und zeitlich begrenzten Forschungsvorhabens. Die Einwerbung von sogenannten Drittmitteln ist anspruchsvoll, denn die Antragstellung ist aufwändig und die Konkurrenz groß. Die Förderquote bei sozialwissenschaftlichen Projekten liegt unter 30 Prozent. Hinweise zum Verfassen von Forschungsanträgen finden sich bei Schwarzer (2001); die Lektüre lohnt sich auch für das Verfassen von Haus- und Abschlussarbeiten.

Grundlagenforschung Sozialwissenschaftliche Grundlagenforschung versucht Lücken im bisherigen Wissensstand zu schließen und/oder Widersprüche im Wissensbestand aufzuarbeiten. Dabei kann es sich einerseits um die wissenschaftliche Untersuchung eines neuen Phänomens handeln (z. B. Einfluss des Internets auf die politische Beteiligung), andererseits kann auch die Leistungsfähigkeit unterschiedlicher (zum Teil sich widersprechender) Theorien zur Erklärung eines sozialwissenschaftlichen Phänomens untersucht werden (z. B. Vergleich theoretischer Ansätze zur Erklärung der Wahlbeteiligung).

5 Die Deutsche Forschungsgemeinschaft (DFG) ist die zentrale Selbstverwaltungsorganisation der Wissenschaft in Deutschland. Sie dient der Wissenschaft durch die Förderung von Forschungsprojekten an Hochschulen und anderen Forschungseinrichtungen (siehe die Homepage unter www.dfg.de).

Anwendungsorientierte Forschung Bei anwendungsorientierter Forschung leiten sich die Fragestellungen häufig aus konkreten Problemen ab (z. B. Wunsch nach Einbindung von Bürgern in lokale Entscheidungsprozesse). Im Mittelpunkt steht in der Regel die Suche nach einer „Lösung" für ein konkretes Problem. Bei der Bearbeitung der Forschungsfrage greift die anwendungsorientierte Forschung auf Erkenntnisse und Methodik der Grundlagenforschung zurück, aber im Vordergrund stehen Vorschläge für das formulierte Problem (z. B. Empfehlungen geeigneter Bürgerbeteiligungsverfahren).

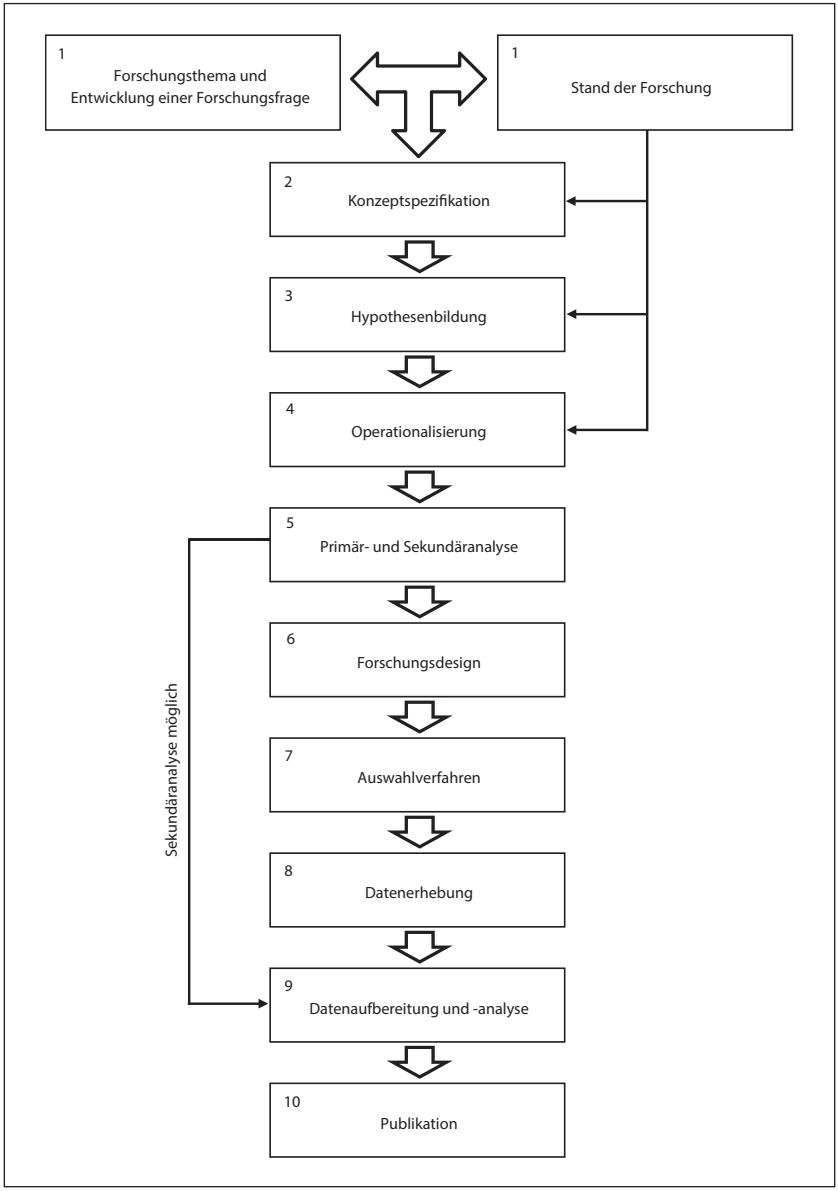
Die Differenzierungen zwischen Auftragsforschung und selbst initiierten Forschungsprojekten sowie zwischen Grundlagenforschung und anwendungsorientierter Forschung können als Endpunkte des jeweiligen Kontinuums betrachtet werden. In der Praxis können Forschungsprojekte sicherlich als „Auftragsforschung" oder „selbst initiiert" bzw. als „Grundlagenforschung" oder „anwendungsorientierte Forschung" klassifiziert werden. Während in anwendungsorientierten Forschungsprojekten die Nutzung wissenschaftlicher Methoden für die Lösung konkreter Probleme betont wird, arbeitet die sozialwissenschaftliche Grundlagenforschung die gesellschaftliche Relevanz des Forschungsprojekts heraus. Sozialwissenschaftliche Grundlagenforschung wird in der Regel durch Wissenschaftler initiiert und durch öffentliche Förderinstitutionen unterstützt (z. B. DFG), bei anwendungsorientierter Forschung handelt es sich in der Regel eher um Auftragsforschung.

3.2 Projektphasen im Überblick

In diesem Kapitel werden die einzelnen Phasen eines sozialwissenschaftlichen Forschungsprojekts vorgestellt (siehe auch Krumm und Westle 2009, S. 115–123; Diekmann 2011, S. 186–200; Schnell et al. 2013, S. 3–11). Die einzelnen Phasen eines Forschungsprojekts werden in den weiteren Kapiteln dieses Lehrbuchs ausführlicher behandelt. In diesem Kapitel steht das erste Kennenlernen dieser typischen Phasen im Mittelpunkt. Eine Übersicht der einzelnen Phasen eines empirischen Forschungsprojekts gibt Abbildung 4. Die durch die Pfeile dargestellte zeitliche Reihenfolge der einzelnen Phasen hat dabei einen idealtypischen Charakter; in der Forschungspraxis sind „Überschneidungen, Sprünge und Rückkoppelungen" (Kromrey et al. 2016, S. 69) zu beobachten.

Forschungsfrage Jedes Forschungsprojekt – angefangen von der Abschlussarbeit im Studium bis hin zur Promotion oder einem internationalen Forschungsprojekt – beginnt mit der Festlegung eines Forschungsthemas (Kasten 1 in Abbildung 4). Das Thema eines Forschungsprojekts grenzt zwar den Inhalt des Projekts

Abbildung 4 Phasen eines Forschungsprojekts



Quelle: Eigene Darstellung

lose ab, es schränkt das Projekt aber noch nicht hinreichend ein (Plümper 2012, S. 16). In Auseinandersetzung mit der Fachliteratur – dem Stand der Forschung zu einem Thema – ist die Entwicklung einer Forschungsfrage erforderlich. Bei einer Haus- oder Abschlussarbeit ist eine Forschungsfrage ausreichend, bei mehrjährigen Forschungsprojekten (z. B. Promotionen) werden ggf. mehrere Forschungsfragen formuliert. Häufig wird die Forschungsfrage noch in Teilfragen aufgeteilt, um das Projekt klarer strukturieren zu können. Die Bearbeitung einzelner Teilfragen zielt schließlich darauf ab, die Forschungsfrage zu beantworten.

Konzeptspezifikation Ist die Forschungsfrage (in einer ersten Version) formuliert, dann müssen die verwendeten Konzepte und Begriffe geklärt werden. Viele Sozialwissenschaftler beschäftigen sich mit schillernden Begriffen wie Legitimität, Identität, Sozialkapital, Vertrauen oder auch Demokratie, die höchst unterschiedlich definiert und verstanden werden. Oft handelt es sich aber auch um Alltagsbegriffe, die erst in die sozialwissenschaftliche Forschungssprache übersetzt werden müssen. Begriffe wie Politikverdrossenheit oder Wutbürger sind keine sozialwissenschaftlichen Konzepte, sondern wurden vielmehr von Journalisten erfunden. Selbstverständlich werden diese Phänomene auch in den Sozialwissenschaften empirisch untersucht, doch müssen solche Begriffe zunächst spezifiziert werden. Die Phase, in der solche Begriffe theoretisch geklärt werden, wird als Konzeptspezifikation bezeichnet (Kasten 2 in Abbildung 4). Kromrey et al. (2016, S. 113) bezeichnen diesen Arbeitsschritt als „dimensionale und semantische Analyse"; in der gängigen Literatur hat sich die Bezeichnung „Konzeptspezifikation" durchgesetzt (z. B. Gerring 2001; Wonka 2007; Diekmann 2011, S. 194; Schnell et al. 2013, S. 7).

Hypothesen Sozialwissenschaftliche Forschung ist insbesondere an Ursache-Wirkungs-Beziehungen interessiert. Diese vermuteten Ursache-Wirkungs-Beziehungen werden in Hypothesen formuliert (Kasten 3 in Abbildung 4). Im sozialwissenschaftlichen Kontext handelt es sich bei Hypothesen um eine begründete und theoriegeleitete Vermutung zwischen mindestens zwei Sachverhalten, die empirisch überprüft wird. Eine zentrale Anforderung an Hypothesen ist daher, dass sie an der Erfahrung (der Realität) scheitern können.

Operationalisierung Viele theoretische Begriffe in den Sozialwissenschaften, wie beispielsweise „Legitimität", „Vertrauen" oder „Soziale Schicht", lassen sich in der Regel nicht direkt beobachten. In einem empirischen Forschungsprojekt müssen diesen theoretischen Begriffen deshalb beobachtbare Sachverhalte zugeordnet werden. Die Zuordnung von einem oder mehreren beobachtbaren Indikatoren zu einem theoretischen Begriff wird als Operationalisierung bezeichnet (Kasten 4 in

Abbildung 4). Die theoretischen Konstrukte werden in einem empirischen Forschungsprozess durch Indikatoren repräsentiert. Ein Indikator ist ein beobachtbarer Sachverhalt, der einen theoretischen Begriff in die Realität „übersetzt".

Auseinandersetzung mit Fachliteratur Die Arbeitsschritte Konzeptspezifikation, Hypothesenbildung und Operationalisierung erfolgen in intensiver Auseinandersetzung mit der Fachliteratur. Zu vielen theoretischen Begriffen existieren bereits Konzeptspezifikationen und Operationalisierungsvorschläge. Im Studium ist es nicht das Ziel, eigene Konzeptspezifikationen oder Vorschläge zur Operationalisierung zu entwickeln. Vielmehr muss sich die Forscherin zunächst mit den verschiedenen existierenden Vorschlägen intensiv auseinandersetzen und begründen, warum die gewählte Konzeptspezifikation oder Operationalisierung für die eigene Forschungsfrage geeignet ist. Für die Arbeitsschritte Konzeptspezifikation, Hypothesenbildung und Operationalisierung ist deshalb ein zeitintensives Literaturstudium erforderlich.

Primär- und Sekundäranalyse Der praktische Teil eines Forschungsprojekts beginnt mit der Frage, ob für das eigene Forschungsprojekt bereits adäquate Daten zur Verfügung stehen, die für eine angemessene Operationalisierung der theoretischen Begriffe und die empirische Überprüfung der formulierten Hypothesen genutzt werden können (Kasten 5 in Abbildung 4). Falls ja, dann entfallen die Phasen Forschungsdesign, Auswahlverfahren und Datenerhebung und der Forscher kann eine Sekundäranalyse durchführen. Bei einer Sekundäranalyse nutzt der Forscher vorhandenes Datenmaterial, um seine Forschungsfrage zu beantworten. Da insbesondere die Entwicklung eines Forschungsdesigns und die Datenerhebung mit einem erheblichen Zeit- und Geldaufwand verbunden sind, ist eine Sekundäranalyse „in akademischen Kontexten eher die Regel als die Ausnahme" (Schnell et al. 2013, S. 242). Falls keine geeigneten Daten zur Verfügung stehen und der Forscher über ausreichende Ressourcen verfügt, dann ist eine Primäranalyse erforderlich und es muss zunächst ein angemessenes Forschungsdesign entwickelt werden.

Forschungsdesign Das Forschungsdesign beinhaltet Entscheidungen über das Vorgehen bei der Bearbeitung der Forschungsfrage (Kasten 6 in Abbildung 4). In Abhängigkeit von der Forschungsfrage sowie der Konzeptspezifikation, der Hypothesenbildung und der Operationalisierung stellt sich erstens die Frage nach der Untersuchungsebene. Sollen Zusammenhänge auf der gesellschaftlichen Ebene (Makro) oder auf der Individualebene (Mikro) untersucht werden? Zu der Festlegung des Forschungsdesigns gehört zweitens auch die Entscheidung, ob man ein Experiment (Labor-, Feld- oder Quasiexperiment) oder ein nicht-experimentelles Vorgehen (Querschnitt-, Trend- oder Paneldesign) wählt. Eine dritte Entschei-

Hausarbeit ist kein Forschungsprojekt

Bei eigenen Haus- oder auch Abschlussarbeiten im Studium werden Studierende praktisch niemals mit allen Phasen eines Forschungsprojekts konfrontiert. Im Gegenteil: Häufig handelt es sich bei Hausarbeiten – insbesondere zu Beginn des Studiums – um sogenannte Literaturberichte, bei denen der Forschungsstand zu einem Thema aufbereitet und eine konkrete Forschungsfrage auf Basis der Fachliteratur beantwortet wird. Dennoch gehört die Kenntnis der einzelnen Phasen eines empirischen Forschungsprojekts zum unverzichtbaren Grundlagenwissen in einem sozialwissenschaftlichen Studium. Warum? Selbst wer niemals ein sozialwissenschaftliches Forschungsprojekt plant und durchführt, benötigt elementare Methoden- und Statistikkenntnisse, um empirische Studien lesen, verstehen und bewerten zu können. Ohne ein Verständnis der Notwendigkeit einer Konzeptspezifikation, der Hypothesenbildung und Operationalisierung sowie des Forschungsdesigns und der Datenerhebung ist eine gehaltvolle Auseinandersetzung mit empirischen Ergebnissen nicht möglich. Nur wer die typischen Schwierigkeiten und „Fallstricke" empirischer Forschungsprojekte kennt, kann bei der Lektüre der Fachliteratur gezielt danach suchen. Ist die Konzeptspezifikation nachvollziehbar? Hat der Autor die Operationalisierung überzeugend begründet? Warum wählt der Autor ein Querschnitts- und kein Paneldesign? Nur wer weiß, wonach er sucht, kann solche Fehler überhaupt finden. Schließlich werden die Autoren der Studien in den seltensten Fällen explizit auf solche Schwierigkeiten hinweisen. Die Auseinandersetzung mit dem typischen Forschungsprozess dient deshalb insbesondere dem Ziel, die reflektierte Auseinandersetzung mit Fachliteratur zu ermöglichen.

dung bezieht sich auf den zeitlichen Aspekt der Datenerhebung: Sollen Daten zu einem oder mehreren Zeitpunkten gesammelt werden? Die Antworten auf die einzelnen Fragen sind einerseits immer abhängig von der konkreten Fragestellung des Forschungsprojekts, andererseits aber auch von den zur Verfügung stehenden Ressourcen (z. B. Zeit und Geld).

Auswahlverfahren Über wen oder was sollen Aussagen gemacht werden? Diese Frage steht bei der Festlegung des Auswahlverfahrens im Mittelpunkt (Kasten 7 in Abbildung 4). Bei der Festlegung des Auswahlverfahrens muss zunächst geklärt werden, ob alle Elemente (z. B. alle Personen), über die eine Aussage gemacht wer-

den soll, untersucht werden oder nur eine Teilmenge dieser Elemente. Aus zeitlichen, finanziellen und forschungspraktischen Gründen dominieren in den Sozialwissenschaften Teilerhebungen. Bei der Teilerhebung kann dann wieder zwischen zufälligen und nichtzufälligen Auswahlverfahren unterschieden werden.

Datenerhebung Die vermutlich bekannteste Phase eines sozialwissenschaftlichen Forschungsprozesses ist die Datenerhebung (Kasten 8 in Abbildung 4). In den Sozialwissenschaften existieren zahlreiche Datenerhebungsmethoden, unter anderem Befragung, Beobachtung und Inhaltsanalyse. Die einzelnen Verfahren lassen sich wieder in mehrere Varianten einteilen (z. B. die Befragung in persönliche, telefonische, schriftliche und internetgestützte Befragung).

Datenaufbereitung und -analyse Die in der Phase der Datenerhebung gesammelten Informationen müssen strukturiert aufbereitet und gespeichert werden (Kasten 9 in Abbildung 4). Dazu gehört auch eine umfangreiche Datenbereinigung, bei der Erfassungs- und Übertragungsfehler gesucht und korrigiert werden. Für die Datenanalyse steht mittlerweile eine Vielzahl an Auswertungstechniken zur Verfügung. Die Entscheidung für ein konkretes Analyseverfahren ist dabei von der Forschungsfrage und dem zur Verfügung stehenden Datenmaterial abhängig.

Publikation Damit die Ergebnisse eines Forschungsprojekts anderen Wissenschaftlern bekannt werden, müssen die Ergebnisse publiziert werden (Kasten 10 in Abbildung 4). Nur eine veröffentlichte Studie kann zum wissenschaftlichen Fortschritt beitragen (Schnell et al. 2013, S. 10). Die Publikation geschieht einerseits durch einen – meist öffentlich nicht zugänglichen und wissenschaftlich daher zu vernachlässigenden – Abschlussbericht und andererseits durch die Veröffentlichungen von Büchern und Zeitschriftenaufsätzen.

Die Darstellung eines Forschungsprojekts in Abbildung 4 bezieht sich auf quantitativ angelegte Studien; die meisten Phasen sind aber Bestandteile von quantitativen und qualitativen Forschungsprojekten. Selbstverständlich müssen sich sowohl quantitativ als auch qualitativ orientierte Forscherinnen für ein Forschungsthema entscheiden und mit dem bisherigen Forschungsstand auseinandersetzen. Auch mit Fragen der Konzeptspezifikation, der Wahl des geeigneten Forschungsdesigns, des Auswahlverfahrens oder der Datenerhebung werden Wissenschaftlerinnen beider Forschungstraditionen konfrontiert. Schließlich wollen sowohl quantitativ als auch qualitativ orientierte Forscher ihre Ergebnisse publizieren. Die Gemeinsamkeiten quantitativer und qualitativer Forschungsprojekte sind daher deutlich größer als die Unterschiede.

3.2.1 Forschungsthema und Entwicklung einer Forschungsfrage

Jedes Forschungsprojekt, unabhängig davon, ob Auftragsforschung oder selbst initiiertes Forschungsprojekt bzw. Grundlagenforschung oder anwendungsorientierte Forschung, beginnt mit der Festlegung des Forschungsthemas. Die Wahl eines Forschungsthemas hängt einerseits von den Interessen des Forschers und andererseits von der jeweils aktuellen Forschungssituation (z. B. Aktualität) ab. Das Thema ist aber lediglich der grobe Rahmen des Forschungsprojekts und bildet noch keine ausreichende Grundlage für die Entwicklung eines Forschungsprojekts. Themen wie Sozialkapital, Vertrauen, Wahlverhalten oder soziale Ungleichheit sind viel zu umfangreich, um als Arbeitsgrundlage dienen zu können. Plümper (2012, S. 16) verdeutlicht dies am Thema „Arbeitslosigkeit“:

> *„Wählt man Arbeitslosigkeit als Thema, kann man ebenso über den Einfluss der Arbeitslosigkeit auf die Entwicklung von Jugendbewegungen, auf die Selbstmordrate, auf Ehescheidungen, auf Unternehmensgewinne arbeiten wie auch über den Einfluss von Kündigungsschutzregeln, Bankinsolvenzen, Investitionen in Humankapital auf die Arbeitslosenquote. Mit anderen Worten: Der Begriff ‚Arbeitslosigkeit‘ allein schränkt das Thema nicht hinreichend ein. Wenn Sie nur ein Wort als Thema haben, haben Sie noch kein Projekt.“*

Thema mit Kernbegriffen einschränken Das Thema grenzt den Forschungsgegenstand nur lose ab. Deshalb empfiehlt Plümper (2012, S. 16), das gewählte Thema mit zwei oder drei Kernbegriffen zu beschreiben. „Arbeitslosigkeit und Wahlverhalten“ ist deutlich präziser als „Arbeitslosigkeit“. Mit der notwendigen Präzisierung des Themas „ist eine Reduktion der Vielfalt und damit eine Eingrenzung des untersuchten Feldes verbunden“ (Schwarzer 2001, S. 148). Dies ist auch deshalb erforderlich, um den Forschungsstand zu einem Thema aufarbeiten zu können. Der Stand der Forschung zu „Arbeitslosigkeit“ kann in einem Forscherleben praktisch nicht aufgearbeitet werden, das Thema „Arbeitslosigkeit und Wahlverhalten“ (Faas 2010) ist zumindest auf den ersten Blick überschaubar.

Das Thema muss aber weiter konkretisiert werden, damit es im Rahmen eines zeitlich befristeten Forschungsprojekts bearbeitet werden kann (z. B. In welchem Land wird der Zusammenhang zwischen Arbeitslosigkeit und Wahlverhalten untersucht? In welchem Zeitraum wird der Zusammenhang untersucht?). Ein zu umfangreiches oder unpräzises Thema ist vermutlich das häufigste Problem eines Forschungsprojekts. Die Ursache dieses Problems liegt darin begründet, dass gerade zu Beginn der Auseinandersetzung mit einem Thema nur geringe Informationen zu diesem Sachverhalt vorliegen. In der Praxis sind aber selbst eng definierte Themen in der Regel deutlich umfangreicher als ursprünglich erwartet.

Forschungsfrage formulieren Nach der (ersten) thematischen Eingrenzung des Forschungsprojekts muss eine präzise Forschungsfrage formuliert oder ein konkretes Forschungsproblem identifiziert werden. Die Forschungsfrage zielt wieder darauf ab, das Thema eines Forschungsprojekts weiter einzugrenzen. Wie weiter oben bereits dargestellt, versucht sozialwissenschaftliche Grundlagenforschung Lücken im bisherigen Wissensstand zu schließen und/oder Widersprüche im Wissensbestand zu klären. Die Untersuchung „neuer" sozialer Phänomene, Widersprüche zwischen theoretischen Erwartungen und empirischen Befunden sowie unterschiedliche theoretische (und häufig sich widersprechende) Ansätze zur Erklärung eines empirischen Sachverhalts sind erfolgversprechende Strategien zur Entwicklung einer Forschungsfrage. Für die Entwicklung einer gehaltvollen Forschungsfrage oder Identifikation eines Forschungsproblems ist die intensive Auseinandersetzung mit der Fachliteratur erforderlich. Nur wer die zentralen Veröffentlichungen zu einem Thema kennt, kann überhaupt Widersprüche zwischen Theorie und Empirie entdecken und eine Forschungsfrage entwickeln. Kapitel 4 beschäftigt sich mit dem Forschungsthema und der Entwicklung von Forschungsfragen.

3.2.2 Konzeptspezifikation

Jede Wissenschaft hat ihre eigene Fachterminologie. Mediziner sprechen von „Glaukom" oder „Angina pectoris", Physiker vom „Dopplereffekt" oder der „Unschärferelation" von Heisenberg und Germanisten von „Jambus" und „Choriambus". Auch Sozialwissenschaftler haben ihre Fachsprache und verwenden Begriffe wie „Sozialkapital", „Legitimität", „Vertrauen" oder „Sozialer Status". Diese Begriffe stehen stellvertretend für komplexe Konzepte und sind häufig nicht „eindeutig definiert und von anderen Begriffen abgegrenzt" (Schnell et al. 2013, S. 118). Selbst für auf den ersten Blick einfache Konzepte wie „Politische Beteiligung" finden sich in der Fachliteratur zahlreiche Definitionen (für eine Übersicht siehe van Deth 2003, S. 170; siehe auch van Deth 2014), bei schillernden Begriffen wie „Sozialkapital" ist die Situation deutlich unübersichtlicher (für eine Übersicht siehe Haug 1997; Franzen und Freitag 2007b; Franzen und Pointner 2007). Häufig wird der gleiche Begriff auch in verschiedenen theoretischen Zusammenhängen verwendet, sodass unterschiedliche theoretische Aspekte betont werden. Schließlich müssen sich Wissenschaftlerinnen auch nicht immer über die korrekte und angemessene Verwendung eines Begriffs einig sein – im Gegenteil: In der Literatur finden sich sehr häufig unterschiedliche und auch widersprechende Definitionen eines – auf den ersten Blick – einfachen Konzepts.

Die Phase, in der solche Begriffe theoretisch geklärt werden, wird als Konzeptspezifikation bezeichnet (Kasten 2 in Abbildung 4). Das Ziel der Konzeptspezifikation beschreibt Wonka (2007, S. 65) wie folgt:

> *„Wir möchten mit einem analytischen Instrumentarium arbeiten, das uns einen möglichst klar verständlichen theoretischen Diskurs sowie die möglichst eindeutige analytische Erfassung und Abgrenzung der für unser Forschungsprojekt relevanten empirischen Objekte erlaubt. Zu diesem Zweck ist es notwendig, die von uns verwendeten Termini möglichst klar zu definieren."*

Klare Konzepte erforderlich Insbesondere in den Sozialwissenschaften ist eine Konzeptspezifikation erforderlich, da die Forscher häufig mit Alltagsbegriffen wie „Politikverdrossenheit" (z. B. Arzheimer 2002) oder „Wutbürger" (z. B. Stürmer 2011) konfrontiert werden. Für die eigene Arbeit ist es zwingend erforderlich, das Verständnis der verwendeten Konzepte und Begriffe offenzulegen. Klar definierte Konzepte sind unabdingbare Voraussetzung für die Formulierung von Hypothesen und die Entwicklung von Operationalisierungen. Empirische Ergebnisse können immer nur im Kontext dieser Konzeptspezifikation interpretiert werden. Die Phase der Konzeptspezifikation setzt deshalb ein intensives Literaturstudium voraus. Existierende Definitionen und Konzeptspezifikationen zu einem Begriff müssen in der Fachliteratur identifiziert und gegenübergestellt werden. Mit Blick auf die erforderliche Zeit ist die Konzeptspezifikation eine stark unterschätzte Phase im Forschungsprozess. Zwar ist es für eine Haus- oder Abschlussarbeit in der Regel ausreichend, eine (etablierte) Konzeptspezifikation explizit auszuwählen. Die „Wahl" dieser Konzeptspezifikation ist aber sorgfältig zu begründen und ohne Kenntnis des Forschungsstands nicht möglich. Kapitel 5 dieses Buchs beschäftigt sich mit der Konzeptspezifikation.

3.2.3 Hypothesenbildung

Im Alltag ist eine Hypothese eine von Widersprüchen freie, aber zunächst unbewiesene Aussage über einen Tatbestand (Kromrey et al. 2016, S. 47). Im sozialwissenschaftlichen Kontext handelt es sich bei einer Hypothese um eine begründete Vermutung zwischen mindestens zwei Sachverhalten (Diekmann 2011, S. 124). Eine Hypothese ist damit eine Aussage über den Zusammenhang von mindestens zwei Merkmalen, zum Beispiel die Ablösesumme eines Fußballspielers und der Anzahl der erzielten Tore in der Bundesliga („Je höher die Ablösesumme, desto mehr Tore erzielt der Spieler in der Saison").

Deterministische und probabilistische Hypothesen Diekmann (2011, S. 124) unterscheidet zwischen deterministischen und probabilistischen Hypothesen. Deterministische Hypothesen behaupten, dass ein Sachverhalt oder ein Ereignis mit Sicherheit eintreten muss, wenn bestimmte Bedingungen vorliegen (z. B. Wasser kocht bei einer Temperatur von 100 Grad Celsius). Probabilistische Hypothesen behaupten, dass ein Sachverhalt oder ein Ereignis nur mit einer gewissen Wahrscheinlichkeit eintritt, wenn bestimmte Bedingungen vorliegen. In den Naturwissenschaften werden in der Regel deterministische, in den Sozialwissenschaften probabilistische Hypothesen formuliert. Eine andere Bezeichnung für probabilistische Hypothesen sind auch statistische Hypothesen. Da in den Sozialwissenschaften probabilistische Hypothesen die Regel sind, wird auf den Zusatz „probabilistisch" oder „statistisch" häufig verzichtet.

Bei der Formulierung von Hypothesen lassen sich zwei Grundformen unterscheiden: Wenn-dann-Hypothesen und Je-desto-Hypothesen. Ein Beispiel für eine Wenn-dann-Hypothese ist: „Wenn eine Person viel raucht, dann steigt die Wahrscheinlichkeit eines Herzinfarkts." Ein Beispiel für eine Je-desto-Hypothese ist: „Je höher die Ablösesumme, desto mehr Tore erzielt der Spieler in der Saison." Kapitel 6 dieses Lehrbuchs gibt einen Überblick über sozialwissenschaftliche Hypothesenbildung.

3.2.4 Operationalisierung

Die meisten Hypothesen enthalten theoretische Konzepte wie Vertrauen oder Legitimität, die sich nicht direkt, sondern bestenfalls indirekt beobachten lassen. Es handelt sich um sogenannte latente Konstrukte, die sich der unmittelbaren Beobachtung entziehen. Hypothesen beziehen sich häufig auf einen vermuteten Zusammenhang zwischen zwei nicht direkt beobachtbaren Konzepten. Diese Überlegung lässt sich an der folgenden Hypothese verdeutlichen: „Je höher der soziale Status, desto größer die Wahrscheinlichkeit der Wahlbeteiligung." Die beiden Konzepte „Sozialer Status" und „Wahlbeteiligung" dieser Hypothese sind in der Regel nicht direkt beobachtbar. Für eine empirische Überprüfung müssen diese Konzepte mit beobachtbaren Indikatoren verknüpft werden, die Rückschlüsse auf diese nicht direkt beobachtbaren Konzepte erlauben. Diese Phase im Forschungsprozess wird als Operationalisierung bezeichnet.

Verknüpfung von Konzept und Indikator Die Verknüpfung zwischen einem oder mehreren Indikatoren und einem theoretischen Begriff geschieht über Korrespondenzregeln. Die Korrespondenzregel verbindet den (beobachtbaren) Indikator mit dem (nicht beobachtbaren) Konzept. Ein möglicher Indikator der Wahlbetei-

ligung könnte beispielsweise die Frage nach der Wahlbeteiligung sein. Personen werden gefragt, ob sie sich an der nächsten Bundestagswahl beteiligen (Wahlbeteiligungsabsicht). Alternativ können Personen aber auch gefragt werden, ob sie sich an der letzten Bundestagswahl beteiligt haben (Erinnerungsfrage). Beide Varianten sind mögliche – und im Forschungsalltag häufig verwendete – Indikatoren der Wahlbeteiligung. Zur Operationalisierung des sozialen Status finden sich in der Literatur verschiedene Vorschläge (für eine Übersicht siehe z. B. Wolf 1995), bei denen häufig Informationen zur Bildung, zur beruflichen Stellung und zum Einkommen einer Person kombiniert werden. Die verschiedenen Operationalisierungen der Wahlbeteiligung und des sozialen Status führen zwar in der Regel zu vergleichbaren, nicht aber zu identischen Ergebnissen. Dies deutet darauf hin, dass die Indikatoren jeweils unterschiedliche Aspekte des nicht direkt beobachtbaren Konzepts abbilden. Die Korrespondenzregel, die Indikator und Konzept verbindet, entspricht damit einer Hypothese. Es wird ein Zusammenhang zwischen Indikator und Konzept postuliert. Eine Korrespondenzregel kann auch falsch sein, das heißt, der gewählte Indikator erlaubt keinen Rückschluss auf das theoretische Konzept. Kapitel 7 behandelt die Phase der Operationalisierung im Forschungsprozess.

3.2.5 Primär- und Sekundäranalyse

Existiert Datenmaterial zur Beantwortung der Forschungsfrage? Spätestens nach der Formulierung der Forschungsfrage, der Konzeptspezifikation und der Hypothesenbildung sowie der Operationalisierung sollte der Forscher prüfen, ob für seine konkrete Forschungsfrage bereits empirisches Datenmaterial zur Verfügung steht. Falls ja, dann kann und sollte auf die Entwicklung eines Forschungsdesigns, die Entscheidung für ein Auswahlverfahren und die in der Regel sehr teure und langwierige Datenerhebung verzichtet werden. Durch die Nutzung bestehender Datensätze können Zeit und Geld gespart und die Nerven geschont werden. Deshalb sollte intensiv geprüft werden, ob sich bereits erhobene Daten zur Untersuchung der gewählten Forschungsfrage eignen. Eine empirische Analyse auf Basis vorhandener Daten wird Sekundäranalyse bezeichnet. Bei einer Primäranalyse führt der Forscher die Datenerhebung (und die damit verbundene Entwicklung eines Forschungsdesigns sowie die Entscheidung für ein Auswahlverfahren) selbst durch. Die Bezeichnungen „Primäranalyse“ und „Sekundäranalyse“ beinhalten kein wertendes Urteil über die Qualität des Forschungsprojekts. Im Gegenteil: Sekundäranalysen auf Basis „hochwertiger“ Daten sind Primäranalysen auf Grundlage minderwertiger Daten vorzuziehen.

Mittlerweile steht eine große Anzahl an nationalen und internationalen Datensätzen zur Verfügung, mit denen (auch angehende) Sozialwissenschaftlerinnen

Standardisierte Befragungen bei Haus und Abschlussarbeiten

Bei quantitativ orientierten Studien im Rahmen von Haus- und Abschlussarbeiten ist eine Sekundäranalyse die einzig sinnvolle Forschungsstrategie. Weder bei einer Haus- noch bei einer Abschlussarbeit stehen ausreichend Ressourcen zur Verfügung, um alle Schritte einer standardisierten Befragung sorgfältig durchführen zu können. „Daher ist die Empfehlung für die selbstständige Durchführung von Primärdatenerhebungen mit Surveys innerhalb von Qualifikationsarbeiten ebenso eindeutig und einfach wie unerwünscht: Niemals" (Schnell 2012, S. 204; siehe auch Kopp und Lois 2014, S. 31–32).

unterschiedlichste Forschungsfragen empirisch untersuchen können. Diese Datensätze stehen häufig kostenlos zum Download zur Verfügung; in Deutschland ist GESIS – Leibniz-Institut für Sozialwissenschaften eine Einrichtung, die Datensätze sammelt und der interessierten (akademischen) Öffentlichkeit zugänglich macht (www.gesis.org). In Kapitel 8 findet sich eine Übersicht wichtiger sozialwissenschaftlicher Datensätze.

3.2.6 Forschungsdesign

Das Forschungsdesign stellt den Rahmen des empirischen Forschungsprojekts dar. Bei der Entwicklung des Forschungsdesigns muss der Forscher klären, welche methodische Vorgehensweise für die Bearbeitung der Forschungsfragen gewählt werden soll. Nach Diekmann (2011, S. 194) lassen sich drei zentrale Entscheidungen innerhalb eines Forschungsdesign unterscheiden: Erstens die Untersuchungsebene (Mikro vs. Makro), zweitens die Untersuchungsform (experimentell vs. nicht-experimentell) sowie drittens die Häufigkeit der Datenerhebung (einmalig vs. mehrmalig).

Untersuchungsebene Durch die Forschungsfrage, Konzeptspezifikation, Hypothesenbildung und Operationalisierung ist die Entscheidung der Untersuchungsebene meist schon (implizit) getroffen. Die Analyse von Zusammenhängen auf der gesellschaftlichen Makroebene wird als Aggregatdatenanalyse bezeichnet, die Untersuchung von Zusammenhängen auf der Mikroebene wird Individualanalyse genannt. Studien, die den Einfluss des gesellschaftlichen Umfelds auf die Individualebene untersuchen, nennt man Kontextanalysen. Die Festlegung der konkre-

ten Analyseebene ist wichtig, da eine Forschungsfrage durchaus auf verschiedenen Analyseebenen untersucht werden kann. Geys (2006) gibt beispielsweise eine Übersicht zu den Faktoren der Wahlbeteiligung auf der Makroebene, Westle et al. (2013) analysieren die Determinanten der Wahlbeteiligung auf der Mikroebene und Faas (2013) untersucht den Einfluss des (lokalen) Umfelds auf die Wahlbeteiligung (Kontextanalyse). Mit der Entscheidung für eine Analyseebene sind Vor- und Nachteile verbunden, die die Forscherin – aber auch der Leser empirischer Studien – kennen sollte.

Untersuchung der Wahlbeteiligung auf der Mikro- oder Makroebene

Bei einer Untersuchung der Wahlbeteiligung auf der Makroebene können Daten der amtlichen Statistik genutzt werden, die relativ einfach zugänglich sind. Diese Daten können dann mit anderen Merkmalen der Makroebene in Beziehung gesetzt werden. Bei einem Vergleich der nationalen Wahlbeteiligung in Europa könnten das beispielsweise die Wahlpflicht, der Wahltag, das Wahlsystem oder auch existierende Sperrklauseln sein (z. B. Freitag 1996). Allerdings erlauben solche Makrodaten keine Rückschlüsse über die individuellen Motive der Wahlbeteiligung. Für die Analyse individueller Bestimmungsfaktoren der Wahlbeteiligung (z. B. Bildung, politisches Interesse) sind zwingend Individualmerkmale erforderlich. Diese Merkmale werden in Befragungen erhoben. In solchen Befragungen werden die Personen auch nach ihrer Wahlabsicht gefragt (Beteiligungsabsicht) oder, ob sie bei der letzten Wahl gewählt haben (Erinnerungsfrage). Beide Fragevarianten sind problematisch: Bei der ersten Frage handelt es sich lediglich um eine Verhaltensabsicht, bei der zweiten Frage können Erinnerungsfehler auftreten. Je größer der zeitliche Abstand zwischen Wahl und Befragung, desto stärker dürfte die Abweichung vom tatsächlichen Verhalten sein. Darüber hinaus kann das Antwortverhalten der Befragten auch durch den Faktor „Soziale Erwünschtheit“ beeinflusst werden. „Wählen gehen“ gilt als sozial erwünschtes Verhalten, sodass Personen auch nicht immer wahrheitsgemäß antworten. Schließlich könnte es auch sein, dass Nichtwähler grundsätzlich seltener an solchen Befragungen teilnehmen. Diese Probleme haben Konsequenzen: Bei einem Vergleich der Beteiligungsraten zeigt sich regelmäßig, dass die per Individualdaten ermittelte Wahlbeteiligung deutlich höher liegt als die amtliche Wahlbeteiligung (eine Übersicht der Vor- und Nachteile bietet Keil 2009, S. 425). Die Wahl einer Analyseebene hat deshalb immer Konsequenzen, die der Forscher berücksichtigen sollte.

Experiment als Idealtyp empirischer Forschung Bei der Untersuchungsform bietet sich die Unterscheidung zwischen einem experimentellen und nicht-experimentellen Vorgehen an. Als Idealtyp empirischer Forschung (Behnke et al. 2010, S. 52) zielt das Experiment darauf ab, kausale Zusammenhänge zu entdecken. Dabei lassen sich Labor-, Feld- und Quasi-Experimente unterscheiden. In den Naturwissenschaften und der Psychologie stellt das Experiment die Standard-Untersuchungsform dar. In der Soziologie und in der Politikwissenschaft sind experimentelle Untersuchungsformen seltener, aber in jüngerer Zeit ist auch in diesen Disziplinen eine stärkere Aufmerksamkeit gegenüber experimentellen Untersuchungsformen festzustellen (z. B. Faas und Huber 2010; Druckman et al. 2011; Keuschnigg und Wolbring 2015; Jäckle 2015). Dies unterstreicht auch die Ende 2014 neugegründete Fachzeitschrift „Journal of Experimental Political Science" (Morton und Tucker 2014). In der Soziologie und Politikwissenschaft sind derzeit allerdings nicht-experimentelle Untersuchungsformen der Standard. Dabei kommt sogenannten Querschnittsstudien eine besondere Bedeutung zu.

Häufigkeit der Datenerhebung Die dritte Unterscheidung innerhalb des Forschungsdesigns bezieht sich auf die Häufigkeit der Datenerhebung (einmalig vs. mehrmalig). Eine nicht-experimentelle Querschnittsstudie ist durch eine einmalige Datenerhebung zu einem Zeitpunkt bzw. innerhalb eines Zeitraums charakterisiert. Bei Trend- und Panelstudien handelt es sich ebenfalls um nicht-experimentelle Untersuchungsformen, allerdings sehen diese Varianten mehrere Erhebungen vor. Bei einer Trendstudie werden die Daten bei unterschiedlichen Untersuchungseinheiten zu mehreren Zeitpunkten erhoben, bei einer Panelstudie werden die Daten bei den gleichen Untersuchungseinheiten mehrmals erhoben. Bei einer Panelstudie liegen daher Informationen zu gleichen Merkmalen und gleichen Untersuchungseinheiten zu mehreren Zeitpunkten vor. Dies erlaubt kausale Rückschlüsse, weshalb eine Panelstudie einem experimentellen Vorgehen am nächsten kommt. Trend- und Panelstudien werden häufig Längsschnittstudien genannt, da sie mehrere Datenerhebungen umfassen. Kapitel 9 behandelt die Phase des Forschungsdesigns und die damit verbundenen Untersuchungsformen bei quantitativ orientierten Forschungsprojekten.

3.2.7 Auswahlverfahren

Voll- oder Teilerhebung? Vor der Datenerhebung muss die Forscherin das Auswahlverfahren festlegen. In einem ersten Schritt muss er entscheiden, ob eine Voll- oder Teilerhebung erforderlich ist. Bei einer Vollerhebung werden Informationen aller Elemente (z. B. Bürger) der Grundgesamtheit erhoben. Bei der Grundgesamt-

heit handelt es sich vereinfacht formuliert um alle Elemente, für die die Aussagen der Untersuchung gelten sollen (Schnell et al. 2013, S. 255). Die konkrete Definition einer Grundgesamtheit ist damit kontextspezifisch und von der konkreten Forschungsfrage abhängig. Deshalb muss bei jedem empirischen Forschungsprojekt die Grundgesamtheit präzise definiert werden. Bei einer Studie zur Bundestagswahl bilden beispielsweise alle wahlberechtigten Bürger und Bürgerinnen die Grundgesamtheit, bei einer Studierendenbefragung sind alle (eingeschriebenen) Studierenden der Universität die Grundgesamtheit und bei einer Studie zur Pressevielfalt in Deutschland stellen alle (gedruckten) Zeitungen eine mögliche Grundgesamtheit dar. Basiert die Datenerhebung eines Forschungsprojekts auf einer Vollerhebung, dann werden Informationen aller Elemente dieser Grundgesamtheit gesammelt (z. B. alle wahlberechtigten Personen werden befragt). Solche Vollerhebungen haben zwar einige Vorteile, sie sind aber in der Regel erstens sehr aufwändig, zweitens oft nicht zu realisieren und drittens häufig auch nicht erforderlich. Deshalb dominiert in den Sozialwissenschaften die Teilerhebung.

Stichprobe Bei einer Teilerhebung werden bei der Informations- bzw. Datensammlung nicht alle Elemente einer Grundgesamtheit berücksichtigt, sondern nur eine Auswahl. Werden die Elemente der Grundgesamtheit (z. B. Bürger) nach festen Regeln ausgewählt, dann handelt es sich um eine Stichprobe. Existieren für die Auswahl keine festen Regeln, dann handelt es sich um eine willkürliche Auswahl. Stichproben, also Auswahlen nach festen Regeln, können zudem danach unterschieden werden, ob die Auswahlregel auf einem Zufallsprozess oder einer bewussten Auswahl basiert. Bei quantitativen Forschungsprojekten mit einer allgemeinen Personenbefragung (z. B. wahlberechtigte Bürgerinnen und Bürger in Deutschland) wird häufig eine Stichprobe angestrebt, die auf einem Zufallsprozess basiert. Der Begriff „Zufall“ bezieht sich dabei aber nicht auf den Alltagsbegriff „Zufall“, sondern auf das mathematische Zufallsprinzip. Dieser mathematische Zufallsbegriff bedeutet, dass alle Elemente einer Grundgesamtheit eine bestimmte, theoretisch berechenbare und von 0 verschiedene Chance haben, in die Stichprobe zu gelangen. Solche Stichprobenverfahren werden auch Random-Stichproben genannt (Löffler und von der Heyde 2014, S. 19). Kapitel 10 informiert über klassische Auswahlverfahren in den Sozialwissenschaften.

3.2.8 Datenerhebung

Befragung In den Sozialwissenschaften existiert eine Vielzahl an Datenerhebungsverfahren. Die bekannteste Datenerhebungstechnik ist sicherlich die Befragung, die als „Standardinstrument empirischer Sozialforschung“ (Schnell et al.

2013, S. 314) oder auch als „Königsweg der praktischen Sozialforschung“ (König 1974, S. 27) bezeichnet wird. „Befragung“ ist dabei nur ein loser Oberbegriff für eine Vielzahl von Befragungstechniken. Es lassen sich persönliche, telefonische, schriftliche und internetgestützte Befragungen unterscheiden. Befragungen können zudem nach dem Ausmaß der Strukturiertheit (gering vs. hoch) und der Anzahl der gleichzeitig zu befragenden Personen (einer vs. mehrere) differenziert werden. Die häufigste Befragungsart ist die Befragung einer Einzelperson im Rahmen eines standardisierten Interviews, das heißt, die Interviewerin stellt jedem Befragten die gleichen Fragen in gleicher Formulierung und Reihenfolge.

Beobachtung Neben der Befragung gilt die Beobachtung als klassisches Verfahren der Datenerhebung. Bei der Beobachtung unterscheidet Diekmann (2011, S. 564) erstens zwischen teilnehmender und nicht-teilnehmender Beobachtung, zweitens zwischen offener und verdeckter Beobachtung, drittens zwischen Beobachtung im Alltag und Beobachtung im Labor, viertens zwischen strukturierter und unstrukturierter Beobachtung und fünftens zwischen Fremd- und Selbstbeobachtung. Bereits diese fünf Dimensionen deuten auf die Vielzahl an Beobachtungstechniken hin.

Inhaltsanalyse Als dritte Grundtechnik der Datenerhebung befasst sich die Inhaltsanalyse „mit der systematischen Erhebung und Auswertung von Texten, Bildern und Filmen“ (Diekmann 2011, S. 576). Solche Texte können die Programme von Parteien (Klingemann et al. 2006; Bräuninger und Debus 2012) oder auch veröffentlichte Todesanzeigen in der Lokalpresse sein (Schmied 2002, S. 211). Durch eine inhaltsanalytische Auswertung solcher Dokumente kann beispielsweise der ökonomische, soziale, politische und/oder kulturelle Wandel einer Gesellschaft analysiert werden. Mit der Frequenz-, Valenz-, Intensitäts- und Kontingenzanalyse unterscheiden Schnell et al. (2013, S. 398–399) vier Grundformen der Inhaltsanalyse (siehe auch Kromrey et al. 2016, S. 322–324). Die drei zentralen Formen der Datenerhebung – Befragung, Beobachtung und Inhaltsanalyse – werden in Kapitel 11 ausführlicher dargestellt.

3.2.9 Datenaufbereitung und -analyse

Die in der Phase der Datenerhebung gesammelten Informationen müssen zunächst systematisch erfasst, auf Fehler überprüft und schließlich analysiert werden. Insbesondere die sogenannte Datenaufbereitung und Datenprüfung ist ein äußerst aufwändiger Prozess, bei Befragungen werden in der Regel auch sogenannte Plausibilitätstest durchgeführt (z. B. kann die Wohndauer in einem Ort

nicht größer als das Alter des Befragten sein). Bei umfangreichen Datenerhebungen kann die Datenaufbereitung durchaus mehrere Monate dauern. Bei nationalen und internationalen Datensätzen (z. B. Allgemeine Bevölkerungsumfrage der Sozialwissenschaften (ALLBUS) oder European Social Survey (ESS)) werden auch Jahre nach der Erstveröffentlichung der Datensätze noch (kleinere) Fehler gefunden und aktualisierte Datensätze veröffentlicht.

Große Anzahl an Analyseverfahren Nach der (ersten) Datenaufbereitung kann die Datenanalyse beginnen. In den Sozialwissenschaften steht mittlerweile eine große Anzahl an Analyseverfahren zur Verfügung (für eine Übersicht siehe z. B. Behnke et al. 2006; Wolf und Best 2010a; Wenzelburger et al. 2014). Je nach Fragestellung und gesammelten Daten sind geeignete Analyseverfahren auszuwählen. Bei der statistischen Datenanalyse werden Programme wie Stata, SPSS oder R genutzt. In der Regel stehen zunächst deskriptive Analysen im Vordergrund, ehe die formulierten Hypothesen empirisch geprüft werden. Insbesondere in der Phase der Datenanalyse kommt es zwangsläufig zu „Rückkopplungen" zwischen Theorie und Empirie. Schließlich müssen Theorien überarbeitet und Hypothesen neu formuliert werden. Die Phase der Datenaufbereitung und -analyse wird in Kapitel 12 dargestellt.

3.2.10 Publikation

Die Ergebnisse eines Forschungsprojekts werden veröffentlicht. Bei einer Finanzierung des Forschungsprojekts durch einen externen Auftraggeber (z. B. Stiftungen, Verbände, Privatwirtschaft) oder eine öffentliche Förderinstitution (z. B. DFG) ist ein Abschlussbericht obligatorisch.[6] Ein Arbeits- und Ergebnisbericht für die DFG umfasst zehn DIN A4-Seiten und enthält unter anderem die Ausgangsfragen und Zielsetzung des Projekts, die Entwicklung der durchgeführten Arbeiten einschließlich Abweichungen vom ursprünglichen Konzept, ggf. wissenschaftliche Fehlschläge, Probleme in der Projektorganisation oder technischen Durchführung sowie die Darstellung der erreichten Ergebnisse und Diskussion im Hinblick auf den relevanten Forschungsstand.

6 Mit GEPRIS (Geförderte Projekte Informationssystem) stellt die DFG eine Datenbank im Internet zur Verfügung, die über laufende und abgeschlossene Forschungsvorhaben der DFG informiert. Dort findet sich jeweils eine kurze Projektbeschreibung, eine Zusammenfassung der Projektergebnisse und Hinweise auf Publikationen, die im Kontext des Projekts entstanden sind.

Für den wissenschaftlichen Diskurs und Fortschritt ist die Veröffentlichung von Projektergebnissen in Monographien und Fachzeitschriften allerdings deutlich wichtiger. Einen besonderen Stellenwert hat die Veröffentlichung von Aufsätzen in Qualitätszeitschriften. Dabei handelt es sich um Fachzeitschriften mit einem Review-Verfahren. Eingereichte Beiträge werden vor der Veröffentlichung von Gutachtern bewertet. Dabei wird selbstverständlich nicht jeder eingereichte Aufsatz auch publiziert. Die Annahmequote bei deutschsprachigen Fachzeitschriften liegt zwischen 30 und 40 Prozent. Bei hochkarätigen englischsprachigen Fachzeitschriften (z. B. Political Analysis oder American Political Science Review) ist die Ablehnungsquote deutlich höher. Die Publikation von Forschungsergebnissen ist das zentrale Ziel von Wissenschaftlern. Schließlich hängt die Reputation eines Wissenschaftlers von der Anzahl und der Qualität seiner Publikationen ab. Kapitel 13 beschäftigt sich mit der Publikationsphase.

Forschungsthema 4

Die Grundlage für ein erfolgreiches Forschungsprojekt bilden das Forschungsthema und die Forschungsfrage. Die Festlegung des Forschungsthemas und die Formulierung der Forschungsfrage setzen eine intensive Auseinandersetzung mit dem existierenden Forschungsstand voraus. Deshalb werden Kriterien für die Auswahl der Fachliteratur und Recherchetechniken vorgestellt. In diesem Kapitel werden ebenenfalls allgemeine Strategien zur Entwicklung einer Forschungsfrage dargestellt. Zudem werden Kriterien vorgestellt, um die Relevanz einer Forschungsfrage zu bewerten.

4.1 Vom Forschungsthema zur Forschungsfrage

Ein Forschungsprojekt beginnt mit der Festlegung des Gegenstands der Forschung, dem sogenannten Forschungsthema. Das Forschungsthema stellt aber nur den groben Rahmen eines Forschungsprojekts dar und bildet nur die Grundlage für die Entwicklung einer Forschungsfrage. Plümper (2012, S. 16) betont:

> *„Selbstverständlich geht es nicht ohne Thema, aber zu denken, ein Thema sei bereits eine ausreichende Arbeitsgrundlage und sei es auch nur, um die Literatur zu sichten, ist schlicht falsch und häufig der Beginn einer unnötigen Leidenszeit.“*

Das Thema grenzt den Inhalt eines Forschungsprojekts lediglich lose ab. Sozialwissenschaftliche Themen wie „Soziale Ungleichheit“, „Vertrauen“ oder „Sozialer Wandel“ sind zu umfangreich, um sie vollständig in einem Forschungsprojekt oder gar einer Haus- oder Abschlussarbeit bearbeiten zu können. Deshalb muss das Thema eingegrenzt werden, um es im Rahmen eines (zeitlich befristeten) Forschungsprojekts bearbeiten zu können.

© Springer Fachmedien Wiesbaden GmbH, ein Teil von Springer Nature 2018

M. Tausendpfund, *Quantitative Methoden in der Politikwissenschaft*,
Grundwissen Politik, https://doi.org/10.1007/978-3-658-20698-7_4

Thema eingrenzen: Wahlbeteiligung

Die empirische Wahlforschung gehört zu den theoretisch und methodisch bestentwickelten Disziplinen der Politikwissenschaft (für eine Übersicht siehe z. B. Schmitt-Beck 2012; Falter und Schoen 2014). Wird Wahlforschung als Thema gewählt, kann sich die Wissenschaftlerin mit der Wahlbeteiligung oder mit dem Wahlverhalten beschäftigen. Entscheidet sich der Forscher für die Wahlbeteiligung, dann muss er im nächsten Schritt klären, ob er sich mit der lokalen, nationalen oder europäischen Wahlbeteiligung auseinandersetzen möchte. Denkbar, aber auch deutlich umfangreicher, wäre natürlich auch eine vergleichende Betrachtung der Wahlbeteiligung auf verschiedenen Ebenen. Möglich wäre aber auch ein Forschungsprojekt, das die zeitliche Entwicklung der Wahlbeteiligung in den Blick nimmt. Schließlich muss die Forscherin auch klären, ob sie Wahlbeteiligung auf der Mikro- oder Makroebene untersuchen möchte.

Das Beispiel zeigt: Bereits die erste Eingrenzung des Themas ist mit zahlreichen Entscheidungen verbunden. Diese Entscheidungen sind erforderlich und werden von Gutachtern und/oder Betreuern auch akzeptiert, solange sie (nachvollziehbar) begründet sind und sich der Forscher darüber im Klaren ist, welche Konsequenzen diese Entscheidungen für das Forschungsprojekt haben.

Thema mit Kernbegriffen beschreiben Da das Thema den Forschungsgegenstand nur lose abgrenzt, empfiehlt Plümper (2012, S. 16) das gewählte Thema mit zwei oder drei Kernbegriffen zu beschreiben. „Wahlbeteiligung auf lokaler Ebene" beschreibt das Thema deutlich besser als „Wahlforschung". Das Forschungsthema kann aber noch präziser formuliert werden: „Lokale Unterschiede der Wahlbeteiligung in der Bundesrepublik Deutschland" grenzt beispielsweise den Forschungsgegenstand deutlich klarer ab, da zum einen das konkrete Problem (Unterschiede der lokalen Wahlbeteiligung) und zum anderen das räumliche Gebiet (Bundesrepublik Deutschland) benannt werden. Je enger das Forschungsprojekt durch die Kombination aus Kernbegriffen beschrieben wird, desto leichter wird es fallen, die Grenzen des Projekts abzustecken (z. B. Plümper 2012, S. 17; Huemer et al. 2012, S. 54–60; Schlichte und Sievers 2015, S. 21–27).

Frage beantworten Jedes Forschungsprojekt muss eine konkrete Frage beantworten oder ein Problem lösen (z. B. Diekmann 2011, S. 187; Schnell et al. 2013, S. 3). Die Fragestellung oder das konkret zu bearbeitende Problem ergibt sich in der Re-

gel aus dem bisherigen Wissensstand und/oder den im jeweiligen Fachgebiet noch offenen Fragen. Deshalb ist für die Entwicklung der konkreten Fragestellung und die Zielsetzung des Forschungsprojekts die Auseinandersetzung mit dem bisherigen Forschungsstand erforderlich. Je konkreter die Forschungsfrage formuliert ist, desto zielgerichteter kann die Literaturrecherche durchgeführt werden. Häufig wird gerade bei der Entwicklung eines Forschungsprojekts die bereits existierende Literatur zu einem Thema unterschätzt. Diese Fehleinschätzung ist auf das geringe Wissen bei der Projektentwicklung zurückzuführen, da es sich zu Beginn der Projektentwicklung noch um einen weitgehend unbekannten Sachverhalt handelt. In Wirklichkeit ist aber jedes noch so enge Thema deutlich komplexer als ursprünglich angenommen und der Forschungsstand deutlich umfangreicher als erwartet. Der komplette Forschungsstand zur „Wahlbeteiligung" ist praktisch nicht aufzuarbeiten; ein erster Überblick zur lokalen Wahlbeteiligung sollte dagegen bereits nach der Lektüre von 10 bis 15 Aufsätzen möglich sein.

Relevanz von Fachzeitschriften Ein Forschungsprojekt baut auf die bereits gesicherten wissenschaftlichen Erkenntnisse auf und versucht, offene Fragen und/oder Probleme zu identifizieren. Schließlich ist Wissenschaft ein gemeinschaftliches Projekt und zeichnet sich dadurch aus, dass „viele Menschen gleichzeitig und nacheinander über Phänomene und Dinge nachdenken und Untersuchungen anstellen, ihre Vorgehensweisen und ihre Ergebnisse vergleichen und aus ihrem Wissensfundus heraus Theorien aufstellen und Schlüsse ziehen und diese auch wieder in Frage stellen" (Schwarzer 2001, S. 142). Wissenschaftliches Arbeiten besteht deshalb aus der Reflexion über theoretische Konzepte, dem Nachdenken über die methodische Vorgehensweise und der sorgfältigen Entwicklung von Untersuchungsstrategien. Dabei sind Fehler unvermeidlich und eine kritische Auseinandersetzung mit Theorien, Konzepten und empirischen Ergebnissen erforderlich. Diese Auseinandersetzung findet auf wissenschaftlichen Tagungen sowie in Fachzeitschriften statt.

Forschungsstand auf Basis von Fachaufsätzen aufarbeiten

Aufsätze in Fachzeitschriften sind heute das zentrale Kommunikationsmedium der Wissenschaft. In den Naturwissenschaften gilt dies schon seit vielen Jahren, aber auch in den Sozialwissenschaften gelten Fachzeitschriften heute als unverzichtbar. Neuere Forschungsergebnisse werden fast immer in wissenschaftlichen Fachzeitschriften publiziert (siehe Kapitel 13). Tages- und Wochenzeitungen (z. B. FAZ, Zeit) sind keine Fachzeitschriften.

Die explizite Anbindung an den bisherigen Forschungsstand „entspricht auch dem kumulativen Charakter des Forschungsprozesses, also der Idee, dass Beiträge unterschiedlicher Forscher aufeinander aufbauen und somit als Ganzes zum besseren Verständnis der Materie beitragen“ (Lehnert et al. 2007, S. 42). Eine gründliche Sichtung der vorhandenen theoretischen Ansätze und empirischen Ergebnisse des Forschungsthemas stellt deshalb die Ausgangsbasis für ein Forschungsvorhaben dar (Ebbinghaus 2009, S. 484). Ein solcher Überblick („state-of-the-art“) steht am Anfang eines Forschungsprojekts und spielt auch bei Forschungsanträgen eine zentrale Rolle (Schwarzer 2001, S. 149–150). Im Leitfaden der Deutschen Forschungsgemeinschaft (2014, S. 3) heißt es:

> *„Legen Sie bei Neuanträgen den Stand der Forschung bitte knapp und präzise in seiner unmittelbaren Beziehung zum konkreten Vorhaben dar. In dieser Darstellung sollte deutlich werden, wo Sie Ihre eigenen Arbeiten eingeordnet sehen und zu welchen der anstehenden Fragen Sie einen eigenen, neuen und weiterführenden Beitrag leisten wollen. Die Darstellung muss ohne Hinzuziehen weiterer Literatur verständlich sein.“*

! Deshalb: Ein Forschungsprojekt beginnt mit der Auseinandersetzung mit der existierenden Fachliteratur zum Forschungsthema. Nur wer die zentralen theoretischen Ansätze und empirischen Befunde kennt, kann Probleme im Wissensstand identifizieren und neue Forschungsfragen formulieren. Um sich einen (ersten) Überblick zum Forschungsstand zu verschaffen, sind nach Plümper (2012, S. 16) die 30 bis 40 wichtigen Artikel und die drei bis acht relevanten Bücher zu lesen. Nach Westle (2009a, S. 126) sollten die Klassiker sowie die aktuelle Literatur der vergangenen zehn Jahre zur Thematik bekannt sein.

Forschungsstand präsentieren Der Forschungsstand zu einem Thema stellt den Ausgangspunkt eines Forschungsprojekts dar und wird in einem Literaturbericht präsentiert. Eine solche Literaturübersicht soll dabei den Forschungsstand strukturieren und verdichtet präsentieren. Dieser Literaturbericht kann einerseits ein Bestandteil einer größeren Arbeit (z. B. innerhalb von Qualifikationsarbeiten) oder andererseits eine eigenständige Übersicht zu einem bestimmten Thema sein. Innerhalb einer größeren Arbeit wird der Literaturbericht meist in einem eigenen Kapitel oder als Unterkapitel präsentiert (z. B. Stand der Forschung), gelegentlich werden Forschungsstand und Hypothesenformulierung auch verknüpft. Vor dem Hintergrund der zunehmenden Zahl an wissenschaftlichen Publikationen werden aber auch in Fachzeitschriften immer häufiger eigenständige Literaturberichte veröffentlicht, um Wissenschaftlern einen schnellen Überblick über aktuelle Entwicklungen in einer Disziplin zu geben.

Nach Bryman (2016, S. 90–119; siehe auch Ressing et al. 2009; Prexl 2017, S. 15–21) lassen sich allgemein zwei Arten von Literaturübersichten unterscheiden: narrative review (traditionelle Literaturübersicht) und systematic review (systematische Literaturübersicht). Die beiden Arten von Literaturübersichten unterscheiden sich in erster Linie in der Art und Weise der Literaturrecherche. Bei einer traditionellen Literaturübersicht erfolgt die Auswahl der berücksichtigten Texte eher unsystematisch. Bei einer systematischen Literaturübersicht werden bereits vor der Literaturrecherche Kriterien fest- und offengelegt, welche Literatur berücksichtigt werden soll.

Traditionelle Literaturübersicht Die traditionelle Literaturübersicht bietet einen breiten Überblick zu einem bestimmten Forschungsthema. Der Forschungsstand wird in einer komprimierten und übersichtlichen Darstellung präsentiert sowie Gemeinsamkeiten und Unterschiede der berücksichtigen Texte aufgezeigt (z. B. Unterschiede in der Operationalisierung, widersprüchliche empirische Ergebnisse). Diese Form der Literaturübersicht wird häufig genutzt, um eine Forschungslücke zu identifizieren und die eigene Forschung einzuordnen. Allerdings basiert die traditionelle Literaturübersicht in der Regel auf einer selektiven Literaturauswahl. Dadurch können zum einen „unliebsame" Texte aussortiert werden und zum anderen ist die Literaturrecherche nicht reproduzierbar (Bryman 2016, S. 91–98; Prexl 2017, S. 17–18).

Systematische Literaturübersicht Bei einer systematischen Literaturübersicht werden möglichst alle relevanten Texte zu einem bestimmten Thema berücksichtigt. Dabei werden vor der Literaturrecherche bestimmte Kriterien definiert, um wissenschaftliche Texte ein- oder auszuschließen. Diese Kriterien werden offengelegt und auch die Wahl der Literaturdatenbanken sowie der Suchbegriffe werden offen kommuniziert. Dadurch ist intersubjektiv nachvollziehbar, welche Literatur berücksichtigt wird oder nicht. Insbesondere die Grundlage des Literaturüberblicks ist dadurch neutraler und weniger subjektiv gefärbt. Auch bei einer systematischen Literaturübersicht soll der Forschungsstand selbstverständlich komprimiert und übersichtlich präsentiert werden. Die Darstellung orientiert sich dabei an inhaltlichen Kriterien (z. B. Erklärungsansätze).

! Grundsätzlich ist sowohl eine traditionelle als auch eine systematische Literaturübersicht mehr als eine einfache Aneinanderreihung von Inhaltsangaben verschiedener Texte (Prexl 2017, S. 23). Döring und Bortz (2016, S. 163) betonen:

> *„Eine lineare Darstellung und pure Nacherzählung ausgewählter Forschungsbeiträge in der Form: In Studie A wurde das untersucht, Studie B hat das herausgefunden, Auto-*

rin C hat das festgestellt, Autor D hat das gesagt etc. ist als Präsentation des Forschungsstandes unbrauchbar. Hier fehlt die Leistung einer sinnvollen Strukturierung und Bewertung.“

Für eine Literaturübersicht muss der Forschungsstand nach inhaltlichen Kriterien strukturiert werden, zum Beispiel nach Erklärungsansätzen und/oder analytischen Ebenen (Mikroebene versus Makroebene). Für die Aufarbeitung des (meist sehr umfassenden) Forschungsstands bieten sich häufig Literatursynopsen an.

Wo finde ich Literaturübersichten?

In Qualifikationsarbeiten (Promotionen und Habilitationen) finden sich meist sehr ausführliche Darstellungen zum Forschungsstand. In Fachzeitschriften wie der Politischen Vierteljahresschrift (PVS) oder der Zeitschrift für Vergleichende Politikwissenschaft (ZfVP) werden regelmäßig Literaturberichte veröffentlicht, in denen ein breiter Überblick über den internationalen Forschungsstand, aktuelle Entwicklungen und die hierzu erschienene Literatur in einem bestimmten Forschungsfeld gegeben wird. Empfehlenswerte Literaturübersichten finden sich auch im Annual Review of Political Science und im Annual Review of Sociology (www.annualreviews.org). Eine (neuere) Form der Literaturübersichten bieten auch sogenannte Living Reviews, die häufig in Open-Access-Zeitschriften veröffentlicht werden (z. B. Living Reviews in European Governance).

Die Grundlage einer Literaturübersicht ist die ausgewählte Fachliteratur. Im folgenden Abschnitt werden deshalb zunächst Kriterien zur Auswahl von wissenschaftlicher Literatur formuliert, ehe Strategien zur Literaturrecherche vorgestellt werden.

4.2 Literatur und Literaturrecherche

Jede wissenschaftliche Arbeit – von der Hausarbeit über die Abschlussarbeit bis hin zur Promotion und der Veröffentlichung einer Studie in einer Fachzeitschrift – ist in den aktuellen Forschungsstand eingebettet (Plümper 2012, S. 43–53; Döring und Bortz 2016, S. 163). Deshalb ist für die Entwicklung und Bearbeitung einer gehaltvollen Forschungsfrage die Auseinandersetzung mit der sozialwissenschaft-

lichen Fachliteratur obligatorisch. Dieser Abschnitt informiert zunächst über Kriterien zur Auswahl von wissenschaftlicher Literatur, ehe Strategien zur Literaturrecherche vorgestellt werden. Die Ausführungen können die Lektüre von einführenden Büchern zum wissenschaftlichen Arbeiten nicht ersetzen. Für diesen Zweck sei an dieser Stelle unter anderem auf die Veröffentlichungen von Berninger et al. (2017), Ebster und Stalzer (2017) sowie Prexl (2017) verwiesen.

4.2.1 Was ist wissenschaftliche Fachliteratur?

In einer wissenschaftlichen Arbeit sollen Sie zeigen, dass Sie einerseits Fachbegriffe und Fachsprache sicher verwenden und andererseits begrifflich klar, systematisch, kritisch-reflektiert und ausgewogen argumentieren können. Da Sie im Kontext des Studiums eine wissenschaftliche Arbeit schreiben, sollte ausschließlich wissenschaftliche Fachliteratur genutzt werden. Deshalb sollten Sie zunächst prüfen, ob es sich bei einem vorliegenden Text um einen Informationstext oder um Fachliteratur handelt.

Informationstexte versus Fachliteratur Bei klassischen Nachschlagewerken fällt diese Entscheidung noch leicht. Enzyklopädien, Lexika oder auch der Duden enthalten Basisinformationen wie biographische und historische Fakten sowie knappe Informationstexte. In einzelnen Texten finden sich auch Hinweise auf weiterführende Literatur. Solche Nachschlagewerke sind keine wissenschaftliche Fachliteratur und werden deshalb nicht in wissenschaftlichen Arbeiten (z. B. Hausarbeiten) zitiert (Stykow et al. 2010, S. 196).

Selbstverständlich sind Informationstexte hilfreich, um sich mit einem noch unbekannten Sachverhalt vertraut zu machen. Allerdings zielen solche Informationstexte in erster Linie darauf ab, basale Informationen zu einem Thema zu vermitteln. Sie können den Einstieg in eine Thematik erleichtern, aber niemals wissenschaftliche Fachliteratur ersetzen. Informationstexte haben selbstverständlich ihre Berechtigung, aber in einer wissenschaftlichen Arbeit haben sie nichts verloren. Bei der zitierfähigen und zitierwürdigen Fachliteratur lassen sich grob drei Gruppen unterscheiden: Monographien, Sammelbände und Fachzeitschriften. Diese drei Gruppen werden im Folgenden knapp dargestellt.

Monographien Bei einer Monographie handelt es sich um ein Buch, das sich einem Sachverhalt oder einer bestimmten Fragestellung widmet. In der Regel sind ein, manchmal aber auch mehrere Personen für den Inhalt einer Monographie verantwortlich. Nach Stykow et al. (2010, S. 194) haben wissenschaftliche Monographien „den Anspruch, Wissens- und Forschungsstand zum Zeitpunkt der Ver-

Kriterien für die Literaturbewertung: Zitierfähigkeit und Zitierwürdigkeit

Nach Ebster und Stalzer (2017, S. 66) sowie Prexl (2017, S. 107–109) lassen sich zwei generelle Kriterien für die Bewertung von Fachliteratur unterscheiden: Zitierfähigkeit und Zitierwürdigkeit. Zitierfähigkeit bezieht sich auf die allgemeine Zugänglichkeit des Textes. Ein zentrales Kriterium von Wissenschaftlichkeit ist die intersubjektive Nachvollziehbarkeit. Dies setzt voraus, dass alle verwendeten Quellen auch anderen Forschern zugänglich sind. Bei Büchern und Fachaufsätzen ist die Zitierfähigkeit gegeben, bei grauer Literatur ist die Zitierfähigkeit zumindest eingeschränkt. Zitierwürdigkeit bezieht sich auf die wissenschaftliche Qualität eines Textes. Die Zitierwürdigkeit hängt sicherlich von der Güte des spezifischen Textes ab, doch in Anlehnung an Ebster und Stalzer (2017, S. 67) gibt es generelle Kriterien für die Bewertung der Zitierwürdigkeit. Bei wissenschaftlichen Qualifikationsarbeiten (Promotionen und Habilitationen) kann die Zitierwürdigkeit in der Regel vorausgesetzt werden. Dies gilt auch für Veröffentlichungen mit Qualitätssicherung (z. B. Peer-Review-Verfahren). Bei Veröffentlichungen ohne Literaturverzeichnis bzw. ohne Quellenangaben (z. B. sogenannte Praktikerbücher) ist die Zitierwürdigkeit in der Regel nicht gegeben.

öffentlichung darzustellen“. Monographien können grob in Lehrbücher und Studien eingeteilt werden.

Lehrbücher bieten einen ersten Überblick zu einem Thema. In der empirischen (quantitativen) Sozialforschung gelten die Veröffentlichungen von Diekmann (2011) oder Schnell et al. (2013) als Standardwerke. Solche Lehrbücher enthalten in der Regel umfangreiche Informationen, um sich mit einem Sachverhalt vertraut zu machen. Es werden Grundkenntnisse vermittelt, die bei Studierenden in höheren Semestern vorausgesetzt werden. Ein Lehrbuch bietet damit den Einstieg in eine Thematik und ist geeignet, um weitere Texte zum Thema zu erschließen (z. B. durch die Literaturverweise in diesem Lehrbuch).

Von Lehrbüchern sind Studien zu unterscheiden, die Forschungsergebnisse zu einem Thema veröffentlichen. Häufig handelt es sich dabei um Qualifikationsarbeiten (Promotionen oder Habilitationen), aber auch um Veröffentlichungen, die im Kontext eines Forschungsprojektes entstanden sind. Bei solchen Studien handelt es sich meist um eine sehr konkrete Forschungsfrage, die dafür aber sehr ausführlich behandelt wird. Deshalb bieten gerade (aktuelle) Studien eine wahre Fundgrube für die Literaturrecherche und die Beschäftigung mit einer konkreten

Schlüsselwerke der Sozialwissenschaften

Einzelne Monographien gelten als sogenannte Schlüsselwerke, die in besonderem Maße die Entwicklung des Fachbereichs geprägt haben. Die politische Kulturforschung bezieht sich beispielsweise immer wieder auf die Studie zu „Civic Culture" von Almond und Verba (1963), in den internationalen Beziehungen hat der „Kampf der Kulturen" von Huntington (1996) besondere Aufmerksamkeit erhalten und in der Parlamentarismusforschung gilt die Studie von Lijphart (1999) zu den „Patterns of Democracy" als besonders wegweisend. Ein Überblick zu Schlüsselwerken der Politikwissenschaft gibt der Band von Kailitz (2007). Papcke und Oesterdiekhoff (2001) haben einen ähnlichen Band herausgegeben, der die Schlüsselwerke der Soziologie enthält. Salzborn (2016) bietet ein Überblick über 100 Klassiker der Sozialwissenschaften. Selbstverständlich kann die Zusammenfassung von Schlüsselwerken die Lektüre der einzelnen Monographien nicht ersetzen, sie kann aber eine wertvolle Orientierungs- und Bewertungshilfe sein.

Forschungsfrage. In Kapiteln wie „Stand der Forschung" oder „Forschungsstand" wird die meist sehr umfängliche Literatur mit Blick auf die Forschungsfrage aufbereitet.

Sammelbände Eine Publikation, die mehrere Beiträge von verschiedenen Autoren enthält, wird als Sammelband bezeichnet. Das zentrale Merkmal eines Sammelbands ist die mehr oder weniger vorhandene thematische Verwandtschaft der einzelnen Beiträge, die unter dem „Dach" eines Bands zu einer umfassenden Publikation gebündelt werden. Die Verantwortung für den Band, das heißt, die Auswahl der Autoren und der Aufnahme der Beiträge in den Band, trägt der oder tragen die Herausgeber. Sammelbände erscheinen häufig im Kontext von Tagungen oder Kongressen, aber auch Lehrbücher und Studien werden häufig als Sammelbände veröffentlicht. Das Handbuch von Baur und Blasius (2014a) bietet in 88 (!) Beiträgen einen Überblick zur empirischen Sozialforschung, das Handbuch von Wolf und Best (2010a) stellt Verfahren der sozialwissenschaftlichen Datenanalyse vor.

Sammelbände haben den Vorteil, dass mehrere Experten eines Teilgebiets einer Disziplin (z. B. Nonresponse) Beiträge verfassen. Diesem Vorteil stehen allerdings auch zwei Nachteile gegenüber. Zum einen leidet manchmal die inhaltliche Kohärenz und auch inhaltliche Dopplungen sind nicht immer zu vermeiden. Zentrale Begriffe müssen beispielsweise mehrmals definiert werden, da sie in un-

Was ist Open Access?

Open Access meint den uneingeschränkten und kostenlosen Zugang zu wissenschaftlichen Informationen. Ziel ist neben einem möglichst barrierefreien Zugang zu wissenschaftlichen Materialien, die Maximierung der Verbreitung von wissenschaftlicher Literatur. Gründe für Open Access sind unter anderem der öffentliche Zugang zu öffentlich finanzierten Forschungsergebnissen, die erhöhte Sichtbarkeit und Zitierfähigkeit von Dokumenten, die Förderung der Forschungseffizienz und die gute Auffindbarkeit der Dokumente über Suchmaschinen und Nachweisdienste.

terschiedlichen Kontexten verwendet werden. Zum anderen kann auch die Qualität der einzelnen Beiträge in einem Sammelband variieren.

Aufsätze in Fachzeitschriften Aufsätze in Fachzeitschriften sind heute das zentrale Kommunikationsmedium der Wissenschaft. In den Naturwissenschaften gilt dies schon seit vielen Jahren, aber auch in den Sozialwissenschaften gelten Fachzeitschriften heute als unverzichtbar. Die Bedeutung von Fachzeitschriften beschreibt Bär (2002, S. 11):

> *„Was für den Bäcker das Brot, für den Metzger die Wurst, das ist für den Wissenschaftler das Paper. Ein Paper ist ein Aufsatz in einer wissenschaftlichen Zeitschrift, der neue Forschungsergebnisse beschreibt."*

Zwar haben auch heute noch Monographien und Sammelbände ihre Berechtigung, aber in Zeiten von kumulativen Dissertationen gelten Aufsätze in Fachzeitschriften heute mehr denn je als zentrale Publikationsform. Neuere Forschungsergebnisse werden fast immer in wissenschaftlichen Fachzeitschriften publiziert. In Deutschland sind das beispielsweise die Politische Vierteljahresschrift (PVS) oder die Kölner Zeitschrift für Soziologie und Sozialpsychologie (KZfSS). Die wirklich wichtigen Fachzeitschriften erscheinen allerdings in englischer Sprache – zum Beispiel American Political Science Review, European Journal of Political Research, Annual Review of Sociology oder American Sociological Review. Der Abschnitt 13.2 bietet eine Übersicht über sozialwissenschaftliche Fachzeitschriften. In Abschnitt 14.2 werden wichtige Fachzeitschriften vorgestellt, die sich mit den Methoden der empirischen Sozialforschung beschäftigen.

Aufsätze in Fachzeitschriften sind keine einführende Literatur, sondern knüpfen an aktuelle Debatten und Diskussionen an. Bei Fachzeitschriften wird am

Was ist Digital Object Identifier (DOI)?

DOI steht für Digital Object Identifier und ist ein eindeutiger und dauerhafter digitaler Identifikator für physische, digitale oder abstrakte Objekte. Viele wissenschaftliche Verlage nutzen DOI-Angaben zur dauerhaften Adressierung der Veröffentlichungen. Der DOI-Name führt dabei direkt zum Speicherort des bezeichneten Objekts (z. B. Fachaufsatz, Monografie, Sammelband) und ist etwa mit ISBN oder ISSN vergleichbar. Um einen DOI-Namen zur zugehörigen Internetadresse aufzulösen gibt es verschiedene Möglichkeiten. Zum einen kann der DOI-Name auf der Internetseite www.doi.org eingegeben werden. Zum anderen kann auch das Präfix „https://dx.doi.org/" vor die DOI-Nummer gestellt werden. Häufig führt auch die Eingabe der DOI-Nummer in die URL-Adresse des Browsers zum gewünschten Dokument. Die gefundene Webseite ist aber häufig nicht das referenzierte Dokument selbst, sondern eine Seite über das entsprechende Dokument, mit der Möglichkeit, es herunterzuladen oder zu erwerben. Für den Download ist unter Umständen eine VPN-Verbindung erforderlich.

deutlichsten, dass Wissenschaft ein kumulativer Prozess ist. Jede einzelne wissenschaftliche Publikation zielt darauf ab, das „Wissen" ein kleines Stück zu erweitern. Häufig beziehen sich Autoren eines Beitrags auch auf andere Kollegen, die zu ähnlichen Fragestellungen publiziert haben. Für die Auseinandersetzung mit dem (aktuellen) Forschungsstand sind Fachzeitschriften unverzichtbar. Deshalb setzen Aufsätze in Fachzeitschriften oft auch Vorkenntnisse voraus, um der Auseinandersetzung bzw. Argumentation folgen zu können. Die Vorteile von Aufsätzen in Fachzeitschriften liegen in der Aktualität der Forschungsergebnisse, der Vielfalt der Veröffentlichungen und in der Kürze der Beiträge. Die in relativ kurzen Abständen erscheinenden Fachzeitschriften (meist zwischen vier bis sechs Ausgaben im Jahr) bieten die Möglichkeit in kurzen Abständen auf neue Themen, Debatten oder Kontroversen zu reagieren.

4.2.2 Literaturrecherche

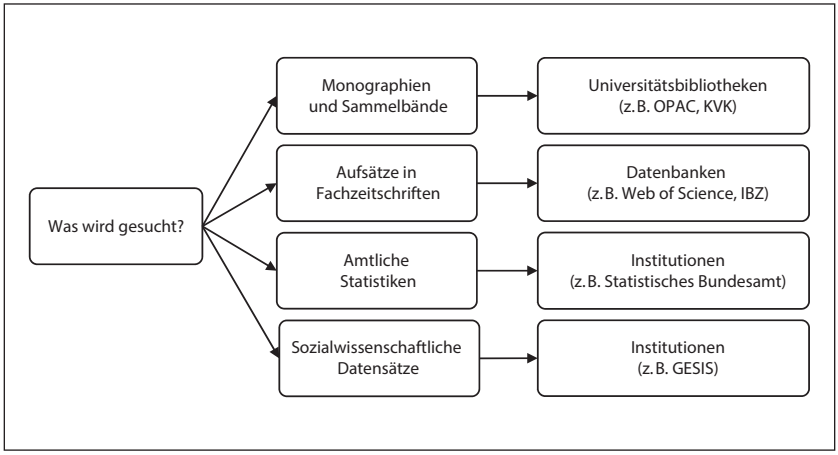
Vor jeder Literaturrecherche stehen zwei Fragen: Was suche ich? Wo suche ich? Im ersten Schritt muss zunächst einmal das Forschungsthema grob festgelegt werden. Dabei ist noch keine konkrete Forschungsfrage erforderlich, aber es sollte klar sein, zu welchem Thema überhaupt Literatur gesucht wird. In Abhängigkeit

der ersten Suchergebnisse muss das Thema aber relativ zügig eingeschränkt und die Fragestellung präzisiert werden, um einen Überblick zum Forschungsstand zu erhalten.

Recherchemittel kombinieren Erst, wenn Sie eine ungefähre Vorstellung Ihrer Forschungsfrage haben, können Sie mit einer systematischen Literaturrecherche beginnen. Als systematische Suche bezeichnen Stykow et al. (2010, S. 202), „wenn man alle verfügbaren Bibliotheks- und elektronischen Recherchemittel konsequent verwendet, um sämtliche Erscheinungsformen wissenschaftlicher Information abzudecken." Mit anderen Worten: Es gibt nicht „ein" Recherchemittel, sondern es müssen verschiedene Recherchemittel kombiniert werden, um die relevante Forschungsliteratur zu einem Thema zu identifizieren.

Nach der (ersten) Eingrenzung der Forschungsfrage stellt sich die Frage, wo nach geeigneter Literatur gesucht werden soll. Dabei empfiehlt sich, die Frage anders zu formulieren: Welche Textform suchen Sie? Wie Abbildung 5 zeigt, unterscheiden sich die verschiedenen Recherchestrategien in erster Linie in Abhängigkeit der gesuchten Textform. Zwar gibt es einzelne Recherchetools (z. B. Google Scholar), mit denen Bücher, Sammelbände und einzelne Fachaufsätze gefunden werden können. Für einen umfassenden Überblick zur Forschungsliteratur zu einem spezifischen Thema ist es aber zwingend erforderlich, mehrere Recherchestrategien zu kombinieren.

Abbildung 5 Recherchestrategien im Überblick



Quelle: Eigene Darstellung

! Für die (erste) Recherche nach Monographien und Sammelbänden bietet sich der Online-Katalog (OPAC) der Universitätsbibliothek an. In den Online-Katalogen sind Monographien und Sammelbände sowie die Titel der Fachzeitschriften verzeichnet. Sie finden die Angaben der Herausgeber zu einem Sammelband oder den Titel einer Fachzeitschrift (z. B. Zeitschrift für Vergleichende Politikwissenschaft), nicht aber die einzelnen Beiträge eines Sammelbands oder einer Fachzeitschrift.

Für die effiziente Recherche nach Artikeln in Fachzeitschriften sind Fachdatenbanken ideal. Solche Datenbanken verzeichnen Autor, Titel des Aufsatzes, Zeitschriftentitel und Erscheinungsjahr. Häufig werden Zusammenfassungen des jeweiligen Texts mitgeliefert. Eine der wichtigsten Fachdatenbanken ist das sogenannte Web of Science, das eng mit dem SSCI eng verknüpft ist.

Wer nach Statistiken sucht (z. B. Wahlbeteiligung), sollte nicht Wikipedia als Quelle ausweisen. Zwar können auch amtliche Daten durchaus fehlerhaft sein (z. B. Tausendpfund und Braun 2008), aber als Grundlage sind offizielle Statistiken aus erster Hand Veröffentlichungen aus zweiter oder dritter Hand vorzuziehen. Zumal sich bei Sekundärquellen häufig noch weitere Fehler einschleichen, da die Datengrundlage häufig die amtlichen Statistiken sind.

Mit Datensätzen sind Forschungsdaten gemeint, die von anderen Institutionen oder Personen gesammelt wurden. Es handelt sich nicht um Texte, sondern um Daten des Politbarometers oder der Allgemeinen Bevölkerungsumfrage der Sozialwissenschaften (ALLBUS). Diese Datensätze stehen für empirische Sekundäranalysen zur Verfügung und können mit entsprechender Software (z. B. SPSS, Stata) ausgewertet werden (z. B. für eine eigene empirische Seminararbeit). GESIS ist die größte deutsche Infrastruktureinrichtung für die Sozialwissenschaften (www.gesis.org). Wer nach Forschungsdaten sucht, der sollte seine Suche mit dem Datenbestandskatalog der GESIS beginnen.

Monographien und Sammelbände Für die (erste) Suche nach wissenschaftlichen Büchern (Monographien und Sammelbände) bietet sich – wie oben dargestellt – der elektronische Bibliothekskatalog (OPAC) einer Universitätsbibliothek an. Die Recherche im OPAC einer Bibliothek ist auf den Bestand der jeweiligen Bibliothek begrenzt. Selbstverständlich variiert der Bestand von Bibliothek zu Bibliothek. Der Bestand der Bibliothek der FernUniversität in Hagen umfasst über 800 000 Titel (www.ub.fernuni-hagen.de), der Bestand der Deutschen Nationalbibliothek (www.dnb.de) betrug Anfang 2017 etwa 32,7 Millionen Einheiten und das Angebot der Library of Congress (www.loc.gov) wird auf etwa 164 Millionen Posten beziffert (darunter 38 Millionen katalogisierte Bücher).

Eine Möglichkeit die Grenzen der lokalen Universitätsbibliothek zu überschreiten, sind sogenannte „regionale“ Zentralkataloge. Die Bibliotheken sind in regionalen Verbünden organisiert, welche die Katalogisierung und Fernleihe koordinieren. Der Karlsruher Virtuelle Katalog (KVK) ist eine Meta-Suchmaschine, welche die eingetragenen Suchanfragen gleichzeitig an mehrere Bibliothekskataloge weiterleitet und entsprechende Trefferlisten anzeigt. Der KVK ist also keine Datenbank, sondern nutzt die Datenbanken anderer Institutionen. Der Vorteil des KVK ist, dass gleichzeitig mehrere Datenbanken (bei Bedarf auch Buchhandlungen) durchsucht werden können. Die Datenbanken können dabei für jede Suchanfrage individuell festgelegt werden. Weitere Informationen finden sich unter http://e.feu.de/kvk.

Aufsätze in Fachzeitschriften Da Aufsätze in Fachzeitschriften mittlerweile eine herausgehobene Bedeutung in den Sozialwissenschaften haben, sollten Sie auch für Ihre Hausarbeit auf entsprechende Fachaufsätze zurückgreifen. Die entsprechende Recherche erfordert aber eine andere Recherchestrategie als die Suche nach Büchern: In den elektronischen Bibliothekskatalogen sind in der Regel die verfügbaren Fachzeitschriften, nicht aber die einzelnen Aufsätze verzeichnet (eine Ausnahme ist der Bibliothekskatalog des Deutschen Bundestags unter http://opac.bibliothek.bundestag.de). Sie könnten jetzt mühsam die einzelnen Inhaltsverzeichnisse der Fachzeitschriften durchgehen oder Sie nutzen Fachdatenbanken.

Fachdatenbanken extrahieren zentrale Informationen der Aufsätze wie Autor, Titel und Zeitschrift sowie eine kurze Zusammenfassung des Aufsatzes (Abstracts) in eine Datenbank. In den meisten Fachdatenbanken finden sich also nur die bibliographischen Angaben zu einem Aufsatz und nicht der komplette Text. Dies ist aber weitgehend unproblematisch, da sich bei den Angaben häufig ein Link zum Volltext finden. Das heißt, Sie können auf den Volltext zugreifen, sofern Ihre Universität die Fachzeitschrift abonniert hat.

Unter den sozialwissenschaftlichen Fachdatenbanken nimmt der Social Science Citation Index (SSCI) eine herausragende Stellung ein. Zum einen werden sehr viele Fachzeitschriften ausgewertet und zum anderen überzeugt die Datenbank mit sehr eleganten Recherchemöglichkeiten. Die Recherche im SSCI erfolgt über die Plattform „Web of Science“. Vereinfacht formuliert, greift der Nutzer über das Web of Science auf die entsprechenden Datenbanken der Firma Thomson Reuters zurück. Die Bedeutung von SSCI bzw. Web of Science für die Sozialwissenschaften dokumentiert eine Aussage von Plümper (2012, S. 46):

„Die Mehrzahl der Bibliotheken besitzt den SSCI in elektronischer Form. Trifft das für Ihre Bibliothek nicht zu, denken Sie – falls die Zeit dafür noch ausreicht – über einen

Wechsel der Universität nach. Das mag übertrieben klingen, aber ich meine das keineswegs nur als Witz: Schlechte wissenschaftliche Arbeitsbedingungen liefern Ihnen einen wichtiges Indiz für eine schlechte Ausbildung.“

Auf den ersten Blick ähnelt die Suche im Web of Science einer Suche im Bibliothekskatalog, aber die Recherchemöglichkeiten sind deutlich umfassender. So bietet das Web of Science vielfältige Möglichkeiten, die gefundenen Treffer zu sortieren. Selbstverständlich können die Treffer nach ihrer Aktualität ausgegeben werden (also von neu zu alt). Interessant ist sicher aber auch, wie oft ein bestimmter Beitrag zitiert wurde (Times Cited). So können relativ schnell zentrale Veröffentlichungen zu einer Fragestellung identifiziert werden. Zudem kann ein „Citation Alert“ erstellt werden.

Der SSCI bzw. das Web of Science ist eine zentrale Recherchemöglichkeit, aber es handelt sich nicht um die einzige Fachdatenbank, die Sie nutzen können bzw. sollen. Auch bei den Fachdatenbanken gilt, dass keine Datenbank alle relevanten Texte erschließt.

Die Internationale Bibliographie der Zeitschriftenliteratur (IBZ) weist derzeit über 3,5 Millionen Zeitschriftenaufsätze aus 11 500 Zeitschriften seit 1983 aus. Jährlich kommen über 120 000 Einträge hinzu. Die Datenbank wird monatlich aktualisiert. Der Schwerpunkt liegt bei deutschen und internationalen Zeitschriften der Geistes- und Sozialwissenschaften. Die Suche nach Schlagwörtern kann in deutscher und englischer Sprache erfolgen. Jeder Eintrag enthält den Aufsatztitel, den Autor und die Zeitschrift (Titel, Erscheinungsjahr, Jahrgang, Heftnummer sowie Seitenzahlen). Für etwa 60 000 Aufsätze liegen auch Abstracts vor. Allerdings entsprechen nicht alle ausgewerteten Zeitschriften wissenschaftlichen Standards.

Journal STORage (JSTOR) ist eine gemeinnützige Organisation, die ein kostenpflichtiges Online-Archiv mit älteren Ausgaben ausgewählter Fachzeitschriften unterhält. In JSTOR finden Sie also keine aktuellen Ausgaben von Fachzeit-

Virtual Private Network (VPN)

VPN steht für Virtual Private Network, ein geschlossenes Netz, in dem private Daten gesichert über ein öffentliches Netzwerk übertragen werden. Die Übertragung der Daten erfolgt über einen sogenannten Tunnel und ist damit in der Regel verschlüsselt. Insbesondere für den Zugriff auf Zeitschriftenaufsätze oder zur Recherche von Fachliteratur (z. B. Web of Science) ist häufig eine aktive VPN-Verbindung notwendig.

schriften, sondern ältere Exemplare. Dafür umfasst JSTOR nicht nur eine, sondern etwa 150 politikwissenschaftliche Fachzeitschriften. Seit einiger Zeit bietet JSTOR auch Bücher an. Es wird eine „Basic Search“ und eine „Advanced Search“ angeboten. Die Benutzersprache ist Englisch.

Amtliche Statistiken Ein Rechercheportal für statistische Daten ist der kommerzielle Anbieter Statista. Darüber hinaus sind die statistischen Landesämter und das Statistische Bundesamt die führenden Anbieter von statistischen Informationen in Deutschland. Europäische Statistiken finden sich bei Eurostat oder der OECD. Auch die Weltbank bietet ein großes Angebot an Daten. Beachten Sie die teilweise sehr ausführlichen Informationen zur Harmonisierung der Datenbestände. Die (internationale) Vergleichbarkeit von Bildungs- und Berufsabschlüssen oder auch ökonomischen Indikatoren (z. B. Arbeitslosenquote, BIP, Inflation) ist deutlich komplexer als es auf den ersten Blick aussieht. Abschnitt 8.3.2 bietet eine Übersicht über wichtige amtliche Datensätze.

Wissenschaftliche Datensätze Mittlerweile steht der interessierten akademischen Öffentlichkeit eine große Fülle an Datensätzen für Sekundäranalysen zur Verfügung, die in der Regel kostenlos erhältlich sind. Die notwendigen statistischen Kenntnisse vorausgesetzt, bietet sich hier ein großes Analysepotenzial für Haus- und Abschlussarbeiten. Wer in Deutschland nach solchen Forschungsdaten sucht, sollte seine Suche beim Datenbestandskatalog der GESIS (www.gesis.org) beginnen. Abschnitt 8.3.1 bietet eine Übersicht über wichtige Mikrodaten.

4.3 Strategien zur Formulierung einer Forschungsfrage

Nach der Eingrenzung des Themas eines Forschungsprojekts und der (ersten) Sichtung der relevanten Fachliteratur ist die Entwicklung einer konkreten Forschungsfrage erforderlich, um das Thema und die Zielsetzung des Forschungsprojekts weiter zu konkretisieren. Bei der Formulierung von Forschungsfragen lassen sich zwei Grundformen unterscheiden: Deskriptive und analytische Forschungsfragen.

Deskriptive und analytische Forschungsfragen Die Beschreibung (Deskription) ist eine wichtige Aufgabe von Wissenschaft. In Forschungsfeldern, in denen noch wenig Wissen vorhanden ist, müssen zunächst Daten gesammelt und Beschreibungen geliefert werden, ehe Theorien formuliert und Hypothesen geprüft werden können (Schwarzer 2001, S. 142–143). Zu den meisten sozialwissenschaftlichen Forschungsfeldern existiert mittlerweile umfangreiche Forschungsliteratur.

Deshalb ist nach King et al. (1994, S. 7–8) die sorgfältige Beschreibung häufig zwar unentbehrlich für wissenschaftliche Forschung, aber die Sammlung von Fakten ist nicht das Ziel sozialwissenschaftlicher Forschung.

Selbstverständlich muss ein Wissenschaftler in einem Forschungsprojekt zur Wahlbeteiligung beschreiben können, wie das Wahlsystem funktioniert, und natürlich müssen sich Forscher in einem Projekt zur Entwicklung des Scheidungsrisikos mit den gesetzlichen Rahmenbedingungen zur Ehescheidung auseinandersetzen. Die deskriptive Beschreibung von Wahlsystemen oder die gesetzlichen Bestimmungen zum Scheidungsrecht stellen allerdings nur das Basiswissen dar, das vorliegen muss, ehe interessantere Fragestellungen entwickelt werden können (Stykow et al. 2010, S. 130). Fragen wie „Welche Funktionen hat der Deutsche Bundestag?" oder „Wie hat sich die Wahlbeteiligung entwickelt?" eignen sich weder für Forschungsprojekte noch für Seminararbeiten. Sozialwissenschaftler sind vielmehr an Wirkungszusammenhängen interessiert: Warum hat das Wahlsystem einen Einfluss auf die Wahlbeteiligung (Blais und Aarts 2006; siehe auch Norris 2004)? Warum hat das Scheidungsrisiko seit den 1950er Jahren zugenommen (Becker 2015; siehe auch Esser 1999b, S. 66–68)? Warum gibt es so wenige Arbeiterkinder an Deutschlands Universitäten (Müller und Pollak 2010; siehe auch Schindler 2014)? Das sind analytische Fragestellungen, die nach Zusammenhängen und kausalen Mechanismen fragen und nach Erklärungen für ein Phänomen suchen. Analytische Fragestellungen gehen deshalb weit über die reine Beschreibung hinaus. Während deskriptive Fragestellungen meist Wie-Fragen sind (z. B. Wie funktioniert das Wahlsystem?), handelt es sich bei analytischen Fragestellungen häufig um Warum-Fragen (z. B. Warum unterscheidet sich die nationale Wahlbeteiligung zwischen den EU-Staaten? (Freitag 1996; siehe auch Hadjar und Beck 2010).

Entwicklung einer Forschungsfrage Nach Westle (2009a, S. 125) gibt es zwar keine festen Regeln, wie man zu einer (analytischen) Forschungsfrage kommt, allerdings lassen sich durchaus Strategien nennen, um eine gehaltvolle Forschungsfrage zu entwickeln. Sämtliche Strategien setzen allerdings Kenntnisse des Forschungsstands zu einem Thema voraus. Nur wer einen (ersten) Überblick des Forschungsstands hat, kann spannende Forschungsfragen entwickeln. Mit der „Identifizierung neuer Phänomene", der „Theorienkonkurrenz" und der „Identifizierung wissenschaftlicher Rätsel" werden drei Strategien unterschieden (siehe auch Stykow et al. 2010, S. 127–130; Sebe-Opfermann 2016).

Identifizierung neuer Phänomene Gesellschaftliche Umbrüche (z. B. Arabischer Frühling), technologische Neuerungen (z. B. Internet) oder wirtschaftlicher Wandel (z. B. Finanzkrise) können Ausgangspunkt für die Entwicklung einer For-

Beispiele für die Identifizierung neuer Phänomene

Erklärung eines Phänomens: In den 1960er Jahren waren in vielen westlichen Industriegesellschaften Protestbewegungen und der Wunsch nach „mehr Demokratie" zu beobachten. Diese Proteste interpretierte Ronald Inglehart (1977) als Symptom eines umfassenden Wertewandels. Danach habe in den 1960er Jahren ein intergenerationaler Wandel von materialistischen zu postmaterialistischen Werten stattgefunden. Materialistische Bedürfnisse wie physisches Wohlergehen und ökonomische Sicherheit verlieren an Bedeutung, während postmaterialistische Anliegen wie Selbstverwirklichung und Lebensqualität wichtiger werden. Als Ursache für diesen Wertewandel verweist Inglehart auf die wirtschaftliche Entwicklung in der Nachkriegszeit, in der die Menschen in einem bisher nicht gekannten Ausmaß ökonomische und physische Sicherheit erlebt hätten.

Entwicklung einer Forschungsfrage auf Grundlage einer technologischen Neuerung: Das Internet ist zu einem festen Bestandteil des gesellschaftlichen Alltags geworden. Es bietet nie zuvor gekannte Möglichkeiten, schnell, günstig und mit vergleichsweise geringem Aufwand an (politisch relevante) Informationen zu gelangen. Da das Internet die Informationskosten senkt, sollte es sich – auf den ersten Blick – förderlich auf die politische Partizipation der Bürger auswirken. Ein Forschungsprojekt könnte daher den Zusammenhang zwischen Internet und politischer Partizipation untersuchen. Mittlerweile liegen zahlreiche empirische Befunde zu dieser Forschungsfrage vor (für einen Überblick siehe z. B. Voss 2014).

schungsfrage sein. Eine Forschungsfrage kann dann zum einen darauf abzielen, das neue „Phänomen" zu erklären, oder zum anderen die Auswirkungen des „Phänomens" auf sozialwissenschaftliche Sachverhalte untersuchen.

Theorienkonkurrenz Für die Erklärung sozialer Phänomene gibt es häufig mehrere theoretische Ansätze. Einige Beispiele: In der Wahlforschung existieren mit dem makro-soziologischen, dem mikrosoziologischen, dem sozialpsychologischen und dem Rational-Choice-Ansatz vier theoretische Ansätze zur Erklärung des Wahlverhaltens (für eine Übersicht siehe z. B. Schoen 2009; Falter und Schoen 2014). Für die Erklärung der kommunalen Identität unterscheidet Bühlmann (2013, S. 330) drei Erklärungsansätze: die Sozialisationshypothese, die Lebensqualitätshypothese sowie den Decline-of-Community-Ansatz. Mit den gesellschaftlichen Rahmenbedingungen, den politischen Rahmenbedingungen, den politi-

schen Orientierungen und der individuellen Lebenslage betrachtet Neller (2004, S. 31) vier verschiedene Faktorenbündel zur Erklärung der Lebenszufriedenheit.

Warum unterscheidet sich die Zustimmung zur EU?

Einen äußerst lesenswerten Theorientest hat Matthew Gabel (1998a) vorgelegt. In seinem Beitrag beschäftigt er sich mit der Frage, warum die Zustimmung zur Europäischen Union zwischen den Bürgern variiert. Er unterscheidet fünf theoretische Ansätze und prüft empirisch, welcher Ansatz am besten geeignet ist, die Forschungsfrage zu beantworten. Der Beitrag überzeugt dabei durch seine klare Struktur und nachvollziehbare Argumentation.

Eine solche Theorienkonkurrenz kann als Ausgangspunkt für die Entwicklung einer Forschungsfrage genutzt werden. Einerseits kann untersucht werden, welcher der theoretischen Ansätze das konkrete soziale Phänomen am besten erklären kann. Andererseits können sich solche Theorien auch widersprechen. Dann kann (und sollte) geprüft werden, welche Theorie sich in der Empire bewährt und deshalb vorzuziehen ist.

Identifizierung wissenschaftlicher Rätsel Eine der ertragreichsten Strategien für die Entwicklung einer Forschungsfrage besteht in der Suche nach wissenschaftlichen Rätseln. Ausgangspunkt ist eine – auf den ersten Blick – überraschende empirische Beobachtung. Dieser empirische Befund ist nicht mit unserem Alltagswissen und den klassischen Erklärungsansätzen vereinbar. Die Einordnung eines empirischen Befunds als „überraschend" erfordert dabei umfangreiche Kenntnisse des Forschungsstands. Nur wer den Forschungsstand kennt, kann Widersprüche zwischen theoretischen Erwartungen und empirischen Befunden ausfindig machen. Im Rahmen des Forschungsprojekts soll dieser Widerspruch erklärt werden. Solche Erklärungen beginnen mit einer Warum-Frage (Esser 1999b, S. 31). Die Erklärung ist die Antwort auf die Frage.

Bei Esser (1999b, S. 31–35) finden sich soziologische Warum-Fragen, Stykow et al. (2010, S. 128–129) nennen exemplarische politikwissenschaftliche Rätsel. Das Ziel des Forschungsprojekts ist es, die Warum-Frage zu beantworten. Es müssen folglich Erklärungen identifiziert und/oder entwickelt werden, die den Widerspruch zwischen „Theorie" und „Empirie" auflösen.

Beispiele für Warum-Fragen

Warum unterscheidet sich die Wahlbeteiligung bei Europawahlen zwischen den EU-Staaten? Warum unterscheiden sich die mathematischen Leistungen von Jungen und Mädchen in der PISA-Studie? Warum ist die Parteiidentifikation in Deutschland rückläufig? Warum nimmt die Anzahl der Ehescheidungen über die Zeit zu? Warum fördern Online-Konsultationen der Europäischen Union nicht die Legitimität der Staatengemeinschaft? Warum fördert das Internet nicht die politische Partizipation der Bürger?

Formulieren einer Forschungsfrage für eine Hausarbeit

Die genannten Strategien eignen sich selbstverständlich auch für die Entwicklung einer Forschungsfrage für eine sozialwissenschaftliche Hausarbeit. Allerdings unterscheidet sich die Zielsetzung zwischen einem Forschungsprojekt und einer Hausarbeit. Während ein Forschungsprojekt darauf abzielt, das Wissen im jeweiligen Fachgebiet zu erweitern, soll in einer Hausarbeit gezeigt werden, dass Sie die zentralen Konzepte, Theorien und empirischen Befunde zu Ihrer Forschungsfrage kennen, anwenden und auch miteinander verknüpfen können. Um diese Fertigkeiten zu trainieren und nachzuweisen, nutzen Sie existierende Fachliteratur. Das heißt, Sie nutzen existierende Fachliteratur, um eine Forschungsfrage zu entwickeln und diese zu beantworten. Deshalb eignen sich tagesaktuelle Themen (z. B. Aufstieg einer Partei oder Wahlbeteiligung der jüngsten Bundestageswahl) in der Regel nicht für eine Hausarbeit, da meist keine ausreichende Literaturgrundlage existiert, um eine entsprechende Forschungsfrage zu entwickeln und zu beantworten (zur Erinnerung: Tageszeitungen sind keine Fachliteratur). Im Mittelpunkt einer Hausarbeit steht die systematische Auseinandersetzung mit der existierenden Fachliteratur. Existiert für eine Forschungsfrage noch keine ausreichende Literaturgrundlage, dann sollten Sie für die Hausarbeit eine andere Forschungsfrage suchen. Mit einer solchen Forschungsfrage können Sie sich ggf. in einer Abschlussarbeit oder im Rahmen einer Promotion beschäftigen.

4.4 Relevanz der Forschungsfrage

Eine Forschungsfrage sollte das Forschungsprojekt klar eingrenzen, ein Problem bearbeiten und als Frage formuliert sein. Darüber hinaus sollte eine Forschungsfrage aber auch relevant sein. Mit Blick auf Haus- und Abschlussarbeiten betonen Stykow et al. (2010, S. 123):

> *„Fragen Sie sich, worin die Relevanz Ihrer Problemstellung besteht. Warum ist es wichtig, dass Ihre Arbeit nicht nur geschrieben, sondern auch gelesen wird?"*

Was für Haus- und Abschlussarbeiten gilt, sollte für wissenschaftliche Forschungsprojekte selbstverständlich sein. Woran lässt sich nun aber ein geeignetes wissenschaftliches Problem erkennen? Einen ersten Hinweis gibt Plümper (2012, S. 22):

> *„Zunächst einmal findet ein geeignetes wissenschaftliches Problem nicht ausschließlich das Interesse des bearbeitenden Wissenschaftlers, sondern interessiert zugleich (möglichst viele) andere Wissenschaftler und vielleicht sogar die Öffentlichkeit."*

Zwei Dimensionen von Relevanz Die Antwort von Plümper (2012, S. 22) verweist auf zwei Dimensionen von Relevanz, die Lehnert et al. (2007, S. 41) einerseits als theoretische und andererseits als gesellschaftliche Relevanz bezeichnen (siehe auch King et al. 1994, S. 14–15). Die theoretische Relevanz bezieht sich auf den Beitrag des Projekts zum wissenschaftlichen Diskurs. Kann das Forschungsprojekt dazu beitragen, „den Grundstock wissenschaftlicher Erkenntnisse über soziale Phänomene zu erweitern" oder zumindest die Voraussetzungen dafür schaffen, „indem Daten gesammelt, Theorien überprüft, Kausalitätsmechanismen nachgewiesen, Prozesse beobachtet oder wissenschaftliche Methoden verbessert werden" (Schwarzer 2001, S. 142)? Mit anderen Worten: Leistet das Projekt einen Beitrag für die Fachwelt? Gesellschaftliche Relevanz bezieht sich dagegen auf die Bedeutung des Forschungsprojekts für die „reale Welt". Das Thema bzw. die Erkenntnisse eines Forschungsprojekts tragen dazu bei, dass Ereignisse und Phänomene besser verstanden werden. King et al. (1994, S. 15) betonen:

> *„The topic should be consequential for political, social, or economic life, for understanding something that significantly affects many people's lives, or for understanding and predicting events that might be harmful or beneficial."*

Lehnert et al. (2007, S. 41) und Schwarzer (2001, S. 142) betonen übereinstimmend, dass ein Forschungsprojekt im idealen Fall beiden Kriterien genügt. Es leistet einerseits einen Beitrag zum Verständnis eines realen Problems und es verbessert

andererseits den innerfachlichen Erkenntnisfortschritt in Theorie und/oder Methode. Lehnert et al. (2007) nennen mehrere Hinweise, um die theoretische und gesellschaftliche Relevanz eines Forschungsprojekts herauszuarbeiten.

Theoretische Relevanz Demzufolge ist die explizite Anbindung des eigenen Beitrags an die einschlägige Literatur der Ausgangspunkt für die Verbesserung der theoretischen Relevanz (Lehnert et al. 2007, S. 42). Dadurch lässt sich zum einen vermeiden, dass eine Idee mehrfach formuliert wird, und zum anderen entspricht die explizite Anbindung an den Forschungsstand auch dem „kumulativen Charakter des Forschungsprozesses, also der Idee, dass Beiträge unterschiedlicher Forscher aufeinander aufbauen und somit als Ganzes zum besseren Verständnis der Materie beitragen" (Lehnert et al. 2007, S. 42). Konkrete Vorschläge zur Förderung der theoretischen Relevanz eines Forschungsprojekts sind (siehe Lehnert et al. 2007, S. 43):

- Die empirische Prüfung bisher nicht überprüfter Hypothesen. Dadurch kann zum einen das empirische Wissen zu sozialen Phänomen verbessert und zum anderen können unbestätigte Hypothesen verworfen und ein weitergehendes theoretisches Verständnis erreicht werden.
- Inkonsistenzen einer Theorie aufdecken. Eine Forschungsfrage, die Inkonsistenzen einer Theorie diskutiert und möglicherweise auch beseitigt, indem im Rahmen des Forschungsprojekts die Theorie ergänzt und/oder überarbeitet wird, stellt einen theoretisch relevanten Beitrag dar.
- Identifizierung empirischer Fälle, die von einer Theorie bisher nicht erklärt werden. Forschungsfragen, die „überraschende" Fälle aufdecken und möglicherweise auch erklären, können dazu beitragen, dass Theorien verbessert werden.
- Die Entwicklung klar spezifizierter Konzepte und/oder die Entwicklung valider Messinstrumente sind sowohl für die Theorieentwicklung als auch für die empirische Prüfung unverzichtbar. Sie tragen dazu bei, das analytische Instrumentarium in einem Fachgebiet zu erweitern.
- Formulierungen von Alternativerklärungen, die entweder bestehende Ansätze verbinden oder alternative Mechanismen vorschlagen, die bisher nicht diskutiert und/oder getestet wurden.
- Anwendung einer Theorie auf ein neues Feld. Wenn eine existierende Theorie auch Beobachtungen erklären kann, die zuvor noch nicht unter dieser theoretischen Perspektive betrachtet wurden, so trägt dies dazu bei, den Geltungsbereich einer Theorie zu verbessern.

Sicherlich kann die Liste an Hinweisen und Vorschlägen für die Verbesserung der theoretischen Relevanz eines Forschungsprojekts noch erweitert werden. Alle

Vorschläge zielen letztendlich darauf ab, das fachspezifische Wissen im jeweiligen Bereich zu erweitern. Allein die Tatsache, dass bestimmte Hypothesen noch nicht überprüft oder spezifische Fälle noch nicht untersucht wurden, stellt allerdings noch keine ausreichende Relevanzbegründung dar. In anderen Worten: Es reicht als Begründung nicht aus, dass irgendetwas bisher „noch nicht gemacht wurde". Es kann durchaus gute Gründe dafür geben, dass das so ist. Vielmehr muss deutlich gemacht werden, warum es wichtig (und möglich) ist, die Hypothese zu überprüfen oder den Fall zu untersuchen.

Gesellschaftliche Relevanz Zur Förderung der gesellschaftlichen Relevanz formulieren Lehnert et al. (2007, S. 48–52) drei aufeinander aufbauende Fragen, die jeder Wissenschaftler an die eigene Arbeit stellen kann, um die gesellschaftliche Relevanz stärker herauszuarbeiten. Erstens: Wer ist betroffen? Zweitens: Wie lassen sich die Effekte bewerten? Drittens: Welche Ratschläge können erteilt werden?

Die erste Frage „Wer ist betroffen?" richtet den Blick auf die Personengruppe, die potenziell von dem untersuchten Gegenstand betroffen ist. Bei Forschungsarbeiten zum Wahlverhalten wären das beispielsweise alle wahlberechtigten Bürger und bei Projekten zum sozialen Vertrauen theoretisch alle Menschen. Deshalb kann und sollte, nachdem die potenziell betroffenen Menschen identifiziert sind, eine Folgefrage gestellt werden: Wie wirkt sich der untersuchte Gegenstand auf die (potenziell) betroffenen Menschen aus? Zur Beantwortung dieser Frage müssen die kausalen Mechanismen identifiziert werden, die zwischen dem Sachverhalt und den Betroffenen wirken.

Die zweite Frage „Wie lassen sich die Effekte bewerten?" widmet sich den Konsequenzen eines Sachverhalts für die betroffene Personengruppe. Nach Lehnert et al. (2007, S. 50) sollte in einem Forschungsprojekt herausgearbeitet werden, „wie Menschen durch ein Phänomen betroffen sind und warum sie sich um diese Effekte kümmern sollten". Die Ergebnisse könnten mit Blick auf ihre Legitimität, Stabilität oder Transparenz bewertet werden. Allerdings weisen die Autoren zu Recht darauf hin, dass es in den Sozialwissenschaften nicht einen akzeptierten normativen Standard gibt. Mit Blick auf die Bewertung von Effekten ist es daher nicht ausreichend, einen einzigen Standard unbegründet zugrunde zu legen.

> *„Vielmehr sollte die Wahl der Standards bewusst, und explizit für den Leser nachvollziehbar, erfolgen! Man sollte stets nach Beziehungen zwischen verschiedenen Maßstäben fragen und gegebenenfalls politische Phänomene und ihre Effekte anhand mehrerer, auch potenziell konfligierender Kriterien bewerten." (Lehnert et al. 2007, S. 51)*

Die dritte Frage „Welche Ratschläge können erteilt werden?" schließt direkt an die Bewertung sozialer Phänomene an. Durch die Diskussion der praktischen Im-

plikationen der Arbeit kann die gesellschaftliche Relevanz der Arbeit gefördert werden.

> *„Allerdings gilt auch hier: Jede politische Empfehlung ist abhängig von den ihr zugrunde liegenden Bewertungsmaßstäben. Auf der Grundlage dieser Bewertungsmaßstäbe argumentieren wir dann für eine bestimmte Idee der ‚guten Gesellschaft' oder des ‚guten Regierens'. Eine Antwort auf die Frage der guten Gesellschaft oder des guten Regierens muss gleichzeitig deutlich machen, welche Ziele sie als erstrebenswert, welche Begebenheiten sie als nicht wünschenswert ansieht, und wie wir die jeweiligen Ziele erreichen beziehungsweise die jeweiligen Begebenheiten verändern können." (Lehnert et al. 2007, S. 51–52)*

! Die normative Bewertung eines Sachverhalts ist deshalb mit vielen Voraussetzungen verbunden und argumentativ anspruchsvoll. Appelle, entrüstete Kommentare und persönliche Bekenntnisse gehören nicht in einen wissenschaftlichen Text (siehe z. B. Stykow et al. 2010, S. 89). Vielmehr sind wissenschaftliche Texte stets distanziert und neutral zu verfassen, da die Bewertung eines gesellschaftlichen Sachverhalts erheblich vom Standpunkt des Betrachters abhängig ist. Dies wird am nachfolgenden Exkurs erläutert.

Exkurs Warum gefährdet mehr Demokratie die politische Gleichheit? Direkte Demokratie bezeichnet ein politisches Entscheidungsverfahren, bei dem die stimmberechtigten Bürger einer politischen Ebene durch Abstimmung selbst und unmittelbar über eine Sachfrage entscheiden (z. B. Schiller und Mittendorf 2002, S. 11–13; Kielmansegg 2006; Walter-Rogg 2008, S. 236). Seit etwa zwei Jahrzehnten gewinnt die direkte Demokratie eine immer größere Bedeutung und hat weltweit Konjunktur (z. B. Hornig 2011, S. 475; Merkel und Ritzi 2017).

Für die verstärkte Einführung direktdemokratischer Elemente auf den verschiedenen politischen Ebenen sprechen mehrere Argumente (für eine Übersicht siehe z. B. Eder 2011, S. 85–91). So kommt die direkte Demokratie dem Grundsatz des Regierens durch das Volk sicherlich am nächsten (Schmidt 2010, S. 350). Abstimmungen über Sachfragen schaffen auch politische Aufmerksamkeit. Dies fördert die Auseinandersetzung mit politischen Inhalten, begünstigt den Erwerb von Kenntnissen und ermöglicht einen öffentlichen Diskurs (Schäfer 2011, S. 55). Direkte Demokratie ermöglicht damit Prozesse des politischen Lernens und die Herausbildung staatsbürgerlicher Verantwortung für das Gemeinwohl. Merkel (2011, S. 49) bezeichnet Volksentscheide deshalb auch als „Schule der Demokratie".

Forschungsergebnisse zur Direktdemokratie deuten allerdings darauf hin, dass die Teilnahme an Volksentscheiden in der Regel geringer ist als die Beteiligung an allgemeinen Wahlen (z. B. Merkel 2011, S. 50; Merkel und Petring 2012, S. 113–114). Die im Vergleich zu allgemeinen Wahlen geringere Beteiligung bei

Abstimmungen begünstigt eine stärkere soziale Selektivität (z. B. Töller et al. 2011, S. 520–522; Schäfer und Schoen 2013; Fatke 2015). Merkel (2011, S. 51) bilanziert: „Nicht das Volk in seiner Gesamtheit, sondern die höheren und mittleren Schichten, die Gebildeten und überproportional die Männer stimmen typischerweise bei Referenden ab. Der politisch aktive Demos ist dabei mehr als halbiert und hat bei Volksabstimmungen damit eine noch größere soziale Schieflage als bei nationalen Parlamentswahlen."

Wie können diese Befunde jetzt bewertet werden? Aus der Perspektive eines liberalen Demokratieverständnisses werden direktdemokratische Verfahren (mittlerweile) problematisiert, da sie offensichtlich gegen das Ideal politischer Gleichheit verstoßen (Dahl 2006). Die Vertreter eines elitären Demokratieverständnisses (z. B. Schumpeter 1950) würden diese „soziale Selektivität" vermutlich begrüßen, da sich eher Personen beteiligen, die in politischen Fragen besser informiert und zu einer rationalen Entscheidung befähigt sind (Brennan 2011, 2017). Vor dem Hintergrund der Befunde ist es fraglich, ob die Verfahren der direkten Demokratie „das Herzstück einer modernisierten Demokratie" bilden, die „durch ihr Mehr an Legitimation und Partizipation einen Ausweg aus der Parteienkrise und Politikverdrossenheit aufzeigen können" (Weixner 2006, S. 18). Entsprechend provokativ fragen Schäfer und Schoen (2013): „Mehr Demokratie, aber nur für wenige?" (siehe auch Merkel 2015, S. 27–41; Schäfer 2015).

5 Konzeptspezifikation

Demokratie, Politikverdrossenheit sowie politische Beteiligung – in den Sozialwissenschaften sind wir häufig mit Begriffen konfrontiert, die Bestandteil des alltäglichen und nichtwissenschaftlichen Sprachgebrauchs sind. Diese Begriffe sind allerdings oft unpräzise und es existiert kein einheitliches Begriffsverständnis (Friedrichs 1990, S. 73–74; Kühne 2013, S. 25). Deshalb müssen die verwendeten Begriffe eindeutig definiert und von anderen Begriffen abgegrenzt werden, ehe die mit diesen Begriffen bezeichneten Sachverhalte empirisch untersucht werden können. Diese Phase im Forschungsprozess wird als Konzeptspezifikation bezeichnet. Dieses Kapitel illustriert die Notwendigkeit der Konzeptspezifikation, stellt die Elemente eines Konzepts dar, diskutiert Herausforderungen und gibt Hinweise zur (erfolgreichen) Konzeptspezifikation.[7]

7 In der neueren Fachliteratur hat sich die Bezeichnung Konzeptspezifikation durchgesetzt (z. B. Gerring 2001; Goertz 2006; Schnell et al. 2013, S. 117). In der älteren Literatur wird gelegentlich noch von dimensionaler und semantischer Analyse anstatt von Konzeptspezifikation gesprochen (z. B. Kromrey et al. 2016, S. 113–162). Die dimensionale Analyse zielt darauf ab, die relevanten Dimensionen des Untersuchungsgegenstands zu identifizieren. Die semantische Analyse beschäftigt sich mit der Bedeutung eines Begriffs. Beide haben gemeinsam, dass „ein zunächst noch unbestimmter Sachverhalt gedanklich strukturiert *und* begrifflich präzisiert werden soll" (Kromrey et al. 2016, S. 141). Die Zielsetzung der dimensionalen und semantischen Analyse sowie der Konzeptspezifikation ist damit vergleichbar.

© Springer Fachmedien Wiesbaden GmbH, ein Teil von Springer Nature 2018

M. Tausendpfund, *Quantitative Methoden in der Politikwissenschaft*,
Grundwissen Politik, https://doi.org/10.1007/978-3-658-20698-7_5

5.1 Über was sprechen wir?

Seit 1977 wählt die Gesellschaft für deutsche Sprache (GfdS) regelmäßig das Wort des Jahres und kürt damit Begriffe und Redewendungen, die das gesellschaftliche Leben eines Jahres in besonderer Weise geprägt haben.[8] 1992 war das Wort des Jahres „Politikverdrossenheit". Mehr als 25 Jahre später begegnet uns der Begriff immer noch regelmäßig in der politischen Berichterstattung der Tagespresse, bei Gesprächsrunden im Fernsehen und auch in wissenschaftlichen Publikationen. „Politikverdrossenheit" wird in unzähligen Veröffentlichungen häufig als Ursache für eine ganze Reihe von politischen Problemen und Entwicklungen genannt. Politikverdrossenheit sei verantwortlich für die sinkende Wahlbeteiligung, die Wahlerfolge rechtspopulistischer Parteien oder auch die Mitgliederverluste der Parteien. Das Problem dabei: Der Begriff „Politikverdrossenheit" wird in den verschiedenen Kontexten höchst unterschiedlich verstanden und verwendet.

Was ist Politikverdrossenheit? Der Politikwissenschaftler Kai Arzheimer (2002; siehe auch Maier 2000) hat sich eingehender mit „Politikverdrossenheit" beschäftigt. In seiner Monographie hat er auf Basis von 176 Publikationen eine Bedeutungsanalyse des Begriffs „Politikverdrossenheit" vorgelegt und auf die Heterogenität der Begriffsverwendung aufmerksam gemacht. Zum einen werden je nach Veröffentlichung Parteien, Politiker, das Prinzip der repräsentativen Demokratie, die Regierung, politische Entscheidungen, die öffentlichen Verwaltungen und auch Kirchen sowie Gewerkschaften als Gegenstand von Politikverdrossenheit genannt. Teilweise richtet sich Politikverdrossenheit auch einfach gegen „die Politik" oder „das Parteiensystem" (Arzheimer 2002, S. 122). Zum anderen stellt Arzheimer (2002, S. 130) neben den vielen – und im Zeitverlauf zunehmenden Objekten – auch eine große Anzahl an Einstellungen fest, die mit Politikverdrossenheit in Zusammenhang gebracht werden. Politikverdrossenheit bedeutet unter anderem Enttäuschung, Unzufriedenheit, Misstrauen, politisches Desinteresse, aber auch politisches Interesse, eine negative Wahrnehmung der Wirtschaftslage und Angstgefühle. Angesichts der uneinheitlichen Verwendung stellt Arzheimer (2002, S. 202) fest:

> *„Bezüglich der Objekte, der Ursachen, der Mikro- und der Makro-Folgen von Verdrossenheit herrscht in der Literatur ebensowenig Konsens wie hinsichtlich der theoretischen Vernetzung des Konzeptes oder der zu verwendenden Begrifflichkeit. Nicht einmal dar-*

8 Für weitere Informationen zum Auswahlverfahren des Wortes des Jahres siehe die Homepage der Gesellschaft für deutsche Sprache (GfdS) unter http://gfds.de.

> *über, welche Einstellungen überhaupt mit Verdrossenheit bezeichnet werden sollen, sind sich die Autoren einig. Der kleinste gemeinsame Nenner der fast 180 untersuchten Arbeiten besteht darin, daß sie mit ‚Verdrossenheit' negative oder zumindest neutrale Einstellungen gegenüber einer Vielzahl von politischen Objekten bezeichnen." (Arzheimer 2002, S. 202)*

Für eine empirische Untersuchung ist ein unscharfer und mehrdeutiger Begriff allerdings sehr problematisch. Wie können empirische Befunde interpretiert werden, wenn sich Politikverdrossenheit einmal auf Regierungen und ein anderes Mal auf das Prinzip der Demokratie bezieht? Wie sollen empirisch überprüfbare Hypothesen formuliert werden, wenn Politikverdrossenheit einerseits mit politischem Interesse und andererseits mit politischer Apathie in Verbindung gebracht wird? Wie soll wissenschaftliche – kumulative – Forschung funktionieren, wenn Begriffe unterschiedlich verstanden werden? Die Antwort gibt Andreas Schedler (2011, S. 370–371):

> *„Concepts are central to the enterprise of political science. The concepts we use shape the world we see. Without solid conceptual foundations, the edifice of political science is insecure. If we fail to develop clear and precise concepts, our theoretical insights and empirical discoveries will fail to be clear and precise, too."*

Klar definierte Konzepte sind die Voraussetzung für die empirische Sozialforschung. Sie stellen die Basis für die Formulierung von gehaltvollen Hypothesen dar und sind Grundlage für eine valide Operationalisierung (Goertz 2006). Empirische Forschung, die auf unklaren Konzepten beruht, ist „zumindest verwirrend, sehr wahrscheinlich unproduktiv und definitiv nicht kumulativ" (Wonka 2007, S. 66). Deshalb müssen die in einem Forschungsprojekt verwendeten Begriffe definiert und von anderen Begriffen abgegrenzt werden. Mit Blick auf den Begriff „Politikverdrossenheit" argumentiert Arzheimer (2002, S. 294):

> *„Im Überblick erscheint Politikverdrossenheit als ein extrem heterogenes Konzept, das in unsystematischer Weise auf eine ganze Reihe von Sachverhalten bezogen wird und den Kriterien für einen wissenschaftlichen Begriff nicht genügt, weil es ihm an Eindeutigkeit und Präzision mangelt."*

Offensichtlich eignet sich der vieldeutige und unscharfe Begriff „Politikverdrossenheit" nicht unmittelbar für die wissenschaftliche Auseinandersetzung (Arzheimer 2002, S. 297). Deshalb plädieren einige Autoren dafür, den Begriff durch präzisere, eindeutigere und adäquatere Konzepte aus der internationalen Forschung zu ersetzen. Für eine gehaltvolle Beschreibung und Erklärung der Haltung von

Klassiker der Sozialwissenschaften: Konzept der politischen Unterstützung

Das Konzept der politischen Unterstützung (Easton 1965) ist heute wohl die einflussreichste Konzeptualisierung von Einstellungen gegenüber politischen Systemen. Easton unterscheidet mit der politischen Gemeinschaft, dem politischen Regime und den politischen Autoritäten drei Objekte eines politischen Systems. Diese Systematisierung basiert auf der Überlegung, dass die abstrakteren Objekte (z. B. Gemeinschaft) eine andere Form der politischen Unterstützung erfahren als die konkreten Objekte (z. B. Bundesregierung) eines politischen Systems. Er differenziert nicht nur zwischen Objekten politischer Unterstützung, sondern mit der spezifischen und diffusen Unterstützung auch zwischen zwei Arten der politischen Unterstützung. Bei der spezifischen Unterstützung handelt es sich um eine leistungsabhängige Form der Unterstützung, die diffuse Unterstützung ist von konkreten Leistungen unabhängig. Das Konzept der politische Unterstützung ist nicht unumstritten, es stellt aber „schon seit mehreren Dekaden eine zentrale Größe in der empirischen Demokratieforschung dar" (Fuchs 2002a, S. 366) und nahezu alle Forschungsarbeiten zur politischen Unterstützung beziehen sich auf seine Überlegungen (z. B. Dalton 2004; Kornberg und Clarke 2011; Norris 2011, 2017). Überarbeitete Konzeptionen politischer Unterstützung haben Fuchs (1989), Westle (1989) oder auch Norris (1999) vorgelegt. Eine knappe Darstellung des Konzepts der politischen Unterstützung sowie seine Weiterentwicklung bietet der Beitrag von Braun und Schmitt (2009).

Bürgerinnen und Bürger gegenüber politischen Objekten bietet sich beispielsweise das Konzept der politischen Unterstützung an (Fuchs 2002b, S. 339).

Der in der gesellschaftlichen Debatte bekannte und bei Journalisten gern verwendete Ausdruck „Politikverdrossenheit" ist allerdings nur ein exemplarisches Beispiel für viele unpräzise Begriffe, die in der alltäglichen Kommunikation verwendet werden. Die Liste lässt sich problemlos fortführen – Wutbürger (z. B. Stürmer 2011), Vertrauen (z. B. Hartmann 2004), Familie (Hill und Kopp 2015) oder auch Euroskeptizismus (z. B. Boomgaarden et al. 2011). Für eine wissenschaftliche – intersubjektiv nachprüfbare – Auseinandersetzung und empirische Untersuchung dieser Phänomene sind diese „blumigen" Begriffe allerdings zu konkretisieren und ihre inhaltliche Bedeutung offenzulegen. Dies ist das Ziel der Konzeptspezifikation.

5.2 Was ist ein Konzept?

Viele Begriffe in den Sozialwissenschaften sind Bestandteil unserer Alltagssprache, die allerdings häufig mehrdeutig und teilweise auch mit (impliziten) Wertungen versehen sind. Dies erschwert nicht nur die Kommunikation mit anderen Personen, sondern auch die empirische Untersuchung dieser mit diesen Begriffen bezeichneten Sachverhalte. Klar spezifizierte Konzepte verringern die Verständigungsschwierigkeiten und bilden das Fundament für die empirische Analyse.

> *„Concept formation thus lies at the heart of all social science endeavor. It is impossible to conduct work without using concepts. It is impossible even to conceptualize a topic, as the term suggests, without putting a label on it." (Gerring 2001, S. 35)*

Mit Gerring (2001, S. 39; siehe auch Wonka 2007, S. 64–65) lassen sich drei Elemente unterscheiden, die gemeinsam ein Konzept bilden: Term, Intension und Extension.

Term Term ist die Bezeichnung eines Konzepts. Es handelt sich um ein sprachliches „Label", das aus einem oder wenigen Wörtern besteht. In diesem Verständnis sind Politikverdrossenheit, Vertrauen oder auch Demokratie zunächst einmal Bezeichnungen für Konzepte, die in ihrer inhaltlichen Bedeutung noch nicht näher spezifiziert sind.

Intension Die inhaltliche Bedeutung eines Konzepts wird durch Attribute (Eigenschaften) festgelegt. Ein Attribut beschreibt eine spezifische Eigenschaft eines Konzepts. Sie dienen der inhaltlichen Präzisierung, der *„inneren Beschaffenheit"* (Westle 2009a, S. 128; Hervorhebung im Original), eines Konzepts. Folgt man Wonka (2007, S. 64), dann füllen Attribute die Bezeichnung eines Konzepts mit Substanz. Für die inhaltliche Präzisierung eines Konzepts sind in der Regel mehrere Attribute erforderlich. Alle Attribute bilden die Intension (Begriffsinhalt) eines Konzepts.

> *„Die Intension eines Konzepts ist nicht nur deshalb wichtig, weil sie die inhaltliche Bedeutung eines Konzepts definiert. Sie grenzt dieses gleichzeitig von anderen Konzepten ab. Starke Überschneidungen in der Intension unterschiedlicher Konzepte führen zu Abgrenzungsproblemen und provozieren inhaltliche Missverständnisse." (Wonka 2007, S. 64)*

Die Intension ist die Menge aller Attribute. Ein Objekt wird nur mit dem Begriff bezeichnet, wenn es alle Attribute erfüllt. Ein Beispiel: Mögliche Attribute für das Konzept „Auto" (Label) könnten „Motor", „Räder" und „Türen" sein. Die Attri-

bute „Motor“, „Räder“ und „Türen“ bilden die Intension des Begriffs „Auto“. Ein (empirisches) Objekt, das diese Attribute besitzt, wird entsprechend als Auto bezeichnet.

Extension Die Extension (Begriffsumfang) bezeichnet den empirischen Geltungsbereich eines Konzepts. Es handelt sich um die Gesamtheit der Objekte, die unter dieses Konzept fallen (Kromrey et al. 2016, S. 146). Die Extension des Konzepts „Auto“ umfasst folglich alle Objekte, die die genannten Attribute – Motor, Räder und Türen – aufweisen. Damit besteht zwischen der Intension und der Extension ein inverses Verhältnis: Je größer die Intension, das heißt, je mehr Attribute zur inhaltlichen Präzisierung eines Konzepts genutzt werden, desto geringer die Anzahl der Objekte, die unter diesen Begriff fallen (Gerring 2001, S. 82).

Gütekriterien Schnell et al. (2013, S. 118) begreifen die Konzeptspezifikation „als eine spezielle Art der Nominaldefinition“ (siehe auch Kromrey et al. 2016, S. 148–157).[9] Deshalb existieren keine objektiven Kriterien, ob eine Spezifikation gelungen ist oder nicht (Schnell et al. 2013, S. 118; Kromrey et al. 2016, S. 155). Atteslander (2010, S. 41) stellt fest:

> *„Begriffe in der empirischen Sozialforschung gründen auf einer Übereinkunft der Forscher, die in jedem einzelnen Fall darüber befinden müssen, ob diese Begriffe theoretisch sinnvoll und empirisch praktikabel sind.“*

Es finden sich in der Literatur allerdings durchaus Vorschläge, um die Qualität einer Konzeptspezifikation zu bewerten. Gerring (2001, S. 40) unterscheidet mit Kohärenz, Operationalisierung, Validität, Nützlichkeit, Resonanz, Kontextgebundenheit, Sparsamkeit sowie analytischer und empirischer Brauchbarkeit insgesamt acht Kriterien. Einzelne Kriterien beziehen sich allerdings nicht unmittelbar auf den theoretischen Klärungsprozess, sondern auf nachgelagerte Phasen im Forschungsprozess. Nach Kromrey et al. (2016, S. 155) ist die „Zweckmäßigkeit“ das zentrale Qualitätskriterium der Konzeptspezifikation. Kühne (2013, S. 26) unterscheidet zwischen Präzision, Eindeutigkeit und Adäquatheit (ähnlich auch Friedrichs 1990, S. 74; Opp 2014, S. 141–147). Die Anwendung dieser drei Kriterien geben nützliche Hinweise, ob die Konzeptspezifikation abgeschlossen ist und das spezifizierte Konzept als Grundlage für eine empirische Untersuchung genutzt werden kann.

---

9 Zur Erinnerung: Bei einer Nominaldefinition wird die Bedeutung eines Begriffs (Definiendum) durch einen bekannten Begriff (Definiens) festgelegt (siehe Abschnitt 2.1).

Präzision Die Präzision bezieht sich auf den Grad der Genauigkeit, ob ein konkretes Objekt oder Ereignis zum empirischen Geltungsbereich des Konzepts gehört oder nicht. Bei einer hohen Präzision ist diese Feststellung unproblematisch. Bei einer geringen Präzision kann diese Feststellung bei vielen Objekten oder Ereignissen nicht getroffen werden. Je geringer die Präzision, desto größer sind die Schwierigkeiten, Objekte einem Konzept zuzuordnen.

Eindeutigkeit Die Eindeutigkeit bezieht sich auf die einheitliche Verwendung eines Konzepts bei verschiedenen Personen. Personen, die das Konzept kennen, verwenden das Konzept konsistent. Zudem wird die Zuordnung von Objekten oder Ereignissen zu einem Konzept von verschiedenen Personen geteilt.

Adäquatheit Die Adäquatheit bezieht sich auf die Angemessenheit eines Konzepts. Dieses Kriterium bezieht sich auf das Verhältnis von Intension und Extension. Bei einer zu geringen Intension ist der empirische Geltungsbereich – die Objekte, die einem Konzept zugeordnet werden – zu groß, bei einer hohen Intension existieren nur wenige Untersuchungsobjekte. Adäquat ist eine Konzeptspezifikation, die die relevanten Objekte und Ereignisse umfasst.

Konzept: Term, Intension und Extension

Der Term gibt einem Konzept einen Namen. Es handelt sich um das „Etikett“ bzw. „Label“ eines Konzepts. Die inhaltliche Bedeutung eines Konzepts wird durch die Intension festgelegt. Die Intension umfasst die Menge der Attribute (Eigenschaften), die ein Objekt (z. B. Personen oder Gegenstände) aufweisen muss, damit das Objekt mit dem „Label“ des Konzepts versehen werden kann. Die Extension eines Konzepts bezieht sich auf die Gesamtheit der (empirischen) Objekte oder Ereignisse, die der Intension des Konzepts entsprechen. Es handelt sich um den empirischen Geltungsbereich eines Konzepts. Nützliche Hinweise für die Zweckmäßigkeit einer Konzeptspezifikation bieten die Präzision, die Eindeutigkeit und die Adäquatheit eines Konzepts.

5.3 Herausforderungen der Konzeptspezifikation

Die Konzeptspezifikation ist eine Phase im Forschungsprozess, die in intensiver Auseinandersetzung mit der Fachliteratur stattfindet. Relevante Konzepte müssen identifiziert, die Attribute dieser Konzepte herausgearbeitet und systematisch gegenübergestellt werden. Im Forschungsalltag lassen sich zwei typische Herausforderungen der Konzeptspezifikation unterscheiden. Erstens finden sich in der sozialwissenschaftlichen Fachliteratur viele Konzepte, die zwar den gleichen Begriff (Term) verwenden, aber eine unterschiedliche Bedeutung (Intension) haben. Zweitens ist die inhaltliche Bedeutung eines Konzepts häufig an spezifische räumliche und/oder zeitliche Kontexte gebunden, die sich ändern (können). Für die Anpassung an einen neuen Untersuchungsgegenstand kann eine Re-Spezifizierung des Konzepts erforderlich sein. Beide Herausforderungen werden im Folgenden an Beispielen illustriert.

Uneinheitliche Verwendung eines Konzepts Sozialkapital gilt als „verwickeltes Thema mit vielen schillernden Facetten" (Esser 2000b, S. 235), das weit über die Sozialwissenschaften hinaus eine enorme Popularität erlangt hat (Kriesi 2007, S. 23). Der Deutsche Bundestag (2002) hat sich bereits mit Sozialkapital beschäftigt und internationale Organisationen wie OECD und Weltbank wollen Sozialkapital fördern. Aus soziologischer Perspektive ist das Konzept „Sozialkapital" insbesondere mit Arbeiten von Bourdieu (1983) verbunden, die politikwissenschaftliche Debatte wird von Putnam (1993, 2000) geprägt.[10] Beide verwenden den Terminus „Sozialkapital" bzw. „Soziales Kapital", aber die definierenden Attribute sind völlig unterschiedlich. So beschreibt Bourdieu (1983, S. 190–191) Sozialkapital als

> *„die Gesamtheit der aktuellen und potenziellen Ressourcen, die mit dem Besitz eines dauerhaften Netzes von mehr oder weniger institutionalisierten Beziehungen gegenseitigen Kennens oder Anerkennens verbunden sind; oder, anders ausgedrückt, es handelt sich dabei um Ressourcen, die auf der Zugehörigkeit zu einer Gruppe beruhen".*

Diese Spezifikation ist nahe am Alltagsverständnis. Sozialkapital ist „Vitamin B" (Roßteutscher 2009, S. 169). Durch Beziehungen zu anderen Personen erhalte ich Zugang zu Ressourcen (z. B. Informationen), die mir einen Vorteil verschaffen

10 Neben Bourdieu und Putnam haben noch weitere Autoren Konzeptspezifikationen für Sozialkapital entwickelt (für einen Überblick siehe z. B. Haug 1997; Franzen und Freitag 2007a; Castiglione et al. 2008). Esser (2000b, S. 235–268; 2008) hat eine umfassende Re-Spezifizierung des Konzepts vorgelegt, die zahlreiche Defizite früherer Konzeptspezifikationen überwindet.

können. Dagegen konzeptualisiert Putnam (1995, S. 664) Sozialkapital als „networks, norms, and trust“ und beschreibt es wie folgt:[11]

> *„… social capital refers to connections among individuals – social networks and the norms of reciprocity and trustworthiness that arise from them.“ (Putnam 2000, S. 19)*

Nach dieser Konzeptspezifikation entsteht Sozialkapital in sozialen Netzwerken. In und durch soziale Interaktionen entwickeln sich soziales Vertrauen sowie Normen der Reziprozität.

Bei einem Vergleich der von Bourdieu und Putnam vorgeschlagenen Konzeptspezifikationen ist auffällig, dass die das Konzept definierenden Attribute unterschiedlich sind. Beide Autoren verwenden zwar den gleichen Term (Sozialkapital), aber die inhaltliche Bedeutung ist unterschiedlich. Deshalb muss bei einer wissenschaftlichen Auseinandersetzung mit dem Konzept „Sozialkapital“ die Intension der verwendeten Konzeptualisierung offengelegt werden, damit die theoretische Argumentation und empirische Befunde von anderen Personen nachvollzogen und kritisiert werden können. Ansonsten sind Irritationen und Missverständnisse programmiert. Die „Zweckmäßigkeit“ (Kromrey et al. 2016, S. 155) einer gewählten Konzeptspezifikation kann immer nur im Hinblick auf das Forschungsthema bzw. die Forschungsfrage diskutiert werden.

Re-Spezifizierung eines Konzepts Häufig sind sozialwissenschaftliche Konzepte an spezifische räumliche und/oder zeitliche Kontexte gebunden. Falls ein Konzept in einem anderen als dem ursprünglich vorgesehenen räumlichen oder zeitlichen Kontext angewendet wird, dann muss geprüft werden, ob die relevanten empirischen Untersuchungsobjekte von dem Konzept erfasst werden.

> *„Wird ein Konzept auf einen veränderten Kontext angewendet, kann der ‚Fit‘ zwischen den untersuchten Fällen und dem Originalkonzept verloren gehen. […] Erfolgt eine solche Anpassung nicht, werden Konzepte auf empirische Untersuchungsgegenstände angewendet, die von dem Konzept nicht erfasst werden können, weil die definierenden Attribute des Konzepts und die empirischen Eigenschaften des Untersuchungsgegenstands nicht korrespondieren. Hieraus könnte eine falsche analytische Erfassung der empiri-*

11 Für die Spezifikation von Sozialkapital nutzt Putnam die Konzepte „networks“, „norms“ und „trust“. Diese Konzepte sind allerdings selbst nicht eindeutig und müssen im nächsten Schritt spezifiziert werden. Die Konzeptspezifikation von Sozialkapital führt also zu weiteren Konzeptspezifikationen („definitorischer Regress“). Die Konzeptspezifikation gilt als abgeschlossen, wenn bestimmte Konzepte als allgemein verständlich akzeptiert werden (Schnell et al. 2013, S. 47).

schen Untersuchungseinheiten resultieren. Sowohl die auf einer Grundlage gewonnenen deskriptiven als auch die kausalen Inferenzen wären fehlerhaft." (Wonka 2007, S. 68–69)

Die Anpassung eines existierenden Konzepts an einen neuen Untersuchungsgegenstand wird Re-Spezifizierung genannt. Eine Möglichkeit der Re-Spezifizierung eines Konzepts ist die Veränderung der Anzahl seiner Attribute (Wonka 2007, S. 69).

Re-Spezifizierung eines Konzepts: Politische Beteiligung

Politische Beteiligung gilt als „heart of democracy" (Verba et al. 1995, S. 1). Ohne politische Beteiligung ist Demokratie undenkbar, da sich Demokratie auf das Regieren durch die Bürger bezieht (van Deth 2009, S. 141). Frühe Konzeptspezifikationen hatten insbesondere Aktivitäten im Blick, die unmittelbar mit Wahlen bzw. der Wahlentscheidung in Zusammenhang stehen (z. B. Verba und Nie 1972, S. 2). Die Ausweitung des Repertoires an politischen Aktivitäten in den 1970er und 1980er Jahren führte zu einer Re-Spezifizierung des Konzepts. Eine besonders einflussreiche Konzeptspezifikation hat Kaase (1995, S. 521) vorlegt. Er bezeichnet mit politischer Beteiligung „alle Tätigkeiten (…) die Bürger freiwillig mit dem Ziel unternehmen, Entscheidungen auf den verschiedenen Ebenen des politischen Systems zu beeinflussen" (Kaase 1995, S. 521). Die Re-Spezifizierung zielt darauf ab, die „neuen" Formen politischer Beteiligung wie Unterschriftenaktion oder Demonstration zu berücksichtigen. Mit der Re-Spezifizierung konnte das Konzept auf Untersuchungsgegenstände angewendet werden, die vorher nicht berücksichtigt wurden. Die weitere Ausweitung politischer Partizipationsformen – insbesondere individualisierte Formen politischer Aktivitäten wie politischer Konsum (van Deth 2012) – haben die Debatte um eine Re-Spezifizierung des Konzepts neu entfacht (Fox 2014), die insbesondere in Fachzeitschriften stattfindet (z. B. van Deth 2014; Hooghe 2014; Hosch-Dayican 2014; Theocharis und van Deth 2016, 2018).

5.4 Hinweise für eine erfolgreiche Konzeptspezifikation

Klar spezifizierte Konzepte bilden die Grundlage für eine gehaltvolle wissenschaftliche Argumentation. Nur auf Basis eindeutig spezifizierter Konzepte kann die inhaltlich „gemeinte" Bedeutung eines Konzepts von allen am Kommunikationsprozess beteiligten Personen erfasst und geteilt werden. Dann wissen wir, „worüber geredet wird" (Schnell et al. 2013, S. 46). Wenn die Akteure den verwendeten Konzepten allerdings unterschiedliche Bedeutungen zuschreiben, dann sind Missverständnisse und falsche Schlussfolgerungen die Konsequenz. Die bloße Nennung der Konzepte ist für den Verständigungsprozess nicht ausreichend, vielmehr muss die Intension eines Konzepts offengelegt werden.

In Anlehnung an Wonka (2007) werden in diesem Abschnitt praktische Hinweise für eine „erfolgreiche" Konzeptspezifikation gegeben. Ausführlichere Betrachtungen zur Konzeptspezifikation finden sich bei Gerring (2001) und Goertz (2006).

1. Was sind die zentralen Konzepte?
Im ersten Schritt sind die zentralen Konzepte eines gewählten Forschungsthemas bzw. einer konkreten Forschungsfrage zu identifizieren. Zentrale Konzepte sind in der Regel die abhängige und die unabhängigen Variablen. Wer die Determinanten des sozialen Vertrauens untersuchen möchte, der muss sich mit dem Konzept „Soziales Vertrauen" auseinandersetzen. Für eine Arbeit zur Europäisierung muss das Konzept „Europäisierung" spezifiziert werden. Wer sich mit Ausländerfeindlichkeit beschäftigt, der benötigt eine Konzeptspezifikation von „Ausländerfeindlichkeit". Die Identifikation der zentralen Konzepte steht am Anfang jeder Konzeptspezifikation.

2. Welche Spezifikationen des Konzepts lassen sich in der Fachliteratur unterscheiden?
Zu vielen sozialwissenschaftlichen Konzepten existieren bereits etablierte Spezifikationen. Dann ist es selbstverständlich sinnvoll, sich an entsprechend einschlägigen Konzepten der Fachdisziplin zu orientieren. Dies sichert den kumulativen Charakter wissenschaftlicher Forschung (Wonka 2007, S. 77). Die erfolgreiche Konzeptspezifikation setzt deshalb ebenfalls eine umfangreiche Auseinandersetzung mit der Fachliteratur voraus (für zentrale politikwissenschaftliche Konzepte siehe z. B. Berenskoetter 2016). Neben der Lektüre von einschlägigen Fachlexika und Handbüchern sind insbesondere Fachzeitschriften zu berücksichtigen, da diese heute das zentrale Kommunikationsmedium der Wissenschaft darstellen.

Falls ein Konzept in unterschiedlichen theoretischen Zusammenhängen und variierender Bedeutung verwendet wird, dann ist das für die Forschungsfrage ge-

eignetste Konzept auszuwählen. Diese Festlegung ist zu begründen. Warum ist die gewählte Konzeptspezifikation für das Forschungsthema bzw. die Forschungsfrage geeignet? Dabei ist die möglicherweise erforderliche Anpassung eines existierenden Konzepts an einen neuen Untersuchungsgegenstand zu prüfen (Re-Spezifikation).

3. Welche Intension hat die ausgewählte Konzeptspezifikation?
Das Beispiel „Sozialkapital" zeigt: Trotz gleicher Bezeichnung (Term) unterscheidet sich die Bedeutung der Konzepte. Deshalb muss die Intension einer verwendeten Konzeptspezifikation offengelegt werden. Die Attribute eines Konzepts sind klar und ausführlich darzustellen. Dies schafft die Voraussetzung, dass die „gemeinte" Bedeutung des Konzepts von allen Personen erfasst und geteilt werden kann. Nur auf dieser Basis können Aussagen intersubjektiv nachvollzogen werden. Dies bildet die Voraussetzung für eine gehaltvolle wissenschaftliche Auseinandersetzung.

4. Wie wird Präzision, Eindeutigkeit und Adäquatheit des Konzepts bewertet?
Zwar existieren keine objektiven Kriterien zur Bewertung einer Konzeptspezifikation, aber die Präzision, Eindeutigkeit und Adäquatheit bieten Hinweise zur Zweckmäßigkeit eines Konzepts. Diese Kriterien bieten Anhaltspunkte, ob die Phase der Konzeptspezifikation (zumindest vorläufig) abgeschlossen ist. Die Zweckmäßigkeit einer gewählten Konzeptspezifikation kann allerdings ausschließlich mit Blick auf eine konkrete Forschungsfrage bewertet werden.

Hypothesen 6

Hypothesen sind vermutete Zusammenhänge zwischen zwei Merkmalen. In diesem Kapitel werden Kriterien der sozialwissenschaftlichen Hypothesenbildung vorgestellt. Neben der Unterscheidung in deterministische und probabilistische Hypothesen lassen sich auch zwei Grundformen von Hypothesen charakterisieren: Wenn-dann-Hypothesen und Je-desto-Hypothesen.

6.1 Was sind Hypothesen?

Im allgemeinen Sprachgebrauch handelt es sich bei einer Hypothese um eine unbewiesene Annahme oder Behauptung über einen Sachverhalt (Westle 2009d, S. 81). Nach diesem Verständnis sind Aussagen wie „Ich vermute, dass es am Donnerstag regnet" oder „Ich meine, dass das Wetter immer unberechenbarer wird" bereits sogenannte Alltagsvermutungen oder Alltagshypothesen. Im sozialwissenschaftlichen Kontext sind mit Hypothesen allerdings (begründete) Aussagen über den Zusammenhang von zwei Merkmalen gemeint. Sozialwissenschaftliche Hypothesen sind „Aussagen über *Merkmalszusammenhänge*" (Diekmann 2011, S. 107; Hervorhebung im Original). Mit anderen Worten: Eine sozialwissenschaftliche Hypothese ist eine Aussage über einen Zusammenhang zwischen mindestens zwei Merkmalen (Diekmann 2011, S. 107). Danach können Aussagen wie „Je höher die Bildung, desto größer das Einkommen" oder „Je höher der Anteil an Arbeitslosen, desto geringer die Wahlbeteiligung" als sozialwissenschaftliche Hypothesen bezeichnet werden. Mit beiden Hypothesen werden wir uns in diesem Kapitel noch exemplarisch auseinandersetzen.

Vier Kriterien einer Hypothese Bortz und Döring (2006, S. 4) formulieren vier allgemeine Kriterien, die erfüllt sein müssen, damit eine Aussage als wissenschaft-

© Springer Fachmedien Wiesbaden GmbH, ein Teil von Springer Nature 2018

M. Tausendpfund, *Quantitative Methoden in der Politikwissenschaft*,
Grundwissen Politik, https://doi.org/10.1007/978-3-658-20698-7_6

liche Hypothese bezeichnet werden kann: Erstens muss sich eine wissenschaftliche Hypothese auf reale Sachverhalte beziehen, die empirisch untersuchbar sind. Zweitens muss es sich um eine allgemeingültige – über den Einzelfall oder ein singuläres Ereignis hinausgehende – Aussage handeln. Drittens muss eine wissenschaftliche Hypothese zumindest implizit die Formalstruktur eines sinnvollen Konditionalsatzes (Wenn-dann oder Je-desto) aufweisen. Viertens muss eine Hypothese falsifizierbar sein. Auf der Basis dieser Kriterien präsentieren Bortz und Döring (2006, S. 4) folgende Definition für eine wissenschaftliche Hypothese:

► **Definition: Hypothesen**
„Wissenschaftliche Hypothesen sind Annahmen über reale Sachverhalte (empirischer Gehalt, empirische Untersuchbarkeit) in Form von Konditionalsätzen. Sie weisen über den Einzelfall hinaus (Generalisierbarkeit, Allgemeinheitsgrad) und sind durch Erfahrungsdaten widerlegbar (Falsifizierbarkeit)."

Unsere zwei Beispielhypothesen erfüllen die vier genannten Kriterien: „Bildung und Einkommen" sowie „Arbeitslosigkeit und Wahlbeteiligung" sind erstens reale Sachverhalte, die empirisch untersuchbar sind. Zweitens handelt es sich um allgemeingültige Aussagen, die drittens eine entsprechende Konditionalstruktur aufweisen („Je-höher-desto"). Viertens können beide Hypothesen auf Basis entsprechender Befunde auch falsifiziert, also widerlegt werden.

Insbesondere Alltagshypothesen genügen diesen formulierten Kriterien häufig nicht. Der Satz „Ich vermute, dass es am Donnerstag regnet" beinhaltet keine Aussage über einen Zusammenhang von zwei Merkmalen. Deshalb handelt es sich auch nicht um eine wissenschaftliche Hypothese. Die Hypothese „Je gläubiger ein Mensch, desto eher kommt er in den Himmel" entspricht auch nicht sozialwissenschaftlichen Kriterien, da sich der „Himmel" einer empirischen Untersuchbarkeit entzieht. Schließlich sind auch analytische Sätze (z. B. „Wenn der Hahn kräht auf dem Mist, ändert sich das Wetter oder es bleibt, wie es ist.") keine Hypothese, da sie nicht widerlegbar sind.

! In einem wissenschaftlichen Text ist es aber nicht ausreichend, einfach eine Hypothese zu formulieren, die den genannten Kriterien entspricht. Vielmehr müssen – auf Basis theoretischer Überlegungen – Argumente präsentiert werden, die den in der Hypothese formulierten Zusammenhang erläutern. Eine Hypothese ist also stets in eine wissenschaftliche Argumentation einzubetten. In der Regel lässt sich eine Hypothese aus der Darstellung des Forschungstands ableiten.

Das Beispiel im Kasten soll deutlich machen, dass Hypothesen stets in den Forschungsstand eingebunden sein bzw. mit dem Forschungsstand in Verbindung ste-

Hypothesen mit dem Forschungsstand verbinden

Die Zustimmung der Bürger zur Europäischen Union gilt als Grundvoraussetzung für die Stabilität und die Entwicklungsperspektiven der Staatengemeinschaft (z. B. McLaren 2006; Kaina 2009). Mittlerweile liegen zahlreiche Arbeiten vor, die sich mit möglichen Faktoren dieser Zustimmung beschäftigen. Einen der einflussreichsten Ansätze, um die Haltung der Bürger zur EU zu erklären, haben Gabel und Palmer (1995, S. 4) vorgelegt. Sie argumentieren auf Basis eines utilitaristischen Erklärungsmodells und behaupten, dass die Bewertung der Union auf den individuellen Kosten und Nutzen der EU-Mitgliedschaft basiert (siehe auch Gabel 1998a; 1998b). Dieser Überlegung liegt die Annahme zugrunde, dass die politischen Entscheidungen der Staatengemeinschaft Gewinner und Verlierer generieren. Personen, die von der EU profitieren, bewerten die Staatengemeinschaft positiver als Personen, die sich als Verlierer des Integrationsprozesses wahrnehmen. Auf Basis dieses Erklärungsmodells lässt sich folgende Hypothese formulieren: „Je größer die Vorteile durch die EU, desto größer die Zustimmung zur EU."

hen sollten. Insbesondere in Fachaufsätzen wird der Forschungstand meist knapp dargestellt und auf Basis der Argumente eine zu prüfende Hypothese präsentiert. Dabei werden Hypothesen im Text häufig auch hervorgehoben (z. B. durch einen Einschub, Nummerierung oder auch kursive Schriftweise). Deshalb: Prüfen Sie beim Lesen (quantitativ) orientierter Studien, ob der Autor bzw. die Autorin die Hypothesen explizit formuliert hat und diese auch den genannten Kriterien entsprechen.

Deterministische und probabilistische Hypothesen Hypothesen können deterministisch oder probabilistisch (statistisch) sein. Deterministische Hypothesen werden vor allem in den Naturwissenschaften formuliert. Formal wird eine deterministische Hypothese ausgedrückt als „Wenn A, dann B". Das Fallgesetz in der Physik wäre ein Beispiel für eine deterministische Hypothese. Unter spezifischen Bedingungen gilt das Fallgesetz in der Physik für alle Körper.

In den Sozialwissenschaften sind deterministische Hypothesen eher die Ausnahme. Dies lässt sich an unserem Beispiel „Je höher die Bildung, desto größer das Einkommen" leicht illustrieren. Sicherlich wird es auch Personen geben, die trotz hoher Bildung nur ein niedriges Einkommen erzielen. Unsere (deterministische) Hypothese wäre bereits widerlegt, wenn nur eine einzige Person mit hohem Bil-

dungsabschluss ein niedriges Einkommen erzielen würde. Allerdings haben wir es in den Sozialwissenschaften mit „Untersuchungsobjekten" zu tun, „die sich unter anderem durch hochgradige Individualität, Komplexität und durch Bewusstsein auszeichnen" (Bortz und Döring 2006, S. 11; ähnlich auch Westle 2009d, S. 82). Deshalb werden in den Sozialwissenschaften in der Regel keine deterministischen, sondern probabilistische (statistische) Hypothesen formuliert. Bortz und Döring (2006, S. 10) sprechen in diesem Zusammenhang auch von Wahrscheinlichkeitsaussagen. Eine statistische Hypothese lässt sich wie folgt formulieren: Die Wahrscheinlichkeit von B ist bei Auftreten von A größer als die Wahrscheinlichkeit von B bei Auftreten von Nicht-A (Diekmann 2011, S. 127). Etwas umgangssprachlicher: Personen mit hoher Bildung haben wahrscheinlich ein größeres Einkommen als Personen mit niedriger Bildung. Probabilistische Hypothesen können durch konträre Einzelfälle nicht widerlegt werden. Bei solchen Hypothesen wird ein vorhergesagter Merkmalswert (hier: Einkommen) nur mit einer gewissen Wahrscheinlichkeit auftreten. In den Sozialwissenschaften handelt es sich praktisch immer um probabilistische Hypothesen, weshalb auf den Zusatz probabilistisch oder statistisch in der Regel verzichtet wird. Zur (empirischen) Prüfung solcher Wahrscheinlichkeitsaussagen kommt deshalb auch der Statistik eine wichtige Rolle zu (siehe auch Abschnitt 12.2).

Alternativ- und Nullhypothese In der empirischen Sozialforschung werden Hypothesen, die untersucht werden, als Alternativhypothesen bezeichnet. Die Alternativhypothese beinhaltet die interessierende Aussage, zum Beispiel „Je höher die Bildung, desto höher das Einkommen" oder „Je stärker das politische Interesse, desto wahrscheinlicher die Wahlbeteiligung". Alternativhypothesen werden häufig auch mit H_1 abgekürzt (Gehring und Weins 2009, S. 273).

Jeder Alternativhypothese lässt sich eine Nullhypothese gegenüberstellen. Die Nullhypothese verneint den in der Alternativhypothese formulierten Zusammenhang. Mit Blick auf die oben formulierten Alternativhypothesen sind die Aussagen „Zwischen Bildung und Einkommen besteht kein Zusammenhang" und „Zwischen dem politischen Interesse und der Wahlbeteiligung besteht kein Zusammenhang" die Nullhypothesen. Die Nullhypothese steht also im Widerspruch zur Alternativhypothese; sie wird mit H_0 abgekürzt (Gehring und Weins 2009, S. 273).

H_0 und H_1 bilden damit ein Hypothesenpaar (Döring und Bortz 2016, S. 660; Braunecker 2016, S. 268), das mögliche empirischen Zusammenhänge von zwei Merkmalen abbildet. In der Forschungspraxis wird allerdings lediglich die Alternativhypothese explizit formuliert, die dann gegen die Nullhypothese getestet wird. Das Vorgehen ist dabei konservativ. Nur, wenn wir uns sehr sicher sind, dass die Alternativhypothese zutrifft, wird die Alternativhypothese vorläufig an-

genommen und die Nullhypothese abgelehnt. Die Inferenzstatistik bietet Kriterien, um zu entscheiden, wann eine Alternativhypothese vorläufig angenommen oder abgelehnt wird.

Grundformen von Hypothesen Nach Bortz und Döring (2006, S. 4) müssen wissenschaftliche Hypothesen zumindest implizit die Formalstruktur eines sinnvollen Konditionalsatzes aufweisen. Mit Diekmann (2011, S. 125–133) lassen sich zwei Grundformen von Hypothesen unterscheiden: Wenn-dann-Hypothesen und Je-desto-Hypothesen.

Wenn-dann-Hypothese Bei einer Wenn-dann-Hypothese haben die beiden Sachverhalte, zwischen denen ein Zusammenhang angenommen wird, nur zwei Ausprägungen. Es handelt sich um sogenannte dichotome Merkmale. Bei Bildung und Alter handelt es sich eigentlich um metrische Merkmale.[12] Allerdings können metrische Merkmale (mit Informationsverlust) in dichotome Merkmale transformiert werden. Wir können beispielsweise bei Bildung zwischen niedriger und hoher Bildung unterscheiden, analog kann Einkommen in niedriges und hohes Einkommen klassifiziert werden. Eine mögliche Hypothese wäre: „Wenn Menschen eine hohe Bildung haben, dann haben sie ein hohes Einkommen." Die Wenn-Komponente wäre die unabhängige Variable (uV), die Dann-Komponente die abhängige Variable (aV). In unserem Beispiel ist Bildung die uV und das Einkommen die aV. Wenn-dann-Hypothesen werden formuliert, wenn uV und aV nur dichotom vorliegen. Typische Beispiele für dichotome Merkmale sind das Geschlecht (Frau/Mann) oder auch die Region (Ost/West). Mögliche Wenn-dann-Hypothesen sind „Frauen haben ein geringeres politisches Wissen als Männer" (z. B. Dow 2009) oder „Westdeutsche beteiligen sich eher an Bundestagswahlen als Ostdeutsche" (z. B. Steinbrecher und Rattinger 2011). Aus sprachlich-ästhetischen Gründen entsprechen beide Hypothesen nicht der klassischen Wenn-dann-Form, aber die erforderliche Konditionalstruktur ist in den Formulierungen erkennbar. Eine alternative Formulierung wäre: „Wenn das Geschlecht weiblich ist, dann ist das politische Wissen geringer als bei Männern" oder „Wenn eine Person in den alten Bundesländern wohnt, dann ist die Wahrscheinlichkeit der Wahlbeteiligung höher, als wenn eine Person in den neuen Bundesländern wohnt".

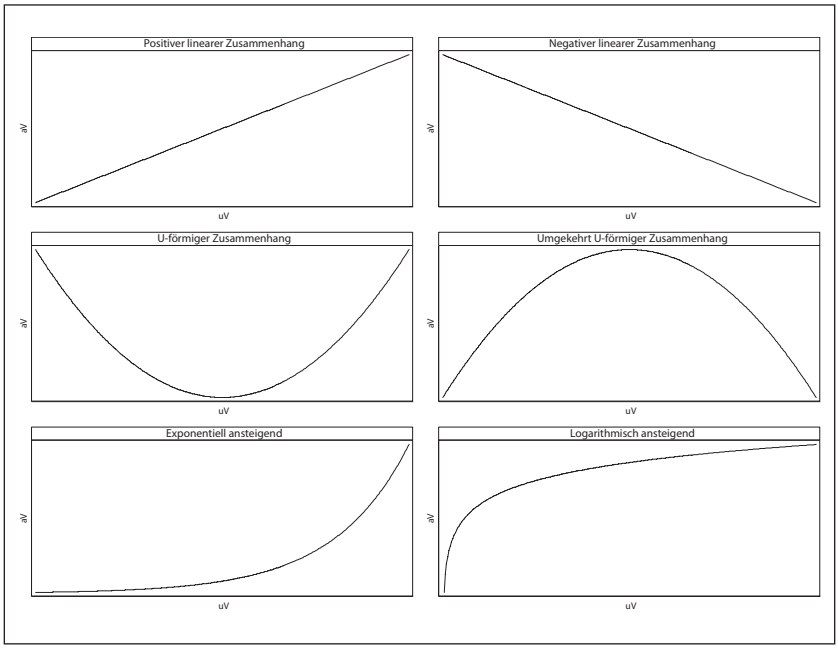
12 Dieses Kriterium bezieht sich auf das Skalenniveau der Variablen. Es lassen sich – vereinfacht – vier Skalenniveaus unterscheiden: Nominal-, Ordinal-, Intervall- und Ratioskalen (siehe z. B. Schnell et al. 2013, S. 132). In Kapitel 7 werden die Skalenniveaus vorgestellt.

Je-desto-Hypothese Bei Bildung und Einkommen handelt es sich in der Regel nicht um dichotome Merkmale, sondern um Merkmale, die als Rangfolge interpretierbar sind. Das Bildungsniveau lässt sich relativ einfach in Bildungsjahren erfassen, und das Einkommen in Euro abbilden. Deshalb ist für die Formulierung eines Zusammenhangs zwischen Bildung und Einkommen die zweite Grundform von Hypothesen besser geeignet: Je-desto-Hypothesen. Bei einer Je-desto-Hypothese müssen die Ausprägungen der unabhängigen und abhängigen Variablen (mindestens) als Rangfolge interpretierbar sein. Dies ist bei Bildung und Einkommen (und vielen anderen Variablen) problemlos möglich. Deshalb bietet sich für unser Beispiel folgende Hypothese an: „Je höher die Bildung, desto höher das Einkommen." Ein Zusammenhang liegt dann vor, wenn das Einkommen in Abhängigkeit von der Bildung systematisch zunimmt. Etwas allgemeiner formuliert: „Je höher die uV, desto größer die aV." Die Bezeichnungen uV und aV sind „Platzhalter" für mögliche interessante Merkmale. Bei unserem Bildung-Einkommen-Beispiel wird ein positiver Zusammenhang unterstellt, denkbar sind natürlich auch negative Beziehungen. „Je höher die Arbeitslosenquote, desto geringer die Wahlbeteiligung" ist eine Je-desto-Hypothese, die einen negativen Zusammenhang zwischen Arbeitslosenquote und Wahlbeteiligung auf der Makroebene zum Ausdruck bringt. Die Art des vermuteten Zusammenhangs (positiv oder negativ) basiert auf theoretischen Überlegungen und ist bei der Formulierung wissenschaftlicher Texte zu erläutern.

Mögliche Zusammenhänge bei Je-desto-Hypothesen In Abbildung 6 sind einige Beispiele für mögliche Zusammenhänge zwischen uV und aV dargestellt. Die am häufigsten postulierten Zusammenhänge sind sicherlich positiv lineare bzw. negativ lineare Beziehungen. „Je höher die Bildung, desto größer das Einkommen" oder „Je größer der Lernaufwand, desto besser das Klausurergebnis" sind Beispiele für (vermutete) positive Zusammenhänge zwischen uV und aV. „Je größer die Angst vor der Statistikklausur, desto geringer die Punktzahl" ist ein Beispiel für einen (vermuteten) negativen linearen Zusammenhang.

Allerdings gibt es auch Zusammenhänge zwischen uV und aV, die nicht als lineare Beziehungen darstellbar sind. Aus der Wahlforschung ist ein umgekehrt U-förmiger Zusammenhang zwischen Alter und Wahlbeteiligung bekannt. Die Wahrscheinlichkeit der Wahlbeteiligung nimmt mit dem Alter langsam zu, erreicht im mittleren Alter das Maximum und nimmt dann wieder ab. Dieser (empirisch vielfach bestätigte) Zusammenhang basiert auf einem angenommenen Lebenszykluseffekt. Vereinfacht formuliert, steigt die Wahlbeteiligung zunächst mit der Einbindung in das Berufsleben und nimmt im höheren Alter aufgrund von Krankheit wieder ab (für eine ausführliche Betrachtung siehe z. B. Falter und Gehring 1998; Abendschön und Roßteutscher 2011; Caballero 2014, S. 468).

Abbildung 6 Mögliche Beziehungen zwischen zwei Merkmalen



Quelle: Eigene Darstellung

Auch lassen sich U-förmige Zusammenhänge zwischen Merkmalen postulieren. Auf Basis der klassischen Studie „American Soldier" (Stouffer 1949) beschreibt Diekmann (2011, S. 27) einen U-förmigen Zusammenhang zwischen Arbeitszufriedenheit und Aufstiegschancen. Bei mittleren Aufstiegschancen in einer Firma könnte die allgemeine Arbeitszufriedenheit niedriger sein als bei geringen oder hohen Aufstiegschancen. Warum?

> *„Wenn die Beförderungsrate von geringem Niveau aus wächst, gibt es natürlich nicht nur beförderte, sondern auch nichtbeförderte Angestellte: Zwar wird im Allgemeinen die Zufriedenheit der aufgestiegenen Angestellten zunehmen, aber auch die Unzufriedenheit bei den übergangenen Arbeitskräften. Bei den Verlierern des Aufstiegswettbewerbs, die sich mit ihren beförderten Kolleginnen oder Kollegen vergleichen, wächst das Ausmaß der sogenannten relativen Deprivation. Solange der Deprivationseffekt den Zufriedenheitsgewinn der beförderten Angestellten überwiegt, wird mit der Beförderungsrate insgesamt das Ausmaß der Zufriedenheit in der Arbeitsgruppe absinken." (Diekmann 2011, S. 27)*

Auch können die Zusammenhänge zwischen uV und aV exponentiell oder logarithmisch steigend (und natürlich auch exponentiell oder logarithmisch fallend). Beim Lernen einer Fremdsprache wäre beispielsweise ein exponentiell steigender Verlauf denkbar. Das Wissen nimmt erst langsam, aber stetig zu. Mit steigendem Wissen können sich allerdings immer mehr Verknüpfungen zwischen einzelnen Wissensbereichen bilden, sodass das Wissensniveau exponentiell zunimmt. Beim Zusammenhang zwischen Bildung und Einkommen ist eine logarithmisch ansteigende Beziehung vermutlich angemessener als eine lineare positive Beziehung. Vermutlich werden die ersten Bildungsjahre einen stärkeren Effekt auf das Einkommen haben als spätere Bildungsjahre.

Individual-, Kollektiv- und Kontexthypothesen Hypothesen formulieren Zusammenhänge zwischen Merkmalen. Diese Merkmale lassen sich analytischen Ebenen zuordnen. Die wichtigsten analytischen Ebenen sind die Mikro- und die Makroebene (siehe Abschnitt 2.4). Bei unserem Beispiel „Je größer die Bildung, desto höher das Einkommen" lassen sich beide Merkmale der Mikroebene (Individualebene) zuordnen. Können beide Merkmale einer Hypothese der Individualebene zugeordnet werden, dann handelt es sich um eine Individualhypothese. Die Merkmale der Hypothese „Je größer der Anteil an Arbeitslosen, desto geringer die Wahlbeteiligung" lassen sich dagegen nicht der Individualebene zuordnen. Es handelt sich um Merkmale der Makroebene (bzw. auch Systemebene oder gesellschaftliche Ebene). Es können damit die Arbeitslosenquote und die Wahlbeteiligung der EU-Staaten (nationalstaatliche Ebene) oder auch die Arbeitslosenquote und die Wahlbeteiligung in den Gemeinden Nordrhein-Westfalens gemeint sein. Die konkrete analytische Ebene muss sich bei solchen Hypothesen aus dem Forschungskontext ergeben. Unabhängig davon, ob Zusammenhänge auf nationaler, regionaler oder lokaler Ebene postuliert werden, werden solche Hypothesen allgemein als Kollektivhypothesen bezeichnet. Wichtig ist nur, dass sich beide Merkmale (also uV und aV) der Makroebene zuordnen lassen. Neben Individual- und Kollektivhypothese existiert mit Kontexthypothesen ein dritter Hypothesentyp, der Makro- und Mikroebene verbindet. In Kontexthypothesen ist die unabhängige Variable der Makroebene zuzuordnen, die abhängige Variable der Individualebene. Für die empirische Forschung ist die Differenzierung zwischen Individual-, Kollektiv- und Kontexthypothesen zentral, da damit eigenständige Forschungsstrategien verbunden sind. Deshalb werden diese Typen von Hypothesen in den folgenden Abschnitten etwas ausführlicher betrachtet.

6.2 Individualhypothesen

Eine Individualhypothese postuliert einen Zusammenhang zwischen zwei Merkmalen auf der Individualebene. Die beiden Merkmale – also die uV und die aV – lassen sich der Individualebene zuordnen. Typische Merkmale auf der Individualebene sind Geschlecht, Alter, Bildung, Einkommen, Einstellungen oder Verhalten. „Je größer die Bildung, desto höher das Einkommen" ist eine klassische Individualhypothese. Bildung ist die unabhängige Variable (uV) und Einkommen die abhängige Variable (aV).

Zusammenhänge auf Individualebene Bei Individualhypothesen stehen Zusammenhänge zwischen Individualmerkmalen im Blickpunkt. Die entsprechende Forschungsstrategie wird Individualdatenanalyse genannt (siehe auch Abschnitt 9.1.1). Bei einer Individualdatenanalyse wird aber nicht ein konkretes Individuum betrachtet, sondern es wird versucht, allgemeine Zusammenhänge zwischen zwei Merkmalen auf der Individualebene zu entdecken. Typische Fragestellungen auf der Individualebene sind: Warum unterscheidet sich das Einkommen von Personen? Warum unterscheidet sich das Wahlverhalten zwischen Bürgern? Warum bewerten einige Menschen die Europäische Union positiv, während andere Personen die Staatengemeinschaft ablehnen?

Bei einer Individualdatenanalyse werden Faktoren gesucht, die die Unterschiede zwischen den Individuen erklären können. Bildung ist ein möglicher Erklärungsfaktor, um Einkommensunterschiede zwischen Personen zu erklären. Sicherlich gibt es einzelne Personen, die auch mit niedriger Bildung ein hohes Einkommen erzielen. Bei höherer Bildung ist ein höheres Einkommen aber wahrscheinlicher als bei niedriger Bildung. Unterschiede der abhängigen Variable (hier: Einkommen) sind in der Regel aber nicht nur auf eine unabhängige Variable (hier: Bildung) zurückzuführen, sondern sind das Ergebnis des Zusammenwirkens vieler unabhängiger Variablen. Das Einkommen wird nicht nur von der Bildung abhängen, sondern vermutlich auch vom Geschlecht, Alter, Beruf und anderen Faktoren. Bei einer Individualdatenanalyse werden – auf Basis theoretischer Überlegungen – mögliche erklärende Faktoren (uVs) gesucht, die die Unterschiede zwischen den Individuen erklären können. Bei einer empirischen Studie zur Einkommensungleichheit werden deshalb auch mehrere Hypothesen formuliert, um mögliche erklärende Faktoren zu identifizieren. Je nach Fragestellung stehen verschiedene „Werkzeuge" (statistische Verfahren) bereit, die bei der konkreten Hypothesenprüfung genutzt werden können (für einen Überblick siehe Wolf und Best 2010a).

6.3 Kollektivhypothesen

Eine Kollektivhypothese postuliert einen Zusammenhang zwischen zwei Merkmalen auf der Makroebene. Beide Merkmale einer Hypothese (uV und aV) lassen sich der Makroebene zuordnen. Die spezifische analytische Ebene – lokal, regional oder national – lässt sich meist aus dem Forschungskontext herleiten. Die Hypothese „Je höher der Anteil an Arbeitslosen, desto geringer die Wahlbeteiligung" kann beispielsweise auf lokaler, regionaler oder europäischer Ebene empirisch untersucht werden. So können die Wahlbeteiligung und Arbeitslosenquote in den 396 Gemeinden Nordrhein-Westfalens (lokale Ebene), in den 16 Bundesländern (regionale Ebene) und in den 28 EU-Staaten (europäische Ebene) zueinander in Beziehung gesetzt werden. Der Zusammenhang zwischen Arbeitslosenquote und Wahlbeteiligung kann aber auch auf Stadtteilebene oder auf Wahlbezirksebene innerhalb einer Gemeinde analysiert werden (z. B. Schäfer 2012). Es müssen lediglich empirische Informationen zur Arbeitslosenquote und Wahlbeteiligung auf der jeweiligen Ebene verfügbar sein.

Aggregatdatenanalyse Die Analyse von Zusammenhängen auf der Makroebene, also die empirische Prüfung von Kollektivhypothesen, stellt eine eigenständige Forschungsstrategie dar, die als Aggregatdatenanalyse bezeichnet wird (siehe z. B. Jahn 2009; Wagschal und Jäckle 2011; Graeff 2014; Lauth et al. 2015, S. 83–117). Die Vorteile dieser Analysestrategie sind offensichtlich: Die Daten sind meist schnell und kostengünstig verfügbar, es existieren (vermutlich) keine systematischen Erhebungsfehler und es besteht die Möglichkeit, langfristige Entwicklungsprozesse

Verfügbarkeit von Aggregatdaten

Statistische Informationen zu den einzelnen regionalen Ebenen sind häufig bei den jeweiligen statistischen Ämtern verfügbar. Für Daten auf europäischer Ebene ist Eurostat der erste Ansprechpartner, auf der nationalen Ebene die entsprechenden nationalen Ämter (in Deutschland z. B. das Statistische Bundesamt). Daten für die regionale Ebene (z. B. Bundesländer) sammeln die statistischen Landesämter; in größeren Städten und Gemeinden gibt es eigene lokale Statistikämter. Bei wissenschaftlichen Texten (auch bei Haus- und Abschlussarbeiten) sind ausschließlich offizielle (amtliche) Daten zu verwenden; insbesondere die Datengrundlage ist dabei offenzulegen. Wikipedia ist hierfür keine geeignete Quelle. Kapitel 8 gibt eine Übersicht über wichtige Bezugsquellen von Daten.

nachzuzeichnen. Allerdings ist die Aggregatdatenanalyse auch mit Problemen behaftet. Ein zentrales Problem ist der Schluss von der Makro- auf die Mikroebene (oder auch von der Aggregat- auf die Individualebene). Häufig werden empirische Zusammenhänge der Makroebene einfach auf die Mikroebene übertragen. Ein solcher Schluss liegt beispielsweise vor, wenn aus dem Ergebnis „Je höher die Arbeitslosenquote, desto geringer die Wahlbeteiligung" gefolgert wird: Arbeitslose wählen seltener als Erwerbstätige. Ein solcher Schluss kann naheliegend, muss aber nicht korrekt sein. Ein falscher Schluss von der Kollektiv- auf die Individualebene wird als ökologischer Fehlschluss bezeichnet.

Ökologischer Fehlschluss Der Begriff „Ökologischer Fehlschluss" stammt von Robinson (1950). Auf der Makroebene belegt er einen Zusammenhang zwischen dem Anteil der Farbigen und dem Anteil der Bevölkerung, der nicht lesen und schreiben kann. Als Kollektivhypothese formuliert: Je größer der Anteil der Farbigen, desto höher die Analphabetenrate. Aus dem (nachgewiesenen) empirischen Befund könnte jetzt folgender Schluss gezogen werden: Afroamerikaner können seltener lesen und schreiben als Angehörige anderer ethnischer Gruppen. Allerdings lässt sich dieser Zusammenhang auf der Individualebene nicht belegen. Robinson (1950, S. 357) bilanziert:

> *„From a practical standpoint, therefore, the only reasonable assumption is that an ecological correlation is almost certainly not equal to its corresponding individual correlation."*

Zusammengefasst: Eine Kollektivhypothese postuliert einen Zusammenhang zwischen zwei Merkmalen auf der Makroebene. Aus bestätigten Kollektivhypothesen folgt allerdings nicht notwendigerweise derselbe Zusammenhang auf der Individualebene. Ein auf der Kollektivebene festgestellter Zusammenhang kann auf der Individualebene zwar auf demselben Zusammenhang beruhen, er kann aber auch auf einem gegenteiligen Zusammenhang beruhen oder es liegt auf der Individualebene überhaupt keine entsprechende Beziehung vor.

Der Schluss von der Kollektiv- bzw. Aggregatebene auf die Individualebene ist mit „Stolpersteinen gepflastert" (Diekmann 2011, S. 136). Ein in Anlehnung an Diekmann (2011, S. 136) entwickeltes Extrembeispiel soll diese „Stolpersteine" illustrieren. In Tabelle 7 finden sich die (fiktiven) Angaben über die Anteile von Arbeitslosen und NSDAP-Wählern (jeweils in Prozent) in einer Gemeinde mit drei Wahlbezirken und jeweils 100 Wählern.

Auf der Kollektivebene (hier: Wahlbezirke) ist der Zusammenhang zwischen Arbeitslosenquote und NSDAP-Wahl offensichtlich: Je größer die Arbeitslosenquote, desto höher der Anteil der NSDAP-Wähler. Daher ist der Schluss verlockend, dass die NSDAP insbesondere von Arbeitslosen gewählt wurde. Das kann

Tabelle 7 Anteile von Arbeitslosen und NSDAP-Wählern (fiktive Daten)

	Bezirk 1	Bezirk 2	Bezirk 3
Anteil Arbeitslose in Prozent	10	20	30
Anteil NSDAP-Wähler in Prozent	10	20	30

sein, muss es aber nicht. In Tabelle 8 ist das Wahlverhalten in den Stimmbezirken in Kreuztabellen auf der Individualebene dargestellt. In jedem Wahlbezirk gab es 100 Wähler. Im ersten Wahlbezirk gab es zehn Arbeitslose, aber keiner davon hat die NSDAP gewählt. Im zweiten Wahlbezirk gab es 20 Arbeitslose, aber keiner davon hat die NSDAP gewählt. Im dritten Wahlbezirk gab es 30 Arbeitslose, aber keiner davon hat die NSDAP gewählt. Zwischen dem Anteil der Arbeitslosen und dem Anteil an NSDAP-Wählern existiert zwar auf der Makroebene (hier: Wahlbezirke) ein Zusammenhang, nicht aber innerhalb der Wahlbezirke.

Deshalb sind Aggregatdaten für die Analyse individueller Zusammenhänge zurückhaltend zu bewerten. In der (lokalen) Wahlberichterstattung finden sich häufiger Spekulationen über Zusammenhänge wie „Je höher der Arbeiteranteil,

Klassiker der Sozialwissenschaften: Arbeitslosenquote und NSDAP-Stimmenanteil

Ein im deutschen Sprachraum bekannter ökologischer Fehlschluss bezieht sich auf folgende Kollektivhypothese: Je größer die Arbeitslosenquote, desto höher der Stimmenanteil der NSDAP. Auf der Makroebene haben Frey und Weck (1981) einen fast perfekten Zusammenhang zwischen Arbeitslosenquote und NSDAP-Stimmenanteil nachgewiesen. Diese Beziehung wurde anschließend – nicht von den Autoren selbst (Frey und Weck 1981, S. 25; siehe aber Bürklin und Klein 1998, S. 37) – häufig als Bestätigung des Zusammenhangs zwischen Arbeitslosigkeit und NSDAP-Wahl auf der Individualebene gewertet. Umfangreiche Analysen von Falter et al. (1983) widersprechen diesem postulierten Individualzusammenhang. Danach haben Arbeitslose überdurchschnittlich häufig KPD und SPD, jedoch nur in geringerem Maße NSDAP gewählt. Zudem unterscheidet sich das Wahlverhalten in Abhängigkeit von der Berufsgruppe. Bei erwerbslosen Arbeitern war die NSDAP-Wahl unterdurchschnittlich, bei erwerbslosen Angestellten dagegen stärker ausgeprägt (Falter et al. 1983, S. 550).

Tabelle 8 Abstimmungsverhalten auf Individualebene (fiktive Daten)

	Bezirk 1		Bezirk 2		Bezirk 3	
	Arbeitslos	Nicht-Arbeitslos	Arbeitslos	Nicht-Arbeitslos	Arbeitslos	Nicht-Arbeitslos
NSDAP-Wahl	0	10	0	20	0	30
andere Parteien	10	80	20	60	30	40

desto höher der Anteil der Stimmen für die SPD“ oder „Je höher der Anteil an Katholiken, desto höher der Anteil der Stimmen für die CDU“. Solche auf der Makroebene nachgewiesenen Zusammenhänge dürfen aber nicht einfach auf die Individualebene übertragen werden: Es besteht die Gefahr eines ökologischen Fehlschlusses. Deshalb nutzt die Wahlforschung für die Analyse des (individuellen) Wahlverhaltens meist auch Individualdaten. Allerdings existieren nicht für alle wahlsoziologischen Fragestellungen entsprechende Individualdaten. Die historische Wahlforschung (z. B. Falter 1991, 2013; Debus 2016) ist zwingend auf Makrodaten angewiesen, da keine Individualdaten existieren. Auch für die Analysen der Wahlbeteiligung und des Wahlverhaltens auf lokaler Ebene ist die Wahlforschung auf Makrodaten angewiesen, da diese Daten einfach und kostengünstig zur Verfügung stehen.

6.4 Kontexthypothesen

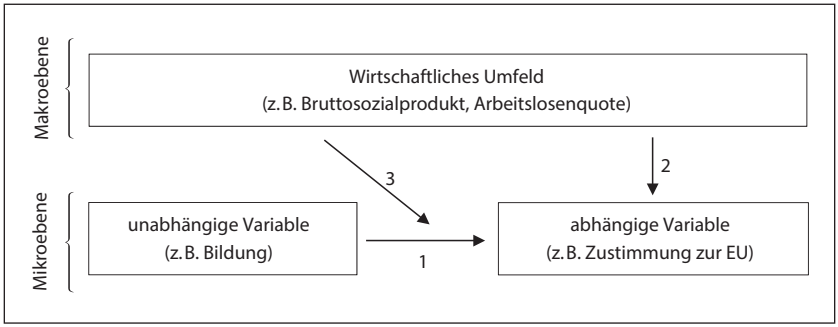
Individualhypothesen postulieren einen Zusammenhang zwischen Merkmalen auf der Individualebene (z. B. Je höher das Alter, desto größer das Einkommen). Kollektivhypothesen formulieren eine Beziehung zwischen Merkmalen auf der Makroebene (z. B. Je höher die Arbeitslosenquote, desto geringer die Wahlbeteiligung). Bei Individual- und Kollektivhypothesen sind die jeweiligen Merkmale jeweils einer Ebene zuzuordnen. Bei Individualhypothesen sind die Merkmale der Mikroebene, bei Kollektivhypothesen der Makroebene zuzuordnen. Kontexthypothesen sind ein dritter Hypothesentyp, der Mikro- und Makroebene verbindet. Die unabhängige Variable (z. B. Arbeitslosenquote) ist ein Makromerkmal, die abhängige Variable (z. B. individuelle Wahlbeteiligung) ist ein Mikromerkmal. Die Grundidee von Kontexthypothesen vermittelt ein Gedankenexperiment von Russell J. Dalton und Christopher J. Anderson (2011, S. 3):

> *„Imagine two people with identical characteristics, such as their age, education, social class, and gender. Also, assume that they hold identical political values. If these two individuals were deciding whether to vote in the next national elections or whom to vote for in that election, we expect they would make identical choices. But now, consider if they lived in two nations with different electoral rules, different numbers of parties, or sets of parties that offered divergent policy programs. How would these variations in political context affect their political behavior? And how would individual characteristics find expression in different political contexts?"*

Die zentrale Annahme dieses Gedankenexperiments ist, dass individuelle Einstellungen und Verhaltensweisen nicht nur eine Folge von individuellen Merkmalen sind, sondern auch das Resultat der Einwirkung von kontextuellen Faktoren (Makrofaktoren). Die empirische Prüfung von Kontexthypothesen stellt – wie die Individual- und Aggregatdatenanalyse – eine eigenständige Forschungsstrategie dar, die als Kontextanalyse bezeichnet wird (z. B. Esser 1999a, S. 426–434). Häufig wird diese Forschungsstrategie auch Mehrebenenanalyse (z. B. Pötschke 2014) oder im englischen Multilevel Analysis genannt (z. B. Hox und Roberts 2011; Snijders und Bosker 2012).

In Abbildung 7 wird die Logik einer Kontextanalyse an einem Beispiel illustriert. Die abhängige Variable ist die (individuelle) Zustimmung zur Europäischen Union. Für die Mikroebene (Ziffer 1) lässt sich eine klassische Individualhypothese formulieren: Je höher die Bildung, desto größer die Zustimmung zur Europäischen Union (Tausendpfund 2013b, S. 114–119). Sicherlich wird es noch weitere Individualmerkmale geben, die einen Einfluss auf die Zustimmung zur Europäischen Union ausüben. „Je negativer die Einstellung gegenüber Immigranten,

Abbildung 7 Kontextanalyse



Quelle: Eigene Darstellung

desto geringer ist die Zustimmung zur Europäischen Union", wäre eine mögliche (zweite) Individualhypothese (Braun und Tausendpfund 2013).

Die Annahme, dass die Zustimmung zur Europäischen Union aber ausschließlich auf individuelle Eigenschaften der Person zurückzuführen ist, wird der sozialen Realität vermutlich nicht völlig gerecht. Schließlich leben die Bürgerinnen und Bürgern in sehr unterschiedlichen Nationalstaaten, die sich in politischer (z. B. Regierungssystem), gesellschaftlicher (z. B. Ausländerquote) und wirtschaftlicher Hinsicht (z. B. Arbeitslosenquote) sehr stark unterscheiden. Solche Merkmale – z. B. Regierungssystem, Ausländerquote und wirtschaftliches Umfeld – haben allerdings nichts mit dem Individuum zu tun. Es handelt sich um Merkmale der sogenannten Kontextebene (konkret: der nationalen Ebene).

Kontextanalyse In unserem Beispiel wird ein Zusammenhang zwischen wirtschaftlichem Umfeld und Zustimmung zur Europäischen Union postuliert (Ziffer 2 in Abbildung 7). Eine mögliche Kontexthypothese wäre: Je besser das wirtschaftliche Umfeld (uV), desto höher ist die Zustimmung zur Europäischen Union (aV). Die unabhängige Variable ist der Kontextebene (Makroebene) zuzuordnen, die abhängige Variable der Individualebene (Mikroebene).

Bei einer Kontextanalyse wird allerdings nicht nur der Einfluss des Kontextmerkmals auf die abhängige Variable geschätzt, sondern auch der Effekt der individuellen Merkmale. Das statistische Verfahren der Kontext- bzw. Mehrebenenanalyse basiert auf einer (linearen) Regression. Für das grundlegende Verständnis einer Kontexthypothese bzw. der Kontextanalyse ist ein Verständnis der linearen Regression an dieser Stelle allerdings nicht erforderlich (siehe z. B. Wolf und Best 2010b; Kopp und Lois 2014).

Cross-Level-Effekte Zur Vollständigkeit ist in Abbildung 7 noch ein Pfeil mit der Ziffer 3 eingezeichnet. Der Pfeil postuliert einen Effekt des Kontexts auf den Zusammenhang zwischen den Merkmalen auf der Individualebene. Übertragen auf unser Beispiel: Der Effekt der Bildung auf die Zustimmung zur Europäischen Union könnte auch vom wirtschaftlichen Umfeld abhängig sein. In der Sprache der empirischen Sozialforschung werden solche Beziehungen als Cross-Level-Effekte bezeichnet (eine einfache Einführung in die Logik von Kontext- bzw. Mehrebenenanalysen bietet die frei zugängliche Publikation von Braun et al. 2010).

Zunahme an Kontextanalysen Die Grundidee der Kontexthypothese bzw. Kontextanalyse ist schon relativ alt (siehe z. B. Hummell 1972; einen Überblick klassischer Kontextanalysen bietet Alpheis 1988). In den vergangenen Jahren hat der Anteil an Kontextanalysen allerdings erheblich zugenommen (Dedrick et al. 2009). Zu dieser Entwicklung hat zweifellos der technische Fortschritt beigetragen; im-

mer leistungsfähigere Computer erlauben die Berechnung von immer komplexeren statistischen Modellen. Dem Forscher stehen verschiedene Statistikprogramme bzw. Module zur Verfügung, die es erlauben, die individuelle und kontextuelle Ebene bei der empirischen Analyse zu berücksichtigen. Allerdings sind solche Kontext- bzw. Mehrebenenanalysen auch besonders aufwändig, da Informationen (Daten) der Mikro- und Makroebene erforderlich sind (siehe auch Friedrichs und Nonnenmacher 2014).

Operationalisierung 7

Die Operationalisierung verbindet die theoretische und empirische Ebene in einem Forschungsprojekt. In diesem Kapitel wird dargestellt, wie theoretischen Konzepten beobachtbare Sachverhalte (Indikatoren) zugeordnet werden, um die theoretisch postulierten Zusammenhänge zwischen Konzepten empirisch überprüfen zu können. Anschließend wird erläutert, was Messen in der empirischen Sozialforschung bedeutet, und es werden traditionelle Skalenniveaus dargestellt. Das Kapitel schließt mit einer Vorstellung klassischer Gütekriterien einer Operationalisierung und erläutert die Vorteile der Indexbildung.

7.1 Konzepte und Indikatoren

Die in den Sozialwissenschaften verwendeten theoretischen Konzepte wie Vertrauen, sozialer Status oder Identität sind in der Regel abstrakt und daher nicht direkt beobachtbar. In der Phase der Konzeptspezifikation werden diese Begriffe zunächst geklärt, ehe mit den Konzepten gearbeitet werden kann (siehe Kapitel 5). Die im Anschluss an die Konzeptspezifikation – auf Basis theoretischer Überlegungen – formulierten Hypothesen beziehen sich auf den (vermuteten) Zusammenhang zwischen zwei (theoretischen) Konzepten. Konzepte und Hypothesen sind der theoretischen Ebene eines Forschungsprojekts zuzuordnen. Für eine empirische Überprüfung der formulierten Hypothesen müssen die theoretischen Konzepte mit der empirischen Ebene verknüpft werden. Diese Phase in einem Forschungsprojekt wird als Operationalisierung bezeichnet.

In Abbildung 8 ist die Operationalisierung allgemein dargestellt. Auf der theoretischen Ebene wird eine Hypothese mit den Konzepten A und B formuliert. Es wird vermutet, dass A einen Einfluss auf B ausübt. Zur empirischen Überprüfung dieser Hypothese müssen den theoretischen Konzepten beobachtbare Sachverhal-

© Springer Fachmedien Wiesbaden GmbH, ein Teil von Springer Nature 2018

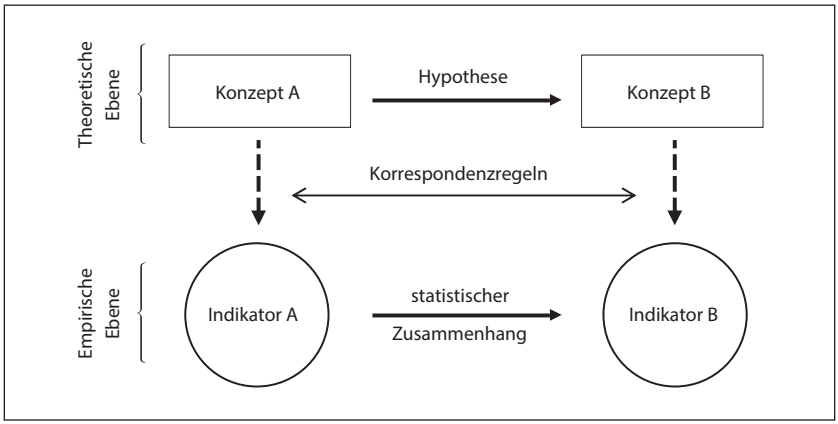
M. Tausendpfund, *Quantitative Methoden in der Politikwissenschaft*,
Grundwissen Politik, https://doi.org/10.1007/978-3-658-20698-7_7

Operationalisierung

Um die theoretischen Konzepte in einem Forschungsprojekt empirisch messbar zu machen, werden die theoretischen Konzepte in beobachtbare Sachverhalte „übersetzt". Dieser Vorgang wird Operationalisierung genannt. Hierfür werden sogenannte Indikatoren herangezogen. Ein beobachtbarer Sachverhalt, der ein theoretisches Konzept abbildet, wird Indikator genannt. Die Verknüpfung zwischen Konzept und Indikator erfolgt über sogenannte Korrespondenzregeln.

te zugeordnet werden. Ein solcher beobachtbarer Sachverhalt wird Indikator genannt. Die Verknüpfung zwischen theoretischem Konzept und beobachtbarem Indikator erfolgt nicht willkürlich, sondern über sogenannte Korrespondenzregeln. Diese Regeln stellen die Verbindung zwischen theoretischer und empirischer Ebene her; sie „regeln" die Zuordnung von Indikatoren zu Konzepten.

Abbildung 8 Konzepte und Indikatoren



Quelle: Eigene Darstellung

Probleme der Operationalisierung Die zentrale Frage bei einer Operationalisierung besteht darin, ob die Zuordnung eines Indikators zu einem theoretischen Konzept überhaupt gerechtfertigt ist (Schnell et al. 2013, S. 121). Nicht immer exis-

tieren unstrittige Indikatoren für ein theoretisches Konzept. Falsche Korrespondenzregeln, also die falsche Zuordnung eines Indikators zu einem theoretischen Konzept, haben schwerwiegende Konsequenzen (siehe Beispiel: Büroklammer-Entwirr-Test). Zum einen kann eine falsche Hypothese aufgrund der empirischen Daten bestätigt und zum anderen eine eigentlich korrekte Hypothese fälschlicherweise empirisch widerlegt werden. Deshalb ist die Suche nach geeigneten Indikatoren für theoretische Konzepte ein ausgesprochen wichtiger Schritt im Forschungsprozess. Die Auswahl der Indikatoren ist daher stets zu begründen und ihre Eignung zu überprüfen.

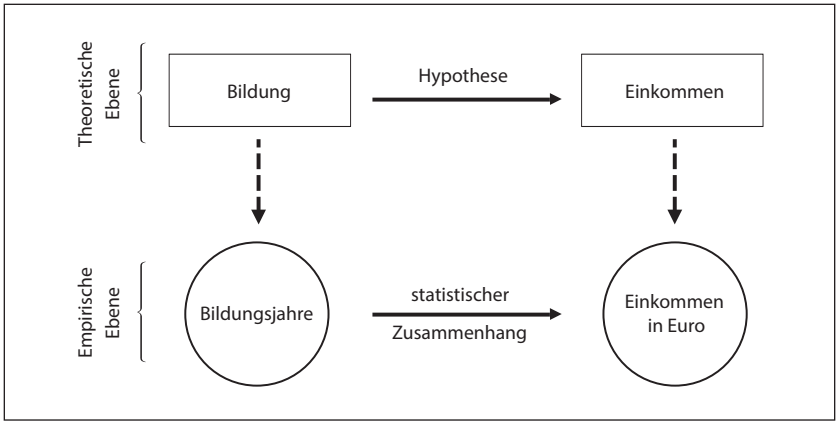
Büroklammer-Entwirr-Test

Bei Diekmann (2011, S. 257) findet sich ein schönes Beispiel für eine offensichtlich nicht optimale Zuordnung eines Indikators zu einem theoretischen Konzept: Die Operationalisierung der Intelligenz auf Basis des Büroklammer-Entwirr-Tests. Der Test funktioniert recht einfach. Die Personen sollen 100 miteinander verknüpfte Büroklammern innerhalb einer bestimmten Zeitspanne entwirren. Die Zahl der entwirrten Klammern ist dann der Indikator für die Intelligenz. Je mehr Klammern eine Person entwirrt hat, desto höher ist die Intelligenz. Es ist leicht nachvollziehbar, dass hier durchaus Zweifel angebracht sind, ob die Zuordnung des empirisch beobachtbaren Sachverhalts („Zahl der entwirrten Klammern“) zu dem theoretischen Konzept („Intelligenz“) gerechtfertigt ist. Leider ist eine nicht optimale Zuordnung eines Indikators zu einem Konzept nicht immer so einfach zu erkennen.

! Für viele theoretische Konzepte existieren verschiedene – auch widersprüchliche – Vorschläge zur Operationalisierung. Deshalb setzt die Operationalisierung eines theoretischen Konzepts wieder eine intensive Auseinandersetzung mit der Fachliteratur voraus. Welche Vorschläge existieren zur Operationalisierung für das Konzept? Was sind Vor- und Nachteile dieser Vorschläge? Unterschiede zwischen verschiedenen Varianten müssen im Hinblick auf die Forschungsfrage diskutiert und gegeneinander abgewogen werden.

In Abbildung 9 wird die Operationalisierung an einem Beispiel illustriert. Auf theoretischer Ebene wird folgende Individualhypothese formuliert: Je höher die Bildung, desto größer das Einkommen. Diese Hypothese basiert auf der Human-

Abbildung 9 Bildung und Einkommen



Quelle: Eigene Darstellung

kapitaltheorie und ist theoretisch plausibel, da Bildung eine zentrale individuelle Ressource darstellt. Höhere Bildung bedeutet demnach „höhere individuelle Fähigkeiten der Informationsverarbeitung, im Umgang mit Problemstellungen bzw. stärkere Handlungskompetenzen" (Hadjar und Becker 2006, S. 14) und damit „verbesserte Teilhabe- und Gestaltungsmöglichkeiten in vielen Bereichen des privaten und öffentlichen Lebens" (Baumert 1991, S. 347; siehe auch Geißler 2014, S. 345). Mit einer höheren (formalen) Bildung verbessern sich die Chancen auf dem Arbeitsmarkt und damit steigt die Wahrscheinlichkeit für ein höheres Einkommen. Die theoretischen Konzepte in unserem Beispiel sind damit „Bildung" und „Einkommen".[13]

Für die empirische Überprüfung der formulierten Hypothese müssen den theoretischen Konzepten nun Indikatoren zugeordnet werden. Als Indikator für das Konzept „Einkommen" könnte beispielsweise das monatliche Nettoeinkommen verwendet werden. Je nach Frageformulierung werden bei dieser Operationalisierung allerdings andere Einkommensarten (z. B. Mieteinnahmen,

13 Was ist Bildung? Bereits die eher populärwissenschaftlichen Bücher von Schwanitz (2002) und Fischer (2003) deuten auf unterschiedliche Antworten hin. Die Veröffentlichung von Precht (2015, S. 25) greift diese Debatte wieder auf. Vor der Operationalisierung muss daher das theoretische Konzept „Bildung" geklärt werden (siehe Kapitel 5). An dieser Stelle wird ein verkürzter Bildungsbegriff verwendet und Bildung mit dem formalen Schulabschluss gleichgesetzt (siehe z. B. Greiffenhagen 2002; Hoffmeyer-Zlotnik und Warner 2014, S. 734–745).

Transferzahlungen) nicht berücksichtigt. Alternativ könnte auch das gesamte Nettohaushaltseinkommen der Haushaltseinheit, in der eine befragte Person eingebunden ist, erfasst werden (Hoffmeyer-Zlotnik und Warner 2014, S. 737; Statistisches Bundesamt 2016).

Auch für die Operationalisierung des Konzepts „Bildung" finden sich in der Literatur mehrere Vorschläge. Naheliegend ist sicherlich der höchste allgemeinbildende Schulabschluss. Alternativ könnte aber auch die Anzahl der Bildungsjahre als Indikator verwendet werden. Schließlich könnte auch ein Test entwickelt werden, um die Bildung jeder Person zu erfassen. Die Beispiele zeigen: Es existieren verschiedene – durchaus vergleichbare, aber nicht identische – Indikatoren, um das theoretische Konzept „Bildung" in einen beobachtbaren Sachverhalt zu übersetzen. Die Wahl eines Indikators ist deshalb stets zu begründen und offenzulegen. In empirischen Studien geschieht dies in der Regel im Abschnitt „Operationalisierung".

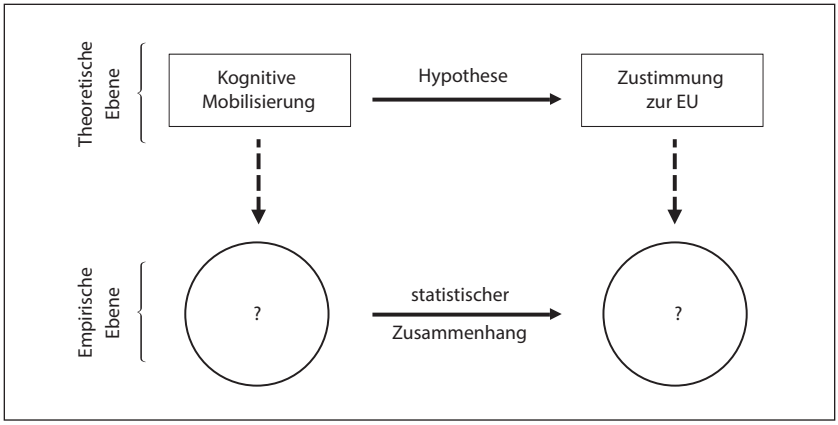
Indikatoren begründen In unserem Beispiel (siehe Abbildung 9) werden die Bildungsjahre als Indikator für das Konzept „Bildung" ausgewählt. Insbesondere bei international vergleichenden Studien (z. B. Eurobarometer) werden Bildungsjahre verwendet, um die verschiedenen Bildungsabschlüsse vergleichen zu können.[14] Um die Hypothese „Je höher die Bildung, desto höher das Einkommen" zu bestätigen, müsste auf der empirischen Ebene ein positiver Zusammenhang zwischen der Anzahl der Bildungsjahre und der Höhe des Einkommens in Euro nachweisbar sein. Dieser Zusammenhang lässt sich mit Hilfe statistischer Verfahren berechnen.

Das Beispiel illustriert, wie anspruchsvoll die Verknüpfung zwischen theoretischem Konzept und empirischem Sachverhalt ist. Die angemessene Operationalisierung von sozialwissenschaftlichen Konzepten ist eine Herausforderung. Das zweite Beispiel (siehe Abbildung 10), bei dem ein Zusammenhang zwischen kognitiver Mobilisierung und Zustimmung zur Europäischen Union postuliert wird, soll dies verdeutlichen. Die Hypothese lautet: „Je höher die kognitive Mobilisierung, desto größer die Zustimmung zur EU."

Bereits das Konzept „Kognitive Mobilisierung" ist schwieriger fassbar als das Konzept „Bildung". Inglehart (1995, S. 419) bezeichnet kognitive Mobilisierung als „Fähigkeiten, die notwendig sind, um sich in einer ausgedehnten politischen Gemeinde zurechtzufinden". Etwas allgemeiner formuliert bedeutet kognitive Mobilisierung eine Erweiterung der Fähigkeiten, mit Politik umzugehen. Bürger mit

14 In der vergleichenden Sozialforschung hat sich mittlerweile die International Standard Classification of Education (ISCED) der UNESCO zur Erfassung des Bildungsniveaus durchgesetzt (z. B. Schneider 2008).

Abbildung 10 Kognitive Mobilisierung und Zustimmung zur EU



Quelle: Eigene Darstellung

einer stärkeren kognitiven Mobilisierung sind mit höheren kognitiven Kompetenzen zur Wahrnehmung, Verarbeitung und Reflexion von Informationen aus der politischen Sphäre ausgestattet (Hadjar und Becker 2006, S. 14) und verfügen daher über einen „einfacheren Zugang zur Politik" (Krimmel 2000, S. 628).

Zusammenhang begründen Mit Blick auf die EU argumentiert Inglehart, dass Personen mit einer höheren kognitiven Mobilisierung eine größere Chance haben „to receive and absorb messages relating to a remote political community" (Inglehart 1970, S. 48). „Cognitive Mobilization should encourage support for European institutions insofar as it helps make them more familiar, less threatening" (Inglehart 1977, S. 338). Danach reduziert kognitive Mobilisierung das Unbehagen vor weit entfernten und komplexen Gebilden wie zum Beispiel dem politischen System der Europäischen Union und verringert Bedrohungsgefühle. Deshalb erwartet Inglehart eine positive Beziehung zwischen dem individuellen Niveau der kognitiven Mobilisierung und der Zustimmung zur EU. Die knappe Darstellung macht deutlich, dass bereits die Konzeptspezifikation und die Begründung der Hypothesen eine Herausforderung ist.

Keine einheitliche Vorgehensweise Im nächsten Schritt muss dem theoretischen Konzept „Kognitive Mobilisierung" ein beobachtbarer Sachverhalt zugeordnet werden. Klingt einfach, ein Blick in die Forschungsliteratur zeigt allerdings, dass es keine einheitliche Vorgehensweise zur Operationalisierung des

Konzepts gibt. Nach Inglehart ist Schulbildung zwar keineswegs dasselbe wie kognitive Mobilisierung, aber „wahrscheinlich doch der beste und am besten zugängliche Indikator" (Inglehart 1995, S. 419). In früheren Veröffentlichungen beurteilte Inglehart (1977, S. 340) die Verwendung der Schulbildung als Indikator für kognitive Mobilisierung allerdings zurückhaltend, da zwischen Bildung und Einkommen bzw. Berufsstatus enge Beziehungen bestehen. Als Indikator für kognitive Mobilisierung nutzte Inglehart (1977, S. 339–340; Inglehart und Rabier 1978, S. 88) auch die Häufigkeit, politische Diskussionen zu führen, sowie die Intensität, Freunde von der eigenen politischen Meinung zu überzeugen. Teilweise ergänzte er beide Indikatoren mit Wissensfragen, die den kognitiven Aspekt des Konzepts abdecken sollten. Der gleiche Autor verwendet zur Operationalisierung des Konzepts verschiedene Indikatoren; in der Fachliteratur existieren noch weitere Vorschläge zur Operationalisierung dieses Konzepts (z. B. Dalton 1984, S. 267; Karp et al. 2003, S. 278). Die Ausführungen illustrieren, wie schwierig es sein kann, geeignete – und weitgehend etablierte – Indikatoren für ein abstraktes Konzept zu finden. Dabei handelt es sich bei kognitiver Mobilisierung nicht einmal um ein besonders komplexes Konzept.

In unserem Beispiel müsste – nach der Operationalisierung von kognitiver Mobilisierung – noch das Konzept „Zustimmung zur EU" spezifiziert und mit einem oder mehreren beobachtbaren Sachverhalten verknüpft werden.[15] Mittlerweile liegt zwar eine große Anzahl an empirischen Analysen zur Zustimmung der Bürgerinnen und Bürger zur Europäischen Union vor, aber es fehlt – bis heute – ein mehr oder weniger akzeptiertes Verständnis dafür, was mit „Zustimmung zur EU" überhaupt gemeint ist. Heißt Zustimmung zur EU, dass die EU eine gute Sache ist? Oder, dass man der EU vertraut? Oder, dass man sich einen schnelleren Integrationsprozess wünscht? Zur Erfassung der „Zustimmung zur EU" gibt es mittlerweile eine erhebliche Bandbreite an Vorschlägen zur Operationalisierung. Dies erschwert natürlich den Vergleich. Die jüngere Forschung (z. B. Weßels 2009; Boomgaarden et al. 2011; Tausendpfund 2013b) greift bei der Operationalisierung der Zustimmung zur Europäischen Union mittlerweile auf das etablierte Konzept der politischen Unterstützung von David Easton (1975) zurück und unterscheidet verschiedene Dimensionen der politischen Unterstützung. Für die Operationalisierung müssten diese verschiedenen Dimensionen offengelegt, spezifiziert und dann mit Indikatoren verknüpft werden. Eine Möglichkeit der Operationalisierung der „Zustimmung zur EU" auf Grundlage des Konzepts der politischen Unterstützung findet sich bei Tausendpfund (2013b, S. 54–83).

15 Die Bezeichnung „Zustimmung zur EU" ist neutral zu verstehen und kann sowohl positive als auch negative Bewertungen umfassen. Personen können der Europäischen Union negativ, neutral oder positiv gegenüber stehen.

Die angemessene Operationalisierung eines theoretisch komplexen Konzepts ist mit Aufwand verbunden. Dabei sind zwei Schritte unbedingt erforderlich. Erstens setzt die Operationalisierung eines Konzepts eine abgeschlossene Konzeptspezifikation voraus. Schnell et al. (2013, S. 116) betonen:

> *„Um eine Operationalisierung (und damit schließlich eine Messung) durchführen zu können, muss eindeutig geklärt sein, auf welche Objekte oder Sachverhalte sich ein Begriff bezieht: Ohne zu wissen, was gemessen werden soll, kann keine sinnvolle Messung erfolgen."*

Zweitens setzt die Operationalisierung eines theoretischen Konzepts die Kenntnis einschlägiger Operationalisierungsvorschläge voraus. Auf dieser Basis können etablierte Vorschläge übernommen oder modifiziert werden. Dabei sind unterschiedliche – zum Teil auch widersprüchliche – Operationalisierungen des gleichen Konzepts in der Literatur eher die Regel als die Ausnahme. Deshalb ist die Wahl eines Indikators immer sorgfältig zu begründen.

Verschiedene Operationalisierungen, unterschiedliche Befunde

Beim Lesen empirischer Studien sollten Sie immer die jeweiligen Operationalisierungen der theoretischen Konzepte im Blick haben. Unterschiedliche, auch sich widersprechende empirische Befunde sind häufig auf verschiedene Operationalisierungen zurückzuführen. Prüfen Sie, ob die Operationalisierung transparent dargelegt wird. Werden Argumente für die Vorgehensweise präsentiert? Werden unterschiedliche Vorschläge gegeneinander abgewogen? Unter anderem dadurch können Studien miteinander verglichen und bewertet werden.

Für viele theoretische Konzepte haben sich in den Sozialwissenschaften aber auch bestimmte Operationalisierungen durchgesetzt. Empfehlungen zur Erfassung sozialstruktureller Merkmale bietet der Band „Demographische Standards" des Statistischen Bundesamts (2016). GESIS – Leibniz-Institut für Sozialwissenschaften bietet mit der Zusammenstellung sozialwissenschaftlicher Items und Skalen (ZIS) eine Übersicht von über 200 sozialwissenschaftlichen Erhebungsinstrumenten, zum Beispiel Indikatoren zur Erfassung politischer Einstellungen oder Persönlichkeit. Rössler (2011) dokumentiert Operationalisierungsvorschläge aus der Kommunikationswissenschaft, und bei Kunter et al. (2002) findet sich eine Übersicht der Erhebungsinstrumente der PISA-Studie.

7.2 Messen in der empirischen Sozialforschung

Empirische Sozialforschung dient der Beschreibung und Erklärung sozialer Sachverhalte. Beide Ziele können – bei einer quantitativen Vorgehensweise – nur durch Messungen erreicht werden. Stevens (1946, S. 677) definiert Messung als „assignment of numerals to objects or events according to rules".

Was heißt messen? Nach der klassischen Definition von Stevens (1946) ist Messung die Zuordnung von Zahlen zu Objekten nach bestimmten Regeln. Zum besseren Verständnis: Gemessen werden nicht die Objekte, sondern Merkmale bzw. Eigenschaften von Objekten. Ein solches „Objekt" könnte der Mensch sein. Ein Merkmal könnte die Körpergröße, die Haarfarbe oder auch das Einkommen sein.

- **Definition: Messen**
  Kromrey et al. (2016, S. 203) definieren Messen als „die Zuweisung von Ziffern zu Objekten entsprechend den Ausprägungen der an diesen Objekten betrachteten Merkmale".

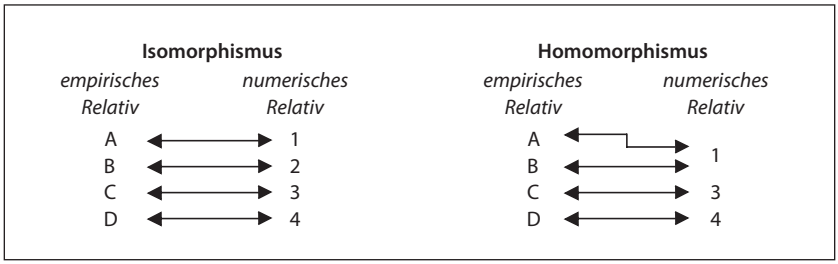
Das Ergebnis einer Messung sind Zahlen. Die Merkmale „Körpergröße" oder „Haarfarbe" werden durch Zahlen abgebildet. Die Zahlen werden den Eigenschaften der Objekte allerdings nicht willkürlich zugeordnet, sondern nach bestimmten Regeln. Günstig wäre eine Zuordnung, die die Beziehung der Eigenschaften der Objekte widerspiegelt. Dies wäre eine strukturtreue Abbildung. Strukturtreu heißt, dass die Beziehung der Merkmale durch die Zahlen widergespiegelt wird.[16]

Morphismus Ziel einer Messung ist die strukturtreue Abbildung eines empirischen Relativs zu einem numerischen Relativ. Eine strukturtreue Abbildung wird Morphismus genannt. In Abbildung 11 werden zwei Arten von Morphismen unterschieden: Isomorphismus und Homomorphismus. Beim Isomorphismus besteht eine wechselseitige eindeutige Zuordnung zwischen empirischem und numerischem Relativ, das heißt jedes Merkmal erhält eine eigene Ziffer. Beim Homomorphismus werden dagegen mehrere Merkmale mit derselben Zahl verknüpft. Dadurch ist kein eindeutiger Rückschluss vom numerischen Relativ auf das empirische Relativ möglich.

Die forschungspraktische Bedeutung von Messen und Morphismen lässt sich leicht illustrieren. Im Eingangsbeispiel untersuchen wir den Zusammenhang zwi-

16 Die Zuordnungs- bzw. Abbildungsregeln werden in der sogenannten axiomatischen Messtheorie entwickelt. Für eine Einführung mit weiteren Literaturhinweisen siehe Diekmann (2011, S. 278–284).

Abbildung 11 Morphismus



Quelle: Schnell et al. (2013, S. 129)

schen Bildung und Einkommen. Ein möglicher Indikator für Bildung ist der höchste allgemeinbildende Schulabschluss. Bei einer Erhebung werden die Personen nach ihrem höchsten Schulabschluss befragt. Dabei wird – vereinfacht – zwischen folgenden Abschlüssen unterschieden: Kein Schulabschluss, Hauptschulabschluss, Realschulabschluss, Fachhochschulreife und allgemeine Hochschulreife.[17]

Die verschiedenen Schulabschlüsse sind unser empirisches Relativ. Die einzelnen Merkmale (Schulabschlüsse) stehen in einer Beziehung. Ein Realschulabschluss ist ein „höherer" Abschluss als kein Schulabschluss; die allgemeine Hochschulreife ist ein „höherer" Abschluss als ein Realschulabschluss. Bei der Messung werden den einzelnen Schulabschlüssen Zahlen zugewiesen. Ziel der Messung ist es, die Struktur des empirischen Relativs zu erhalten, das heißt, die zugeordneten Zahlen sollen die Rangfolge des Schulabschlusses widerspiegeln. Bei einer willkürlichen Zuordnung der Zahlen zu den einzelnen Schulabschlüssen – z. B. dem Hauptschulabschluss eine 3, der allgemeinen Hochschulreife eine 2 und der Fachhochschulreife eine 1 – würde das numerische Relativ nicht die Struktur des empirischen Relativs widerspiegeln. Folglich wäre eine solche willkürliche Zuordnung von Zahlen zu Objekten keine (strukturtreue) Abbildung.

Eine Messung als strukturtreue Abbildung fordert die Zuordnung von Zahlen zu Objekten nach bestimmten Regeln.[18] In Abbildung 12 finden sich zwei Beispiele für strukturtreue Abbildungen (Morphismus). Beim Isomorphismus wird

17 Bei einer echten Erhebung wird der höchste allgemeinbildende Schulabschluss in der Regel differenzierter erfasst (z. B. Statistisches Bundesamt 2016, S. 11)

18 An dieser Stelle ist offensichtlich, dass die klassische Messdefinition von Stevens (1946) nicht optimal ist. Gemäß der oben genannten Definition ist Messung die Zuordnung von Zahlen zu Objekten gemäß festgelegter Regeln. Diese Regeln sind aber zu spezifizieren. Denk-

Abbildung 12 Morphismus am Beispiel formaler Bildungsabschlüsse

Isomorphismus			Homomorphismus		
empirisches Relativ		*numerisches Relativ*	*empirisches Relativ*		*numerisches Relativ*
kein Abschluss	↔	1	kein Abschluss	⇨	1
Hauptschulabschluss	↔	2	Hauptschulabschluss		
Realschulabschluss	↔	3	Realschulabschluss	↔	2
Fachhochschulreife	↔	4	Fachhochschulreife	↔	3
Hochschulreife	↔	5	Hochschulreife	↔	4

Quelle: Eigene Darstellung

jedem Abschluss eine Zahl zugeordnet. Je höher der Abschluss, desto höher die Zahl. Die Zuordnung beim Isomorphismus ist wechselseitig eindeutig, da jedem Abschluss nur eine Zahl zugeordnet wird. Beim Homomorphismus werden dagegen mehrere Abschlüsse mit derselben Zahl verknüpft. Dadurch ist kein Rückschluss vom numerischen zum empirischen Relativ möglich.

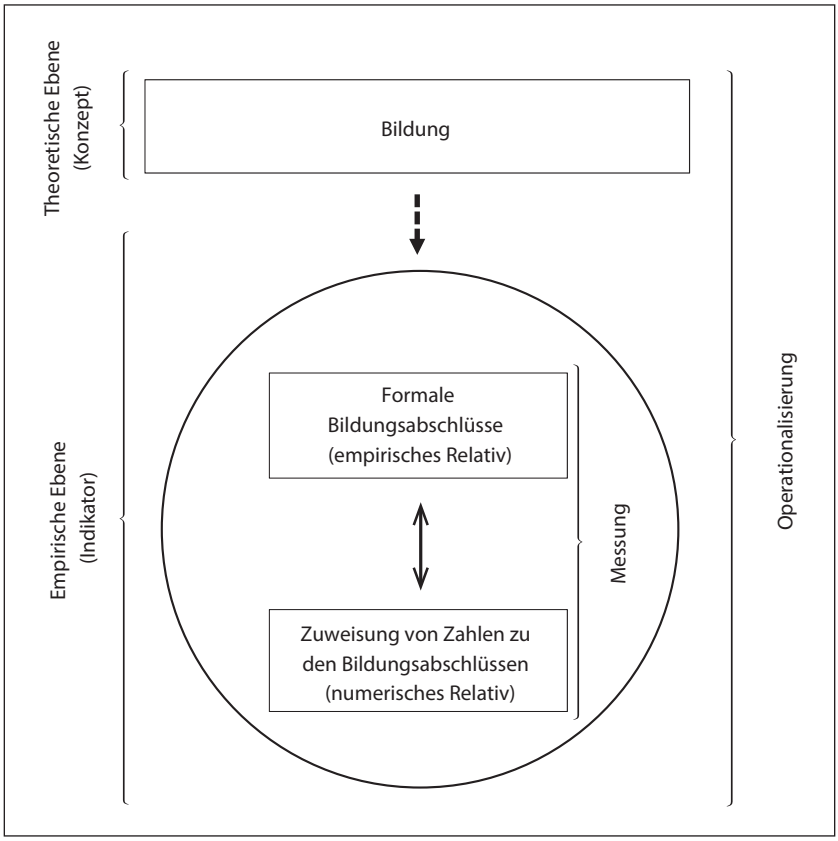
Beim Messen wird ein empirisches Relativ in ein numerisches Relativ „umgewandelt". Dabei soll die Beziehung der Merkmale der Objekte jeweils erhalten bleiben. Eine umkehrbare eindeutige Abbildung wird als Isomorphismus bezeichnet. Sind einer Zahl mehrere Merkmale zugeordnet, dann handelt es sich um eine nicht umkehrbare eindeutige Abbildung (Homomorphismus).

In Abbildung 13 wird die Beziehung zwischen Konzept, Indikator und Messen verdeutlicht. Die Operationalisierung verbindet die theoretische und die empirische Ebene in einem Forschungsprozess. Theoretischen Konzepten werden beobachtbare Sachverhalte (Indikatoren) zugeordnet. In unserem Beispiel wird das formale Bildungsniveau als Indikator für das theoretische Konzept „Bildung" verwendet. Es lassen sich verschiedene Bildungsabschlüsse unterscheiden (z. B. Hauptschulabschluss, Fachhochschulreife, Abitur). Die verschiedenen Bildungsabschlüsse bilden zusammen das empirische Relativ.

Bei der Messung wird dem Merkmal (z. B. Bildungsabschluss) eines Untersuchungsobjekts (z. B. Person) eine Zahl zugeordnet. Die Zuordnung der Zahlen erfolgt nicht beliebig, sondern nach bestimmten Regeln. Denn Messungen

bar wäre eine „Messung" der Schulabschlüsse in Abhängigkeit von der Anzahl der Buchstaben. Schulabschlüsse mit mehr Buchstaben erhalten eine höhere Ziffer als Schulabschlüsse mit weniger Buchstaben. Dies ist mit Blick auf die Forschungsfrage aber keine sinnvolle Relation. In der empirischen Praxis bereiten die Zuordnungsregeln allerdings selten Probleme.

Abbildung 13 Konzept, Indikator und Messung



Quelle: Eigene Darstellung

sollen strukturtreue Abbildungen widerspiegeln. Das Strukturmerkmal im gewählten Beispiel ist das Bildungsniveau. Die einzelnen Bildungsabschlüsse können hinsichtlich des Bildungsniveaus geordnet werden. Einem höheren Bildungsabschluss wird folglich eine höhere Zahl zugeordnet.[19]

19 Grundsätzlich können höheren Bildungsabschlüssen auch niedrigere Zahlen zugeordnet werden. Also: 1 = allgemeine Hochschulreife, 2 = Fachhochschulreife, 3 = Realschulabschluss, 4 = Hauptschulabschluss und 5 = kein Abschluss. Diese Zuordnung wäre auch strukturtreu. Die Festlegung, ob höhere Zahlen einen höheren Abschluss oder höhere Zahlen einen nied-

7.3 Skalen

Eine Skala ist eine strukturtreue Abbildung eines empirischen Relativs in ein numerisches Relativ (z. B. Westle 2009e, S. 183; Schnell et al. 2013, S. 130). In den Sozialwissenschaften werden meist die Skalenarten von Stevens (1946) verwendet (z. B. Diekmann 2011, S. 285; Schnell et al. 2013, S. 132; Porst 2014a, S. 71), der vier Skalenniveaus unterscheidet: Nominal-, Ordinal-, Intervall- und Ratioskala. Intervall- und Ratioskalen werden auch metrische Skalen genannt.

Warum ist die Unterscheidung der Skalen wichtig?

In der Regel ist der Informationsgehalt einer Messung umso höher, je höher die Skala bzw. das Skalenniveau ist. Das Messniveau einer Ordinalskala ist höher als das Messniveau einer Nominalskala und das Messniveau einer Ratioskala ist höher als das Messniveau einer Intervallskala (allgemein: Ratioskala > Intervallskala > Ordinalskala > Nominalskala). Je höher die Skala, desto mehr (statistische) Auswertungsverfahren sind möglich. Die Kenntnis der Skalenniveaus ist wichtig, um bei der Datenanalyse nur die zulässigen Auswertungsverfahren anzuwenden. Viele statistische Verfahren sind nur zulässig, wenn eine Messung mit mindestens Intervallskalenniveau vorliegt. Deshalb wird meist ein möglichst hohes Messniveau angestrebt (siehe auch Schendera 2015, S. 39–66).

Nominalskala Die Nominalskala ist das unterste Messniveau. Stevens (1946, S. 678) beschreibt die Nominalskala wie folgt:

> *„The nominal scale represents the most unrestricted assignment of numerals. The numerals are used only as labels or type numbers, and words or letters would serve as well.“*

Bei einer Nominalskala wird jedes Merkmal eines Objekts genau einer Klasse zugeordnet. Schnell et al. (2013, S. 133) betonen: „Weder darf ein bestimmtes Objekt überhaupt nicht zugeordnet werden, noch darf ein Objekt mehreren Klassen zugeordnet werden.“ Eine Nominalskala erlaubt damit lediglich eine Unterscheidung der Objekte.

---

rigeren Abschluss widerspiegeln, liegt in der Verantwortung des Forschers. Für die Interpretation der empirischen Ergebnisse muss die Zuordnung jedoch berücksichtigt werden.

Ein bekanntes Beispiel für ein nominalskaliertes Merkmal ist das Geschlecht, das aus zwei Kategorien besteht: weiblich oder männlich.[20] Jedes Objekt (Mensch) kann hinsichtlich des Merkmals „Geschlecht" genau einer dieser zwei Kategorien zugeordnet werden. Bei der Messung werden den Kategorien „weiblich" und „männlich" dann Zahlen zugeordnet: 1 für weiblich, 2 für männlich. Aber diese Zuordnung der Ziffern ist reine Definitionssache. Es könnte auch 1 für männlich und 2 für weiblich verwendet werden.

Ein anderes Beispiel für ein nominalskaliertes Merkmal ist die Konfession oder die Zugehörigkeit zu einer Glaubensgemeinschaft. Die Personen könnten etwa in den Klassen „Evangelische Kirche (ohne Freikirchen)", „Evangelische Freikirche", „Römisch-katholische Kirche", „einer anderen christlichen Religionsgemeinschaft", „einer anderen nicht-christlichen Religionsgemeinschaft" oder „keiner Religionsgemeinschaft" zugeordnet werden. Diesen Klassen werden bei der Messung wieder Ziffern zugeordnet, etwa 1 für „Evangelische Kirche (ohne Freikirchen)", 2 für „Evangelische Freikirche" usw. Die Zuordnung einer Ziffer zu einer Klasse ist wieder reine Definitionssache.

Bei einer Nominalskala werden den einzelnen Klassen unterschiedliche Ziffern zugeordnet, die allerdings nur eine Kennzeichnung darstellen. Die konkrete Bezifferung ist aber eine reine Definition, die nicht falsch oder richtig, sondern allenfalls mehr oder weniger sinnhaft ist (Porst 2014a, S. 71–72). In der empirischen Praxis wird häufig 1 für weiblich und 2 für männlich verwendet. Die Möglichkeiten von Auswertungsverfahren bei nominalskalierten Merkmalen sind begrenzt: Die Objekte können nur ausgezählt werden (z. B. Anzahl der Frauen und Männer).

„The ordinal scale arises from the operation of rank-ordering" (Stevens 1946, S. 679).

Ordinalskala Während bei einer Nominalskala die Objekte bzw. die Merkmale der Objekte lediglich in Bezug auf ihre Unterscheidbarkeit untersucht werden (z. B. männlich oder weiblich), können bei einer Ordinalskala die Merkmale der Objekte zusätzlich in eine Rangordnung gebracht werden. Die Ausprägungen

20 In einer Stellungnahme zum Thema Intersexualität schreibt der Deutsche Ethikrat (2012, S. 177) mit Blick auf Personenstandsregister: „Es sollte geregelt werden, dass bei Personen, deren Geschlecht nicht eindeutig feststellbar ist, neben der Eintragung als ‚weiblich' oder ‚männlich' auch ‚anderes' gewählt werden kann." In den demographischen Standards des Statistischen Bundesamts (2016, S. 8–9), die darauf abzielen, soziostrukturelle Erhebungsmerkmale in Bevölkerungsumfragen zu vereinheitlichen, wird auf Intersexualität oder Transsexualität nicht eingegangen. Beim Geschlecht der Zielperson wird zwischen „männlich" und „weiblich" differenziert. In seinem Beschluss vom 10. Oktober 2017 (Aktenzeichen: 1 BvR 2019/16) hat das Bundesverfassungsgericht entschieden, dass das Personenstandsrecht einen weiteren positiven Geschlechtseintrag zulassen muss. Eine Anpassung der demographischen Standards ist abzusehen.

stehen in einer relationalen Beziehung und die zugeordneten Ziffern geben diese Rangordnung wieder. Ein klassisches Beispiel sind Schulnoten. Dabei werden den Noten Ziffern zugeordnet: mangelhaft = 5, ausreichend = 4, befriedigend = 3, gut = 2 und sehr gut = 1. Die Schulnoten geben eine ordinale Rangordnung vor. Eine 2 ist besser als eine 4. Allerdings ist der Abstand zwischen den einzelnen Bewertungen nicht gleich groß. Mit anderen Worten: Eine 2 ist zwar besser als eine 4, aber eine 2 ist nicht doppelt so gut wie eine 4. Ein Blick auf den IHK-Notenschlüssel verdeutlicht dies: Auszubildende müssen 57 von maximal 100 Punkten erreichen, um eine 4,0 zu erhalten. Für eine 3,0 sind 73 Punkte erforderlich, für eine 2,0 dann 85 Punkte. Die Punktabstände zwischen den Noten sind nicht identisch.[21]

In der empirischen Sozialforschung werden viele Merkmale von Objekten mit Ordinalskalen erfasst. In der Allgemeinen Bevölkerungsumfrage der Sozialwissenschaften (ALLBUS) werden die Personen wie folgt nach ihrem politischen Interesse befragt:

> *Wie stark interessieren Sie sich für Politik? Sehr stark, stark, mittel, wenig oder überhaupt nicht?*

Die einzelnen Antwortvorgaben können in eine Rangfolge gebracht werden: „sehr stark" deutet auf ein stärkeres Politikinteresse als „mittel" hin und „mittel" ist sicherlich ein stäkeres Politikinteresse als „überhaupt nicht". Die Antwortvorgaben erlauben damit die Festlegung einer Reihenfolge, aber die Abstände zwischen den einzelnen Antwortvorgaben sind nicht klar. Den einzelnen Antwortvorgaben werden bei der Messung wieder Ziffern zugeordnet. Zum Beispiel: sehr stark = 1, stark = 2, mittel = 3, wenig = 4 und überhaupt nicht = 5. Bei einer Ordinalskala stehen bereits mehr (statistische) Auswertungsverfahren zur Verfügung als bei einer Nominalskala. Es kann der Median bestimmt werden, der eine Verteilung in zwei Hälften teilt.

Intervallskala Eine Intervallskala zeichnet sich dadurch aus, dass die einzelnen Ausprägungen auf einer Skala nicht nur in eine Reihenfolge gebracht werden können, sondern, dass die Abstände zwischen den Ausprägungen auch gleich sind.

> *„Die Intervalle müssen die gleiche Größe besitzen. Die entsprechenden Zahlen müssen daher stets dieselbe Differenz besitzen." (Schnell et al. 2013, S. 133)*

---

21 Ein vergleichbares Beispiel sind Bildungsabschlüsse (siehe Abschnitt 7.2). Die Abstände zwischen den einzelnen Bildungsabschlüssen, denen bei der Messung Zahlen zugewiesen werden, sind nicht identisch.

Ein Beispiel sind Temperaturmessungen in Grad Celsius. Der Abstand zwischen 15 und 20 Grad Celsius ist genau so groß wie der Abstand zwischen 20 und 25 Grad Celsius (jeweils fünf Grad Celsius). Mit anderen Worten: Bei einer Intervallskala sind die Abstände (Differenzen) zwischen den Merkmalsausprägungen gleich. Es kann folglich Äquidistanz der Abstände unterstellt werden. Es gibt allerdings – im Vergleich zu Verhältnis- und Ratioskalen – keinen natürlichen Nullpunkt.

Pseudometrische Skalen

Viele statistische Auswertungsverfahren setzen mindestens intervallskalierte Messungen voraus, allerdings liegen empirische Informationen häufig nur auf ordinalem Skalenniveau vor. Dadurch werden die Analysemöglichkeiten eingeschränkt. Deshalb werden in der empirischen Praxis häufig Skalen ab etwa fünf Ausprägungen als pseudometrische Skalen behandelt (Faulbaum et al. 2009, S. 26; Baur 2011). Für die Behandlung einer Ordinalskala als pseudometrische Skala ist neben der Mindestanzahl von fünf Ausprägungen allerdings entscheidend, dass die Abstände zwischen den Merkmalsausprägungen nicht nur in eine Reihenfolge gebracht werden können, sondern dass die Abstände zwischen den Ausprägungen auch gleich sind (Äquidistanz). Die Behandlung einer ordinalen Skala als pseudometrisch muss deshalb stets begründet werden und sollte nicht leichtfertig erfolgen (für eine kritische Betrachtung siehe Ohr 2010, S. 646).

! Der „Sprung" von der Ordinal- zur Intervallskala ist in den Sozialwissenschaften von großer Bedeutung, da eine Vielzahl von statistischen Auswertungsverfahren intervallskallierte Messungen voraussetzt (Schnell et al. 2013, S. 135). Das Problem: In den Sozialwissenschaften sind solche Skalen eigentlich selten (Porst 2014a, S. 75). Eine Lösung für dieses Problem ist die Entwicklung von Messinstrumenten, die die Voraussetzungen einer Intervallskala erfüllen. Im European Social Survey (ESS) findet sich beispielsweise folgende Frage zur Erfassung des sozialen Vertrauens:

> *Ganz allgemein gesprochen: Glauben Sie, dass man den meisten Menschen vertrauen kann, oder dass man im Umgang mit anderen Menschen nicht vorsichtig genug sein kann? Bitte sagen Sie es mir anhand dieser Skala von 0 bis 10. 0 bedeutet, dass man nicht vorsichtig genug sein kann, und 10 bedeutet, dass man den meisten Menschen vertrauen kann.*

Bei dieser Frage sind die beiden Endpunkte „Man kann nicht vorsichtig genug sein“ (0) und „Den meisten Menschen kann man vertrauen“ (10) mit den Ziffern 0 und 10 verknüpft. Mit der Frageformulierung wird dem Befragten nahegelegt, dass die Abstände zwischen den einzelnen Antwortmöglichkeiten gleich sind. In der empirischen Praxis sind solche endpunktbenannten Skalen sehr verbreitet, aber die wesentliche Forderung einer Intervallskala – die gleichen Abstände zwischen den Skalenpunkten – muss nicht unbedingt erfüllt sein. Im Forschungsalltag haben sich solche Skalen allerdings durchgesetzt (zur Entwicklung von Antwortskalen siehe Rohrmann 1978). Häufig werden 7- oder 11-Punkt-Skalen verwendet, um den Grad der Zufriedenheit oder Zustimmung abzufragen. Für Daten auf Intervallskalenniveau stehen eine Reihe von zusätzlichen statistischen Verfahren zur Verfügung: Mittelwert, Standardabweichung und Korrelationen.

Verhältnis- und Ratioskalen Bei Verhältnis- bzw. Ratioskalen sind schließlich nicht nur die Abstände zwischen den einzelnen Skalenpunkten gleich, sondern es existiert noch ein echter Skalennullpunkt. Zum Vergleich: Bei der obigen Frage zur Erfassung des sozialen Vertrauens gibt es zwar auch einen „Skalennullpunkt“, aber dieser ist künstlich gesetzt. Die Skala könnte auch den Bereich 10 bis 20 abbilden. Eine Verhältnis- oder Ratioskala besitzt aber einen echten Nullpunkt.

Ein typisches Beispiel für eine Verhältnis- oder Ratioskala ist die Messung von Längen mit Hilfe eines Lineals oder Metermaßes. Ein Objekt kann zehn Zentimeter oder auch 1,50 Meter lang sein. Der Abstand zwischen fünf und zehn Zentimetern sowie zwischen 1,50 und 1,55 Metern beträgt jeweils fünf Zentimeter. In den Sozialwissenschaften sind das Nettoeinkommen, die Dauer der Arbeitslosigkeit oder die Anzahl der im vergangenen Jahr gelesenen Bücher Beispiele für Merkmale von Objekten, die mit Verhältnis- bzw. Ratioskalen gemessen werden können.

Absolutskala Eine besondere Variante der Verhältnis- oder Ratioskala ist die Absolutskala. Außer einem natürlichen Nullpunkt gibt es noch eine natürliche Einheit.

> *„Die Anzahl der Personen in einem Raum, die Anzahl der Gasmoleküle in einem Kubikmeter Luft sind demnach Eigenschaften, die mit einer Absolutskala gemessen werden können.“ (Behnke und Behnke 2006, S. 129)*

Tabelle 9 verdeutlicht, dass die genannten Skalenniveaus eine Hierarchie darstellen. Messungen auf Nominalskalenniveau können nur etwas über die Gleichheit bzw. Unterschiedlichkeit der Merkmale der Objekte aussagen. Bei Messungen auf Ordinalskalenniveau können die Merkmale in eine geordnete Reihenfolge ge-

Tabelle 9 Skalenniveaus im Vergleich

Skalenniveau	Nominalskala	Ordinalskala	Intervallskala	Ratioskala
Unterschied	Ja	Ja	Ja	Ja
Ränge	Nein	Ja	Ja	Ja
gleiche Abstände	Nein	Nein	Ja	Ja
Natürlicher Nullpunkt	Nein	Nein	Nein	Ja
Zulässigkeit statistischer Verfahren	Modus	Median	Arithmetischer Mittelwert	Geometrischer Mittelwert
Beispiel	Geschlecht	Schulnoten	Temperatur in Grad Celsius	Einkommen in Euro

bracht werden. Bei einer Messung auf Intervallskalenniveau sind die Abstände zwischen den Ausprägungen bzw. zwischen den Skalenpunkten auch gleich groß. Schließlich besitzt die Verhältnis- bzw. Ratioskala zudem einen natürlichen Nullpunkt.

Die statistischen Eigenschaften eines niedrigen Skalenniveaus sind immer auch Eigenschaften eines höheren Skalenniveaus. Bei einer Ordinalskala können die Messwerte nicht nur in eine Reihenfolge gebracht werden, sondern es können auch Unterschiede klassifiziert werden. Bei einer Intervallskala sind die Abstände der einzelnen Skalenpunkte jeweils gleich groß, die Ränge können in eine Reihenfolge gebracht und die einzelnen Messwerte können mit Blick auf Gleichheit bzw. Unterschiedlichkeit differenziert werden.

Wie bereits oben dargestellt, unterscheiden sich die einzelnen Skalenniveaus auch mit Blick auf die Zulässigkeit statistischer Auswertungsverfahren. Bei einer Nominalskala können die Werte nur ausgezählt werden. Es kann beispielsweise der häufigste Wert (Modus) ermittelt werden. Bei einer Messung auf Ordinalskalenniveau kann der Median bestimmt werden. Bei einer Messung auf Intervallskalenniveau können Differenzen gebildet werden, sodass der Mittelwert, die Standardabweichung oder auch Korrelationen berechnet werden können. Schließlich können bei einer Ratioskala auch sinnvolle Quotienten gebildet werden (Stevens 1946, S. 678).

7.4 Gütekriterien der Operationalisierung

In der empirischen Sozialforschung lassen sich allgemein drei Gütekriterien der Operationalisierung unterscheiden: Objektivität, Reliabilität und Validität (z. B. Rammstedt 2010; Diekmann 2011, S. 247; Krebs und Menold 2014).

7.4.1 Objektivität

Unter Objektivität wird die Unabhängigkeit der Operationalisierung vom jeweiligen Forscher verstanden. Die Objektivität einer Messung ist dann gegeben, wenn verschiedene Personen mit dem gleichen Messinstrument zum gleichen Ergebnis kommen (Häder 2015, S. 104).

> *„Vollständige Objektivität liegt vor, wenn zwei Anwender A und B mit dem gleichen Messinstrument jeweils übereinstimmende Resultate erzielen." (Diekmann 2011, S. 249)*

Die Objektivität eines Messinstruments wird deshalb auch als Anwenderunabhängigkeit bezeichnet. Das Testergebnis ist unabhängig von der Person, die die Messung vornimmt. Mit Döring und Bortz (2016, S. 443) lassen sich drei Aspekte der Objektivität unterscheiden: Durchführungsobjektivität, Auswertungsobjektivität und Interpretationsobjektivität.

Durchführungsobjektivität Die Durchführungsobjektivität bezieht sich auf die Phase der Datenerhebung. Zwei verschiedene Interviewer sollten beispielsweise bei einer Befragung mit dem gleichen Fragebogen bei einer Person zu den gleichen Ergebnissen kommen. Der Interviewer stellt Fragen, die der Befragte beantworten soll. Dieser soll die Frage allerdings unabhängig vom anwesenden Interviewer beantworten, das heißt, der Interviewer soll keinen Einfluss auf die Antworten des Befragten haben. Würde eine Person auf die gleichen Fragen zu seinem Freizeitverhalten bei Interviewer A völlig anders antworten als bei Interviewer B, dann wäre die Objektivität der Messung gering. Denkbar wäre, dass sich ein männlicher Befragter bei einer weiblichen Interviewerin anders präsentieren möchte als bei einem männlichen Interviewer. Durch genaue Handlungsanweisungen wird versucht, den subjektiven Einfluss der untersuchenden Person (hier: der Interviewer) auf das Messergebnis möglichst gering zu halten. Der Grad des möglichen subjektiven Einflusses auf das Messergebnis variiert dabei auch in Abhängigkeit von der Form der Datenerhebung. Bei einem persönlichen Interview ist der mögliche Einfluss des Interviewers auf das Antwortverhalten des Befragten größer als bei einem telefonischen Interview.

Auswertungsobjektivität Die Auswertungsobjektivität bezieht sich auf die Analyse der gesammelten Daten. Die Auswertungsobjektivität ist gegeben, wenn verschiedene Personen mit den gleichen Daten und den gleichen Auswertungsverfahren zu den gleichen Ergebnissen kommen. Ein Beispiel in Anlehnung an Diekmann (2011, S. 249) soll die Auswertungsobjektivität illustrieren. Textinterpretationen von Studierenden in der Germanistik werden von verschiedenen Dozenten vermutlich (bis zu einem gewissen Grad) unterschiedlich bewertet. Der Interpretationsspielraum eines Dozenten verringert die Auswertungsobjektivität. Die Ergebnisse von Statistikklausuren werden auch bei der Korrektur durch verschiedene Dozenten gleich ausfallen, da der Interpretationsspielraum bei der Auswertung geringer ist. Allgemein ist bei den quantitativen Analyseverfahren die Auswertungsobjektivität eher unproblematisch. Bei Anwendung der gleichen statistischen Auswertungsverfahren sollten verschiedene Forscherinnen zu gleichen Ergebnissen kommen.

Interpretationsobjektivität Die Interpretationsobjektivität bezieht sich nach Krebs und Menold (2014, S. 426) auf die Interpretation sozialwissenschaftlicher Forschungsergebnisse. Da die Interpretation empirischer Ergebnisse subjektiven Wertungen (Werturteilen) unterliegen kann, ist die Interpretationsobjektivität häufig eingeschränkt. Ein Beispiel: In den 1990er Jahren wurde die rückläufige Beteiligung bei den Bundestagswahlen höchst unterschiedlich bewertet. Feist (1992) beurteilte die gesunkene Wahlbeteiligung als Krisensymptom der Demokratie, da politischer Protest, Misstrauen in die politische Klasse sowie fehlende politische Einflussmöglichkeiten die zentralen Motive für die Wahlenthaltung seien. Dagegen sah Roth (1992) in der sinkenden Wahlbeteiligung keine Krise der Demokratie. Im Gegenteil: Die Bürger verzichten auf die Stimmabgabe, weil sie mit der Politik und der Funktionsweise der Demokratie zufrieden sind. Entsprechend wurde der Rückgang der Beteiligung als Normalisierung interpretiert. Der gleiche empirische Befund – Rückgang der Wahlbeteiligung – wurde von Wissenschaftlern höchst unterschiedlich bewertet.

7.4.2 Reliabilität

Mit Reliabilität ist die Zuverlässigkeit einer Messung gemeint. Das Gütekriterium bezieht sich damit auf die Replizierbarkeit von Messungen und beschreibt das Maß der Genauigkeit, mit der ein Konzept erfasst werden kann. Wiederholte Messungen eines unveränderten Konzepts mit dem gleichen Messinstrument sollen die gleichen Ergebnisse liefern.

„Die Reliabilität eines Messinstruments ist ein Maß für die Reproduzierbarkeit von Messergebnissen." (Diekmann 2011, S. 250)

Was sind Messfehler? Perfekte Reliabilität bedeutet, den „wahren Wert" eines Konzepts ohne jeden Messfehler zu erfassen (Döring und Bortz 2016, S. 442–445). Ein Beispiel: Von einem Balkon wird zehnmal ein Tennisball geworfen. Ein Forscher notiert sich jeweils die Zeit, die der Tennisball bis zum Aufprall auf die Erde benötigt. Auch bei identischen Würfen wird sich – bedingt durch die menschliche Reaktionszeit – die gemessene Zeit bei den einzelnen Würfen leicht unterscheiden. Diese Abweichung vom „wahren Wert" wird als Messfehler bezeichnet. Ein Messverfahren ist reliabel, wenn diese Messfehler möglichst gering sind. In der empirischen Praxis lassen sich Messfehler allerdings nicht völlig vermeiden.

In der Literatur werden drei Vorgehensweisen unterschieden, um die Reliabilität eines Messinstruments einschätzen zu können: Test-Retest-Methode, Paralleltest-Methode und Methode der Testhalbierung (siehe z. B. Diekmann 2011, S. 250–256; Schnell et al. 2013, S. 141–143; Döring und Bortz 2016, S. 465–469).

Klassische Testtheorie

Das Grundmodell der klassischen Testtheorie basiert auf der Annahme, dass sich ein Messwert (X) aus einem wahren Wert (T = True Score) und dem Messfehler (E = Error Score) zusammensetzt. Formal betrachtet:

$$X = T + E$$

Die Messung der kognitiven Mobilisierung einer Person setzt sich beispielsweise aus der „wahren" kognitiven Mobilisierung (T) und einem Fehlerterm (E) zusammen. Ein möglicher Fehler könnte sein, dass eine Befragte beim Ausfüllen des Fragebogens in der Spalte verrutscht. Eine Annahme der klassischen Testtheorie ist jetzt, dass sich die „Fehler" bei wiederholten Messungen ausgleichen. Bei einer Messung wird kognitive Mobilisierung etwas überschätzt, bei der nächsten Messung wird die kognitive Mobilisierung etwas unterschätzt. Werden jetzt „viele" Messungen durchgeführt, dann gleichen sich die Ungenauigkeiten „irgendwann" aus. Als „wahrer Wert" kann der Mittelwert einer großen Anzahl unabhängiger Messungen desselben Konzepts aufgefasst werden (Schnell et al. 2013, S. 139). Für eine Darstellung der klassischen Testtheorie siehe Döring und Bortz (2016, S. 463–482).

Test-Retest-Methode Bei der Test-Retest-Methode wird das Konzept, das erfasst werden soll, bei den gleichen Personen zu zwei Zeitpunkten gemessen. Danach werden die Ergebnisse verglichen. Das Problem dieser Methode ist, dass eine zeitliche Stabilität der Konzepte unterstellt wird. Ein Beispiel: Die Personen werden zu ihrer Haltung gegenüber Kernenergie befragt. Zwischen den beiden Zeitpunkten der Messung ereignet sich ein Reaktorunglück. Es ist relativ wahrscheinlich, dass sich zumindest bei einem Teil der Befragten die Haltung zur Kernenergie nach dem Reaktorunglück verändert hat. Die Veränderung ist also nicht auf die mangelnde Zuverlässigkeit des Messinstruments zurückzuführen, sondern vielmehr auf eine Veränderung des „wahren" Werts. Die Methode ist daher bei instabilen bzw. zeitabhängigen Merkmalen wenig brauchbar. Ein weiterer Nachteil der Test-Retest-Methode besteht in ihrem relativ großen Aufwand, da die Messung zweimal durchgeführt werden muss.

Paralleltest-Methode Bei der Paralleltest-Methode werden zwei unterschiedliche Messinstrumente verwendet, die jeweils das gleiche Konzept messen und identische Messeigenschaften haben sollen. Diese Vorgehensweise ist in der Praxis ebenfalls nicht sehr verbreitet, da einerseits die Entwicklung ähnlicher Indikatoren schwierig und andererseits die doppelte Datensammlung auch sehr aufwändig ist.

Methode der Testhalbierung Die Methode der Testhalbierung setzt voraus, dass ein Konzept mit mehreren Indikatoren gemessen wird (siehe Abschnitt 7.5). Im European Social Survey wird beispielsweise das soziale Vertrauen mit drei Fragen (Items) erfasst. Diese lauten:

- Ganz allgemein gesprochen: Glauben Sie, dass man den meisten Menschen vertrauen kann, oder dass man im Umgang mit anderen Menschen nicht vorsichtig genug sein kann?
- Glauben Sie, dass die meisten Menschen versuchen, Sie auszunutzen, wenn sie die Gelegenheit dazu haben, oder versuchen die meisten Menschen, sich fair zu verhalten?
- Und glauben Sie, dass die Menschen meistens versuchen, hilfsbereit zu sein, oder dass die Menschen meistens auf den eigenen Vorteil bedacht sind?

Cronbachs Alpha Alle Fragen beantworten die Personen mit Zahlen von 0 bis 10. Höhere Werte entsprechen dabei einem größeren Vertrauen. Die Methode der Testhalbierung basiert nun auf der Idee, dass Personen mit einem größeren sozialen Vertrauen bei allen drei Fragen eher höhere Zahlen angeben und Personen mit einem geringeren sozialen Vertrauen bei allen Fragen eher niedrigere Werte nennen. Auf dieser Basis lassen sich dann sogenannte Reliabilitätskoeffizienten be-

rechnen. Am bekanntesten ist Cronbachs Alpha, der Werte zwischen 0 und 1 annehmen kann. Höhere Werte lassen auf eine höhere Reliabilität schließen. Nach Schnell et al. (2013, S. 143) können Werte über 0,8 als akzeptabel betrachtet werden. In der empirischen Praxis werden meist auch niedrigere Koeffizienten akzeptiert. Die Berechnung von Cronbachs Alpha ist mit jedem Standard-Statistikprogramm problemlos möglich (siehe zum Verfahren Cronbach 1951; Cortina 1993).

7.4.3 Validität

Mit Validität ist die Gültigkeit eines Messinstruments gemeint. Ein Messinstrument ist valide (gültig), wenn es tatsächlich das misst, was es messen soll. Es geht um den Grad der Genauigkeit, mit der ein bestimmtes Konzept gemessen werden soll. Im Mittelpunkt des Gütekriteriums Validität stehen damit die Indikatoren, die ausgewählt und/oder entwickelt wurden, um das Konzept empirisch zu erfassen. Wie weiter oben bereits dargestellt, besteht das zentrale Problem der Validität darin, ob die Zuordnung eines Indikators zu einem theoretischen Konzept überhaupt gerechtfertigt ist (Schnell et al. 2013, S. 121). Deshalb ist die Auswahl der Indikatoren stets zu begründen bzw. in Bezug zum Forschungsstand offenzulegen.

In der Literatur werden drei Formen der Validität unterschieden: Inhaltsvalidität, Kriteriumsvalidität und Konstruktvalidität (z. B. Diekmann 2011, S. 258; Schnell et al. 2013, S. 144–148; Häder 2015, S. 109–110).

Inhaltsvalidität Die Inhaltsvalidität bezieht sich darauf, dass möglichst alle Dimensionen des Konzepts, das gemessen werden soll, bei der Messung auch berücksichtigt werden. Eine valide Messung des Konzepts „Kognitive Mobilisierung“ müsste folglich alle Dimensionen des Konzepts abdecken. Eine gültige Messung des Konzepts „Intelligenz“ müsste entsprechend alle Dimensionen abdecken, die bei der Spezifikation des theoretischen Konzepts „Intelligenz“ offengelegt wurden. Für jede Dimension sind entsprechende Indikatoren auszuwählen, sodass alle Dimensionen bei der Messung berücksichtigt werden können. Nach Schnell et al. (2013, S. 145) sollte die Inhaltsvalidität als Idee aufgefasst werden, die bei der Konstruktion eines Messinstruments nützlich sein kann. Nur wenn in der Phase der Konzeptspezifikation die Dimensionen spezifiziert wurden, können bei der Operationalisierung entsprechende Instrumente entwickelt werden.

Kriteriumsvalidität Die Kriteriumsvalidität bezieht sich auf einen Vergleich der empirischen Ergebnisse des eigenen Messinstruments mit den Ergebnissen eines anderen (etablierten) Messinstruments. Die Kriteriumsvalidität ist gegeben, wenn die Ergebnisse der beiden Instrumente vergleichbar sind. Mit der Kriteriumsvali-

dität sind zwei offensichtliche Probleme verbunden: Erstens muss ein geeignetes (und allgemein etabliertes) Messinstrument existieren, um das eigene Instrument damit vergleichen zu können. Falls ein solches Instrument existiert, stellt sich allerdings zweitens die Frage, warum nicht gleich dieses etablierte Messinstrument verwendet wird. Wie Häder (2015, S. 110) zu Recht betont, dürften damit „der Nutzung der Kriteriumsvalidität in der sozialwissenschaftlichen Praxis relativ enge Grenzen gesetzt sein".

Konstruktvalidität Inhalts- und Kriteriumsvalidität sind einerseits wenig aussagekräftig und andererseits selten anwendbar (Schnell et al. 2013, S. 146). In den Sozialwissenschaften ist insbesondere die Konstruktvalidierung von großer Bedeutung. Was ist Konstruktvalidität?

> *„Konstruktvalidität verlangt, dass das von einem Messinstrument erfasste Konstrukt mit möglichst vielen anderen Variablen in theoretisch begründbaren Zusammenhängen steht und hieraus Hypothesen ableitbar sind, die einer empirischen Prüfung standhalten." (Diekmann 2011, S. 258)*

Für die Prüfung der Konstruktvalidität werden mehrere theoretisch plausible Hypothesen über den Zusammenhang zwischen dem zu messenden Konzept und anderen Konzepten formuliert und empirisch überprüft. Bei der Konstruktvalidierung des Konzepts „Kognitive Mobilisierung" müssten entsprechend Hypothesen über den Zusammenhang zwischen der kognitiven Mobilisierung und anderen Konzepten entwickelt werden. Mögliche Hypothesen könnten sein:

- Je höher die kognitive Mobilisierung, desto größer das politische Vertrauen.
- Je höher die kognitive Mobilisierung, desto stärker das politische Interesse.
- Je höher die kognitive Mobilisierung, desto geringer die politische Entfremdung.

Diese Hypothesen werden empirisch überprüft. Bei einer Bestätigung der Hypothesen wird Konstruktvalidität unterstellt. Da die Konstruktvalidierung ein kumulatives Forschungsprogramm ist, können einzelne Hypothesentests allerdings nur Teilergebnisse liefern. Erst die Gesamtheit der Hypothesentests gibt über die Konstruktvalidität eines Messinstruments Auskunft (Diekmann 2011, S. 259). Kann eine Hypothese allerdings nicht bestätigt werden, dann sind mehrere Ursachen möglich (Schnell et al. 2013, S. 147):

1) Es liegt tatsächlich keine Konstruktvalidität vor.
2) Die formulierte Hypothese ist falsch.

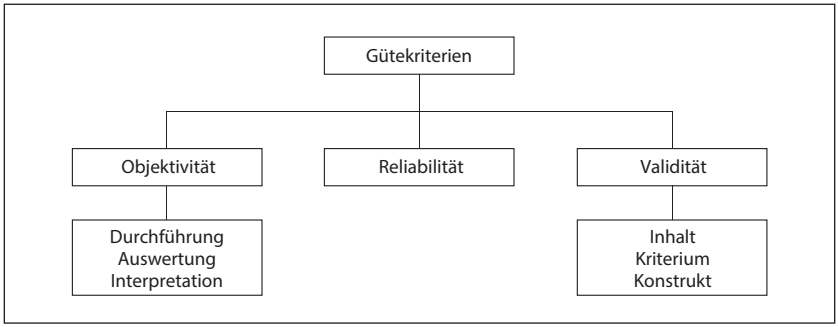
3) Die empirische Untersuchung ist fehlerhaft.
4) Die im Rahmen des Validierungsprozesses verwendeten anderen Konzepte (z. B. politisches Interesse oder politisches Vertrauen) sind selbst nicht valide erfasst.

Bei der Prüfung der Konstruktvalidität sollten insbesondere die zweite, dritte und vierte Fehlerursache ausgeschlossen werden. Deshalb sollten insbesondere etablierte Hypothesen formuliert und valide Konzepte verwendet werden.

7.4.4 Zusammenfassung

In Abbildung 14 sind die vorgestellten Gütekriterien dargestellt. Messungen sollen möglichst objektiv, zuverlässig (reliabel) und gültig (valide) sein. Bei den einzelnen Gütekriterien lassen sich zudem Unterformen unterscheiden. Bei der Objektivität wird zwischen der Durchführungs-, Auswertungs- und Interpretationsobjektivität differenziert. Bei den quantitativen Methoden ist insbesondere die Durchführungsobjektivität sicherzustellen, das heißt, das Messergebnis soll unabhängig von der Person sein, die die Messung vornimmt. Bei der Validität werden Inhalts-, Kriteriums- und Konstruktvalidität unterschieden. In der empirischen Forschungspraxis kommt insbesondere der Konstruktvalidität eine besondere Bedeutung zu, da Inhaltsvalidität nicht objektiv überprüfbar und Kriteriumsvalidierung nur bei (existierenden) geeigneten externen Messinstrumenten anwendbar ist.

Abbildung 14 Übersicht der Gütekriterien



Quelle: Eigene Darstellung

Diekmann (2011, S. 261) betont das hierarchische Verhältnis der drei Gütekriterien Objektivität, Reliabilität und Validität:

> *„Objektivität ist eine notwendige, aber nicht hinreichende Bedingung für Reliabilität. Die Reliabilität ist wiederum eine notwendige, aber nicht hinreichende Bedingung der Validität. Ziel ist die Konstruktion valider Messinstrumente.“*

Ein Messinstrument kann zwar objektiv und reliabel sein, muss aber noch lange nicht valide sein. Ein Beispiel ist die digitale Personenwaage, die jeweils zehn Kilogramm zu wenig misst. Das Testergebnis wäre objektiv, da verschiedene Personen (z. B. Ärzte) mit dem gleichen Messinstrument (z. B. Waage) zum gleichen Ergebnis kommen. Das Instrument wäre auch reliabel, da es das Körpergewicht systematisch um zehn Kilogramm unterschätzt. Aber das Messinstrument ist nicht valide, da der wahre Wert des Konzepts (Körpergewicht) nicht korrekt gemessen wird. Deshalb sind Objektivität und Reliabilität notwendige, aber noch keine hinreichenden Bedingungen für ein valides Messinstrument.

7.5 Indexbildung

Für die Operationalisierung einfacher Konzepte wie Alter oder Geschlecht ist in der Regel ein Indikator ausreichend. In den Sozialwissenschaften sind wir allerdings häufig mit komplexen Konzepten konfrontiert, die nicht nur mit einem Indikator erfasst werden können. In der empirischen Praxis werden deshalb mehrere Indikatoren verwendet, um Konzepte wie sozialer Status, Vertrauen oder auch die Haltung gegenüber Immigranten zu erfassen. Die einzelnen Indikatoren werden dann in einem Index zusammengefasst.

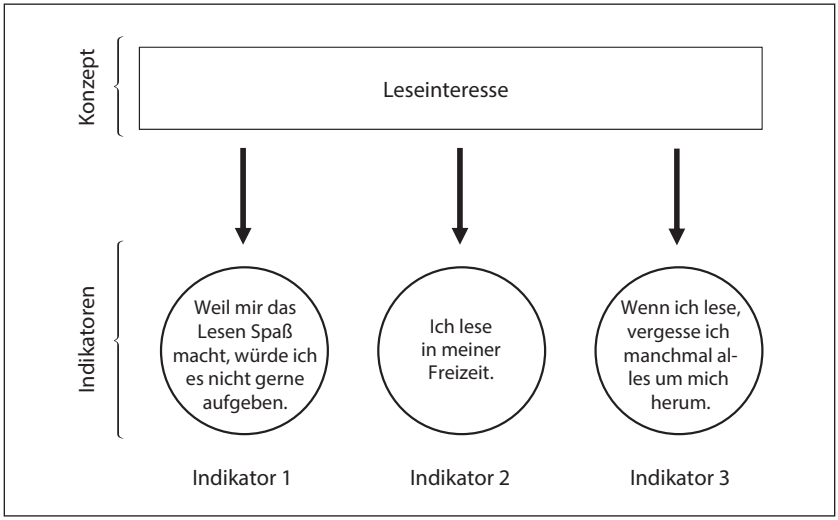
- **Definition: Index**
  Ein Index ist eine Zusammenfassung von mehreren Einzelindikatoren. Miller (2007, S. 138) definiert Indizes als „Maße (‚composite measures‘), die nach klar definierten Regeln aus zwei oder mehr Indikatoren zusammengesetzt sind“.

Die Indexbildung lässt sich an einem Beispiel aus der Bildungsforschung erläutern. Bei der PISA-Studie werden nicht nur die Kompetenzen von Schülern „gemessen“, sondern auch motivationale Merkmale der Jugendlichen erfasst.[22] Ein

22 PISA ist die Abkürzung für „Programme for International Student Assessment“. Dabei handelt es sich um eine vergleichende Studie, welche die Basiskompetenzen erfasst, die in modernen Gesellschaften für eine aktive Teilnahme am gesellschaftlichen Leben erforderlich sind. Für weitere Informationen zur PISA-Studie siehe: www.oecd.org/de/pisa.

solches motivationales Merkmal ist das Leseinteresse. Beim Leseinteresse handelt es sich um ein theoretisches Konzept, das nicht direkt beobachtbar ist. Für die empirische Messung sind Indikatoren erforderlich, die das Konzept abbilden. Wie Abbildung 15 zeigt, haben die PISA-Forscherinnen aber nicht nur ein, sondern gleich drei Indikatoren zur Messung des Leseinteresses entwickelt (Kunter et al. 2002, S. 165).

Abbildung 15 Konzept mit mehreren Indikatoren (Beispiel: Leseinteresse)



Quelle: Eigene Darstellung

Bei der PISA-Studie waren die Jugendlichen aufgefordert, zu folgenden Aussagen Stellung zu beziehen:

- Weil mir das Lesen Spaß macht, würde ich es nicht gerne aufgeben.
- Ich lese in meiner Freizeit.
- Wenn ich lese, vergesse ich manchmal alles um mich herum.

Dabei konnten die Schüler zwischen folgenden Antwortmöglichkeiten wählen: trifft nicht zu (1), trifft eher nicht zu (2), trifft eher zu (3) oder trifft zu (4). Die Antwortmöglichkeiten sind das empirische Relativ, das durch die Vergabe der Ziffern in ein numerisches Relativ „übersetzt“ wird. Um aus diesen Indikatoren einen In-

dex zu bilden, werden in der empirischen Praxis die genannten Antwortmöglichkeiten (also die Ziffern) zusammengezählt. Bei einer Person, die alle Aussagen mit „trifft zu“ beantwortet hat, wäre dies 12. Bei einem Jugendlichen, der die erste Aussage mit „trifft nicht zu“ (1) und die beiden anderen Statements mit „trifft eher zu“ (3) beantwortet hat, würde sich eine 7 ergeben. Ein Schüler, der alle Aussagen mit „trifft nicht zu“ (1) beantwortet, hätte eine 3. Man kann aber auch die durch die Addition der einzelnen Antworten gebildete Summe noch durch die Anzahl der Indikatoren dividieren. Dies hat den Vorteil, dass der Indexwert wieder mit dem ursprünglichen Wertebereich vergleichbar ist. In Tabelle 10 finden sich die Angaben für vier fiktive Jugendliche, die die Aussagen beantwortet haben.

Tabelle 10 Beispiele für additive Indexbildung

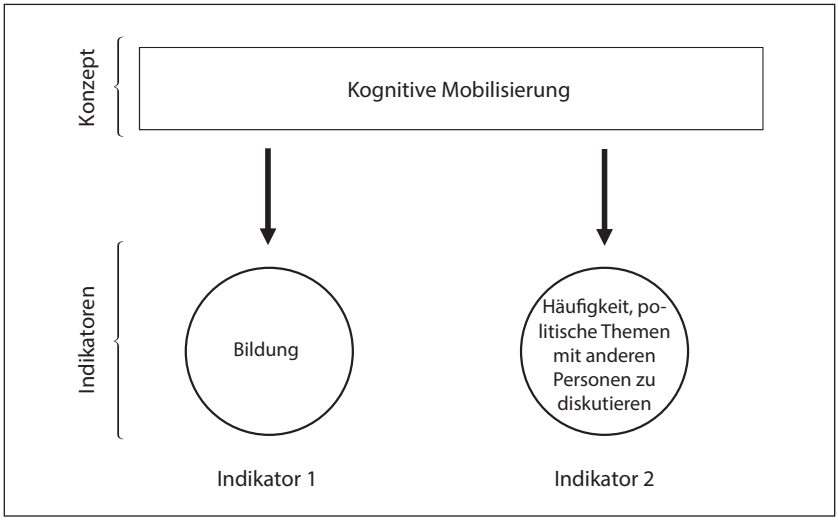
ID	Aussage 1	Aussage 2	Aussage 3	Indexwert I	Indexwert II
1	trifft zu (4)	trifft zu (4)	trifft zu (4)	12	4
2	trifft nicht zu (1)	trifft eher zu (3)	trifft eher zu (3)	7	2,33
3	trifft nicht zu (1)	trifft nicht zu (1)	trifft nicht zu (1)	3	1
4	trifft eher zu (3)	trifft nicht zu (1)	trifft nicht zu (1)	5	1,66

In unserem PISA-Beispiel wird der Indexwert einfach durch die Addition der einzelnen Indikatorenwerte ermittelt. Die Verknüpfung der einzelnen Indikatoren ist damit additiv. In der empirischen Praxis werden die meisten Indizes durch die (einfache) Addition der einzelnen Indikatoren gebildet. Dadurch kann ein niedriger Punktwert auf einem Indikator durch einen hohen Punktwert auf einem anderen Indikator ausgeglichen werden.

Verknüpfung der Indikatoren Indizes müssen allerdings nicht immer additiv verknüpft werden. Gelegentlich ist gerade die Eigenschaft, dass ein niedriger Punktwert auf einem Indikator durch einen hohen Punktwert auf einem anderen Indikator ausgeglichen wird, unerwünscht. In einem solchen Fall wäre eine multiplikative Verknüpfung der einzelnen Indikatoren möglicherweise vorzuziehen. Betrachten wir das Eingangsbeispiel „Kognitive Mobilisierung“. Für die Operationalisierung des Konzepts existieren in der Fachliteratur viele unterschiedliche Vorschläge. Offensichtlich umfasst das Konzept (mindestens) zwei Aspekte: „Kognition“ und „Mobilisierung“. Daher ist es sicherlich angemessen, jeweils (mindestens) einen Indikator für einen Aspekt zu verwenden (siehe Abbildung 16). Für den Aspekt „Kognition“ könnte beispielsweise das formale Bildungsniveau ge-

nutzt werden; für den Aspekt „Mobilisierung“ könnten die Personen gefragt werden, wie häufig sie politische Themen mit anderen Personen diskutieren. Zur Messung der beiden Indikatoren werden jeweils Zahlen von 0 bis 5 vergeben; höhere Zahlen deuten jeweils auf eine höhere Bildung bzw. auf eine häufigere Diskussion über Politik hin.

Abbildung 16 Indexbildung des Konzepts „Kognitive Mobilisierung“



Quelle: Eigene Darstellung

Additive versus multiplikative Verknüpfung Bei einer additiven Indexbildung wird jetzt ein niedriger Punktwert auf einem Indikator (niedriger Bildungsabschluss) durch einen hohen Punktwert auf dem anderen Indikator (häufige Diskussion über Politik) ausgeglichen. Wenn jetzt aber eine Person niemals über Politik spricht bzw. über „keine“ Bildung verfügt, dann scheint eine kognitive Mobilisierung zumindest fraglich. Eine multiplikative Verknüpfung der beiden Indikatoren ist daher eine alternative Operationalisierungsstrategie. Wenn einer der beiden Indikatoren den Wert 0 (keine Bildung bzw. keine politische Diskussion) annimmt, dann ist der Indexwert ebenfalls 0. Exemplarische Beispiele für eine multiplikative Verknüpfung finden sich in Tabelle 11.

In der Forschungspraxis werden häufig mehrere Operationalisierungen entwickelt und gegeneinander abgewogen. Hypothesen können auf Basis unter-

Tabelle 11 Beispiele für multiplikative Indexbildung

ID	Bildungsniveau	Diskussionen über Politik	Indexwert
1	kein Abschluss (0)	nie (0)	0
2	Hochschulreife (5)	täglich (5)	25
3	Hochschulreife (5)	nie (0)	0
4	Realschulabschluss (2)	gelegentlich (2)	4

schiedlicher Operationalisierungen geprüft werden. Führen verschiedene Operationalisierungen eines Konzepts zu vergleichbaren empirischen Ergebnissen, dann stärkt dies unser Vertrauen in die empirischen Ergebnisse.

Europäische Parlamente als Wahlorgane

Die Entwicklung von überzeugenden Messinstrumenten ist eine theoretische und empirische Herausforderung. Ein gelungenes Beispiel für eine äußerst transparente Vorgehensweise bei der Entwicklung eines Messinstruments bietet Sieberer (2008). Er präsentiert einen Vorschlag, um die Stärke von Parlamenten als Wahlorgane im europäischen Vergleich empirisch abbilden zu können. Dabei werden mehrere Indikatoren zu einem Index gebündelt. Eine Übersicht zu Indizes in der Vergleichenden Politikwissenschaft bietet der Beitrag von Pickel und Pickel (2012).

Vorteile von Indizes Für die Verwendung von Indizes sprechen mindestens zwei Vorteile: Erstens eine bessere Abbildung der theoretischen Konzepte und zweitens die Verringerung von Messfehlern.

Einerseits sind viele Konzepte zu komplex, um sie nur mit einem Indikator empirisch abbilden zu können. Es werden daher mehrere Indikatoren entwickelt, um das Konzept empirisch zu erfassen. Damit stellt sich unmittelbar die Frage, wie viele Indikatoren zur Erfassung eines theoretischen Konzepts wünschenswert sind. Es ist kaum überraschend, dass es bei der Beantwortung dieser Frage häufig zu Zielkonflikten zwischen theoretisch Wünschenswertem und empirisch Machbarem kommt. Mehr Indikatoren sind in der Regel mit einem größeren Ressourcenaufwand verbunden. Zu den Ressourcen gehört auch die Geduld des Befragten beim Ausfüllen des Fragebogens. Wer möchte schon (freiwillig) stundenlang

ähnlich klingende Aussagen ankreuzen (oder beantworten), wenn zur gleichen Zeit das Finale der Champions League im Fernsehen übertragen wird? In der Forschungspraxis kommt es daher oft zu (notwendigen) Kompromissen zwischen Theorie und Empirie. Diese Kompromisse sind allerdings kein Freifahrtschein für die Entwicklung „unsauberer" Messinstrumente.

Andererseits können Indizes die Reliabilität einer Messung verbessern. Wie in Abschnitt 7.4.2 dargestellt, lassen sich bei der Messung von Konzepten Messfehler praktisch nicht vermeiden. Durch eine Zusammenfassung mehrerer Indikatoren zu einem Index kann der „wahre" Wert eines Konzepts zuverlässiger abgebildet werden.

Zusammenhang zwischen Bildung und Einkommen

Unsere Hypothese „Je höher die Bildung, desto größer das Einkommen" ist sicherlich plausibel. Für Bildung gibt es allerdings verschiedene Möglichkeiten der Operationalisierung: Den formalen Bildungsabschluss, die Anzahl der Jahre im Bildungssystem oder auch einen eigens entwickelten Bildungstest. Unabhängig von der konkret gewählten Operationalisierung sollte empirisch aber immer ein positiver Zusammenhang zwischen Bildung und Einkommen zu beobachten sein: Wenn die Bildung steigt, dann sollte tendenziell auch das Einkommen steigen. Verschiedene Operationalisierungen führen damit zwar nicht zu identischen, aber doch zu vergleichbaren empirischen Ergebnissen.

Primär- und Sekundäranalyse

8

In diesem Kapitel werden die Begriffe Primär- und Sekundäranalyse erläutert sowie die Vor- und Nachteile dieser Forschungsstrategien aufgezeigt. Bei den meisten Forschungsprojekten werden keine eigenen Daten erhoben, sondern es wird bereits existierendes Datenmaterial zur Überprüfung der Hypothesen genutzt. Deshalb werden in diesem Kapitel auch zentrale Datensätze bzw. Institutionen, die sozialwissenschaftliche Datensätze zur Verfügung stellen, vorgestellt. Das Kapitel schließt mit Hinweisen zur Recherche nach sozialwissenschaftlichen Datensätzen.

8.1 Was ist eine Primäranalyse? Was ist eine Sekundäranalyse?

Die Forschungsfrage ist formuliert, die theoretischen Konzepte sind geklärt, Hypothesen abgeleitet und geeignete Operationalisierungen gefunden. Diese vier Phasen eines quantitativ orientierten Forschungsprojekts bilden das Fundament für den eher praktischen, aber weiterhin theoriegeleiteten empirischen Teil eines Forschungsprojekts. Dabei ist in allen vier Phasen eine explizite Anbindung an den existierenden Forschungsstand wichtig. Wer nicht die Fachliteratur zu einem Forschungsthema kennt, der kann keine (theoretisch und gesellschaftlich) relevante Forschungsfrage entwickeln, der wird theoretische Diskussionen bei der Konzeptspezifikation nicht berücksichtigen, der wird langweilige Hypothesen formulieren und bei der Operationalisierung Fehler wiederholen. Deshalb ist die intensive Auseinandersetzung mit dem Forschungsstand die Grundlage für die Entwicklung und Durchführung eines Forschungsprojekts.

Im Mittelpunkt des praktischen Teils eines Forschungsprojekts steht die Datenerhebung. Die Festlegung des Forschungsdesigns und des Auswahlverfahrens

© Springer Fachmedien Wiesbaden GmbH, ein Teil von Springer Nature 2018

M. Tausendpfund, *Quantitative Methoden in der Politikwissenschaft*,
Grundwissen Politik, https://doi.org/10.1007/978-3-658-20698-7_8

Zur Erinnerung: Hausarbeit ist kein Forschungsprojekt

In einer Haus- oder Abschlussarbeit steht die intensive Auseinandersetzung mit dem existierenden Forschungsstand im Mittelpunkt. Auf Basis des existierenden Forschungsstands soll eine interessante Forschungsfrage formuliert und beantwortet werden. Zur Beantwortung der Forschungsfrage müssen empirische Studien gelesen und verstanden werden. Die Kenntnis der einzelnen Phasen eines quantitativen Forschungsprozesses mit ihren spezifischen Herausforderungen und Fallstricken soll helfen, empirische Studien lesen, reflektieren, bewerten und systematisch mit der Forschungsfrage verknüpfen zu können.

sind in diesem Kontext „nur" vorbereitende Schritte, die den Rahmen für eine möglichst effiziente Datenerhebung vorgeben. Unabhängig davon, ob eine Befragung durchgeführt wird, eine Beobachtung vorgesehen ist oder Texte inhaltsanalytisch ausgewertet werden, ist jede Form der Datenerhebung teuer, zeitaufwändig und personalintensiv (Behnke 2009, S. 392). Deshalb macht die Planung und Durchführung der Erhebung empirischer Daten in Forschungsprojekten einen großen Anteil der kalkulierten Ressourcen aus. Mit Ressourcen sind dabei unter anderem Zeit, Personal und Geld gemeint. In einem Standardwerk zu Befragungen erläutert Schnell (2012, S. 23):

> *„Der notwendige Aufwand für Befragungen wird fast immer unterschätzt. Dies beginnt mit der erforderlichen Zeit. Selbst wenn der seltene Fall einer präzise definierten inhaltlichen Fragestellung gegeben sein sollte, werden allein für die Konstruktion und den Test eines Fragebogens in der Regel mehrere Wochen erforderlich sein."*

Neben der Konstruktion und der Prüfung (Pretest) eines Fragebogens müssen die Zielpopulation definiert und ausgewählt, die Erhebung durchgeführt und die Daten für eine spätere Analyse aufbereitet werden. Dies beinhaltet auch eine umfangreiche Datenkontrolle, die bei einem größeren Projekt (z. B. ALLBUS) durchaus mehrere Monate dauern kann. Je nach Forschungsfrage ist es deshalb weder sinnvoll noch erforderlich eine eigene Datenerhebung durchzuführen, um die eigene Forschungsfrage zu beantworten. Bei der Mehrheit der wissenschaftlichen Forschungsprojekte werden auch nicht eigene Datenerhebungen durchgeführt, sondern die Wissenschaftlerin greift auf bereits existierendes Datenmaterial zurück. In der Sprache der empirischen Sozialforschung werden solche Studien dann als Sekundäranalysen bezeichnet.

- **Definition: Primär- und Sekundäranalyse**
  Bei einer Primäranalyse werden zur Überprüfung der formulierten Hypothesen neue Daten erhoben. Bei einer Sekundäranalyse werden bereits vorhandene Daten genutzt, um Hypothesen zu überprüfen. Solche Daten werden als Sekundärdaten bezeichnet (Porst 2014b).

Um Missverständnissen vorzubeugen: Mit Primär- und Sekundäranalyse ist kein wertendes Urteil verbunden. Primäranalysen sind nicht etwa besser als Sekundäranalysen. Im Gegenteil: Eine Sekundäranalyse auf Basis sorgfältig erhobener Daten ist einer Primäranalyse, die Daten nutzt, bei denen das Auswahlverfahren unklar ist und die eigentliche Erhebung „unsauber" durchgeführt wurde, eindeutig vorzuziehen.

8.2 Vor- und Nachteile von Primär- und Sekundäranalysen

In der Literatur lassen sich drei Argumente für die Verwendung von Sekundärdaten unterscheiden: Erstens stehen für die Durchführung einer eigenen Erhebung keine ausreichenden Ressourcen zur Verfügung. Zweitens existieren geeignete Daten, die zur Beantwortung der Forschungsfrage genutzt werden können. Drittens erfordert die Kontrolle wissenschaftlicher Befunde (mehr) Sekundäranalysen. Diese drei Vorteile werden im Folgenden genauer betrachtet.

Datenerhebung erfordert Ressourcen Eine Datenerhebung erfordert Zeit, Personal und Geld. Stehen diese Ressourcen nicht zur Verfügung, dann sollte von der Durchführung einer Erhebung Abstand genommen werden (Schnell 2012, S. 25). Diese Empfehlung gilt für alle Forschungsprojekte, aber insbesondere für geplante Erhebungen im Rahmen von Haus- und Abschlussarbeiten. Mit Blick auf quantitative Studien – insbesondere der standardisierten Befragung von Personen – betonen Kopp und Lois (2014, S. 31):

> *„Es ist eine Illusion, dass man im Rahmen eines Lehrforschungsprojekts oder gar einer Bachelor- oder Masterarbeit genügend Zeit hat, alle Schritte einer empirischen Erhebung so sorgfältig durchzuführen, dass man mit tragfähigen Ergebnissen rechnen darf. Datenerhebung ist keine Laienarbeit, die man einfach mal so schnell erledigen kann!"*

Auch andere Erhebungsformen (z. B. Inhaltsanalysen) sind aufwändiger als es auf den ersten Blick scheint. Bei einer Inhaltsanalyse müssen beispielsweise die entsprechenden Dokumente beschafft, ein Kodierschema erstellt und die Dokumente kodiert werden. Ein Blick in die klassischen Einführungen in die empirische So-

zialforschung (z. B. Diekmann 2011; Schnell et al. 2013) verdeutlicht die Herausforderungen, mit denen sich Forscher bei der Planung und Durchführung einer quantitativen Datenerhebung konfrontiert sehen. Viele dieser Herausforderungen und Probleme werden häufig erst in der eigentlichen Phase der Datenerhebung (der sogenannten Feldphase) sichtbar, sodass die Datenerhebung meist länger dauert als maximal veranschlagt.

Verfügbarkeit existierender Daten Falls keine Ressourcen für eine eigene Datenerhebung zur Verfügung stehen, dann sollte keine eigene Datenerhebung durchgeführt werden. Dieser Ratschlag dürfte den meisten Personen keine Probleme bereiten, da in vielen Fällen eine eigene Datenerhebung für die Beantwortung der Forschungsfrage überhaupt nicht erforderlich ist. Erstens existieren mittlerweile umfangreiche Datenbestände, die zur Beantwortung unterschiedlichster Fragestellungen verwendet werden können. Zweitens werden zahlreiche sozialwissenschaftliche Erhebungen nicht allein zur Beantwortung einer Forschungsfrage durchgeführt, sondern erlauben die Bearbeitung verschiedener Forschungsfragen. Solche sozialwissenschaftlichen Erhebungen sind etwa die Allgemeine Bevölkerungsumfrage für Sozialwissenschaften (ALLBUS) oder der European Social Survey (ESS). Diese allgemeinen Erhebungen sind ein Kollektivgut der wissenschaftlichen Gemeinschaft, das allen Forschern (auch Studierenden) zur Verfügung steht.

Die Nutzung existierender Daten für die Überprüfung eigener Hypothesen bzw. zur Beantwortung eigener Forschungsfragen wird Sekundäranalyse genannt. Für empirische – quantitative – Haus- und Abschlussarbeiten sind Sekundäranalysen die einzige sinnvolle Strategie, da im Rahmen einer Haus- oder Abschlussarbeit nicht ausreichend Ressourcen zur Verfügung stehen, um selbst Daten zu erheben. Das einzige Problem bei einer Sekundäranalyse besteht für den Forscher darin, dass er für seine Forschungsfrage geeignete Daten finden muss:

> *„Es müssen dann Indikatoren bzw. Variablen aus vorhandenen Datenbeständen für die nunmehr interessierenden Konstrukte gefunden werden. Zugleich muss das Auswahlverfahren, welches für die bei der Sekundäranalyse benutzten Daten angewendet wurde, auch mit dem nunmehr angestrebten Aussagebereich kompatibel sein. So nützt es wenig, wenn geeignete Indikatoren vorliegen, diese jedoch nur an männlichen Jugendlichen erhoben wurden und die Sekundäranalyse Aussagen über Erwachsene anstrebt." (Schnell et al. 2013, S. 242)*

Die Kenntnis existierender Datenbestände ist wichtig, um einerseits vorhandenes Datenmaterial mit Blick auf die eigene Forschungsfrage finden zu können und andererseits unnötige Erhebungen – auch mit Blick auf eine Überlastung von Be-

fragten – zu vermeiden. In Abschnitt 8.3 werden wichtige sozialwissenschaftliche Datensätze vorgestellt, in Abschnitt 8.4 finden sich allgemeine Hinweise zur Recherche von Datensätzen.

Notwendigkeit von Sekundäranalysen Der dritte Vorteil von Sekundäranalysen bezieht sich auf die Kontrolle von Forschungsergebnissen. Wer heute Auto, Fernseher oder Handy kauft, der wird vor dem Kauf wahrscheinlich Bewertungen zu dem bevorzugten Produkt recherchieren und lesen. In umfangreichen Tests werden die unterschiedlichsten Produkte auf „Herz und Nieren" geprüft. In den Sozialwissenschaften wird die Kontrolle empirischer Studien Replikation oder auch Replikationsstudie genannt. Solche Replikationsstudien dienen der Qualitätskontrolle empirischer Studien und sind für die empirische Sozialforschung unentbehrlich.

> *„… the only way to understand and evaluate an empirical analysis fully is to know the exact process by which the data were generated and the analysis produced." (King 1995, S. 444)*

Selbstverständlich kann nicht jeder Forscher alle gelesenen Studien nochmals überprüfen. Dies wäre weder sinnvoll noch angemessen. Allerdings sollte jede Forscherin die Chance haben, eine gelesene Studie überprüfen zu können. Damit eine Überprüfung aber überhaupt möglich ist, muss eine empirische Studie sorgfältig dokumentiert, die Datengrundlage verfügbar und die einzelnen Analyseschritte (z. B. Operationalisierung) nachvollziehbar beschrieben sein. King (1995) hat entsprechende Empfehlungen formuliert, die in der Forschungspraxis aber erst in Ansätzen umgesetzt sind (z. B. Gherghina und Katsanidou 2013).

Replikationsstudien sind wichtig, um Nachlässigkeiten, Fehler oder auch bewusste Fälschungen aufdecken zu können. Bei Diekmann (1998) finden sich mehrere Beispiele für falsche oder zumindest „unsauber" publizierte Forschungs-

Wissenschaftliche Replikationen ermöglichen: Replikationsserver.de

Replikationsserver.de ist eine gemeinsame Initiative der Zeitschrift für Soziologie, der Sozialen Welt und GESIS zur Bereitstellung sozialwissenschaftlicher Daten zu Replikationszwecken. Autorinnen und Autoren der genannten Zeitschriften verpflichten sich dazu, quantitative Daten und dazugehörige Analyseskripte zu ihrer Publikation öffentlich zugänglich zu machen.

ergebnisse. Eine gelungene Sekundäranalyse hat Schnaudt (2013) vorgelegt. Die Arbeit von Schnaudt (2013) zeigt, dass auch in begutachteten Fachzeitschriften durchaus empirische Befunde publiziert werden, die kritisch zu hinterfragen sind. In der Kölner Zeitschrift für Soziologie und Sozialpsychologie findet sich eine Kontroverse zwischen Reuband (2001, 2002) und Schnell (2002), die sich thematisch mit postalischen Befragungen beschäftigt. Die Beispiele unterstreichen die Bedeutung der Sekundäranalyse zur Kontrolle von Forschungsergebnissen.

Autoren empirischer Studien sind bei der Veröffentlichung eines Beitrags in einschlägigen Fachzeitschriften mittlerweile verpflichtet, die Dokumentation und die verwendeten Daten für Replikationszwecke verfügbar zu machen. Nach Diekmann (1998) wird der Trend zu (unbeabsichtigten) Fehlern bei empirischen Studien allerdings eher zu- als abnehmen. Für diese These sprechen die steigende Komplexität von Datensätzen und statistischen Analyseverfahren, die zunehmende Arbeitsteilung bei einer Publikation sowie der gestiegene Publikationsdruck. Folgt man Diekmann (1998, S. 140), dann haben Sekundäranalysen sowohl eine „heilende als auch eine präventive Wirkung". Eine „heilende Wirkung", weil dadurch Irrtümer und Fehler bei empirischen Studien aufgedeckt werden. Eine „präventive Wirkung", weil Forscher vermutlich ihre Vorgehensweise genauer dokumentieren und ihre Ergebnisse besser kontrollieren, wenn die Wahrscheinlichkeit einer Sekundäranalyse hoch ist. Die Überschrift des Beitrags von King (1995) „Replication, Replication" ist daher als Aufforderung für alle (angehenden) Sozialwissenschaftler zu verstehen.

Nachteile von Sekundäranalysen Die Vorteile von Sekundäranalysen sind enorm, allerdings können natürlich nicht alle sozialwissenschaftlichen Fragestellungen mit Sekundärdaten bearbeitet werden. Insbesondere bei neuen „Themen", überarbeiteten theoretischen Konzepten sowie alternativen Operationalisierungen sind weiterhin Primärerhebungen erforderlich. Zudem können auch Sekundäranalysen problematisch sein, da bei der Operationalisierung der abhängigen und unabhängigen Variablen nicht immer die (theoretisch) optimale Lösung realisiert werden kann. Nach Roose (2013) können Sekundäranalysen auch zu einer Fehlermultiplikation und einer Pfadabhängigkeit der Erkenntnis führen. Deshalb muss die Wissenschaftlerin bei jedem Forschungsprojekt die Vor- und Nachteile von Primär- und Sekundärerhebung abwägen und – mit Blick auf die eigene Forschungsfrage – die Notwendigkeit einer Datenerhebung begründen (Primäranalyse) bzw. Argumente für den ausgewählten Datensatz präsentieren (Sekundäranalyse). Auch die Wahl eines Datensatzes für eine Sekundäranalyse ist deshalb zu erläutern (Warum ist gerade der ausgewählte Datensatz geeignet, um die Forschungsfrage zu beantworten?).

DFG zur Replizierbarkeit von Forschungsergebnissen

In einer Stellungnahme bewertet die Deutsche Forschungsgemeinschaft (DFG) die Replikation als ein „sehr wichtiges Verfahren zur Prüfung experimentalwissenschaftlich und quantitativ begründeter empirischer Wissensansprüche in der Medizin, in den Natur-, Lebens-, Ingenieur- sowie den Sozial- und Verhaltenswissenschaften und auch den Geisteswissenschaften". Allerdings, so die DFG in der Stellungnahme, ist Replizierbarkeit kein generelles Kriterium wissenschaftlicher Erkenntnis. In der Stellungnahme heißt es: „Die Erwartung, alle wissenschaftlichen Ergebnisse müssten replizierbar sein, ist schon deshalb nicht erfüllbar, weil zahlreiche Forschungsgebiete einmalige Ereignisse wie Klimawandel, Sternenexplosionen, Vulkanausbrüche oder vergangene Vorkommnisse untersuchen." Die Nicht-Replizierbarkeit sei deshalb auch kein genereller Falsifikationsbeweis. Den Anstoss zur öffentlichen Debatte um die „Replikationskrise" hatte eine Serie zur Replizierbarkeit von biomedizinischen Forschungsergebnissen in der Fachzeitschrift „Lancet" im Jahr 2014 gegeben. Die Pressemitteilung der DFG findet sich unter http://e.feu.de/replikation.

8.3 Sozialwissenschaftliche Datensätze

In diesem Kapitel werden zentrale sozialwissenschaftliche Datensätze bzw. wichtige Institutionen der empirischen Sozialforschung vorgestellt. Bei der Darstellung der Datensätze bzw. Institutionen wird nach der analytischen Ebene (Mikro- versus Makroebene) differenziert. In Abschnitt 8.3.1 werden sozialwissenschaftliche Erhebungen dargestellt, die auf der Individualebene erhoben werden (in der Regel durch Befragungen). Diese Daten erlauben es, die Einstellungen und Verhaltensweisen einer Person mit zahlreichen anderen individuellen Merkmalen in Beziehung zu setzen. In Abschnitt 8.3.2 werden Datensätze bzw. Institutionen dokumentiert, die Informationen über größere Analyseeinheiten wie Gemeinden, Länder und Staaten bereithalten (z. B. Arbeitslosenquote, Staatsverschuldung). Solche Daten werden häufig auch Aggregatdaten genannt.

Neben den aufgeführten Datensätzen bzw. Institutionen gibt es selbstverständlich noch zahlreiche weitere Datensätze, die sich für die Sekundäranalyse eignen: Wirth und Müller (2006) sowie Schimpl-Neimanns und Weiss (2014) beleuchten das Potenzial von amtlichen Daten für die empirische Sozialforschung. Bei Keil (2009) findet sich eine Übersicht der Datengrundlage der politischen So-

Infrastruktureinrichtung der Sozialwissenschaften: GESIS

Eine zentrale Infrastruktureinrichtung der Sozialwissenschaften ist GESIS – Leibniz-Institut für Sozialwissenschaften. Mit seinen 250 Mitarbeitern an den Standorten Mannheim, Köln und Berlin erbringt GESIS wichtige Dienstleistungen. Dazu gehört die Beratung bei sozialwissenschaftlichen Forschungsprojekten, die Archivierung von Datensätzen sowie das Angebot an Publikationen und Weiterbildungsveranstaltungen. GESIS ist Herausgeber von Fachzeitschriften (z. B. MDA) und Arbeitspapieren, die praktische Informationen zur Datenerhebung vermitteln. Die Homepage der Infrastruktureinrichtung ist unter www.gesis.org zu erreichen.

ziologie, und Behnke (2009) stellt nationale sowie internationale Datenarchive vor. Koop und Lois (2014, S. 31–43) informieren über wichtige soziologische Erhebungen, eine Darstellung von zentralen Datenquellen der vergleichenden Politikwissenschaft findet sich bei Stark und Mohamad-Klotzbach (2016). Hensel (2012) bietet eine Übersicht von Datensätzen in den internationalen Beziehungen. Eine knappe Darstellung zur Sekundäranalyse und dem Zugang zu sozialwissenschaftlichen Daten bietet auch der Beitrag von Porst (2014b), und der Band von Mallock et al. (2016) bietet einen allgemeinen Überblick über Informationsressourcen in den Sozialwissenschaften.

8.3.1 Mikroebene

Mittlerweile existieren zahlreiche Individualdatensätze, die sich für die unterschiedlichsten Fragestellungen eignen. Bei der Nutzung solcher Daten sind gelegentlich kleinere Abstriche zu machen, da möglicherweise interessierende Merkmale nicht bzw. nicht optimal erhoben wurden. Für die meisten Fragestellungen – nicht nur, aber insbesondere im Rahmen von Haus- und Abschlussarbeiten – sind solche Datensätze allerdings mehr als ausreichend. Zudem basiert auch ein Großteil der heute veröffentlichten Studien nicht auf Primärdaten, sondern nutzt bereits existierendes Datenmaterial. Die Arbeit mit solchen Datensätzen erfordert – neben einem gängigen Analyseprogramm (z. B. SPSS, Stata) – allerdings zumindest statistische Grundkenntnisse bzw. Kenntnisse in einem Statistikprogramm. Solche Datensätze erlauben es auch, die spannende Welt der empirischen Sozialforschung selbst kennenzulernen (und nicht nur darüber zu lesen). Wer selbst einmal mit echten Daten gearbeitet hat, kann die praktischen Probleme der

Datenaufbereitung und Datenanalyse meist besser nachvollziehen als ohne entsprechende praktische Erfahrung. Eine Übersicht wichtiger Mikrodaten der Sozialwissenschaften ist auch auf der Homepage der Arbeitsstelle „Quantitative Methoden" verfügbar: http://e.feu.de/mikrodaten.

ALLBUS Mit der Allgemeinen Bevölkerungsumfrage der Sozialwissenschaften (ALLBUS) werden seit 1980 alle zwei Jahre Daten über Einstellungen, Verhaltensweisen und Sozialstruktur der Bevölkerung in der Bundesrepublik Deutschland erhoben. Bei der ALLBUS handelt es sich um eine langfristig angelegte Befragung mit einem teils gleichen, teils variablen Fragenprogramm. Das heißt, einzelne – in der Regel demographische – Fragen etwa zu Geschlecht, Alter, Bildung oder beruflicher Stellung werden in jeder Erhebung gestellt. Jede Befragung hat aber auch einen thematischen Schwerpunkt (z. B. politische Beteiligung, Religion, soziale Ungleichheit), der in regelmäßigen Abständen wiederholt wird. Dadurch können Veränderungen und Entwicklungen über die Zeit verfolgt werden (Koch und Wasmer 2004).

Die ALLBUS ist eine sogenannte Querschnittstudie, das heißt, für jede Erhebung wird eine neue Stichprobe gezogen. Bis 1990 wurden etwa 3000 Personen in persönlich-mündlichen Interviews befragt. Seit 1992 werden in den alten Bundesländern etwa 2400, in den neuen Bundesländern 1100 Interviews durchgeführt. Dadurch sind ostdeutsche Befragte in der ALLBUS-Stichprobe überrepräsentiert. Diese Überrepräsentation wird als „oversample" bzw. „oversampling" bezeichnet. Dadurch können in Ostdeutschland für kleinere Bevölkerungsgruppen statistisch vertretbare Analysen durchgeführt werden. Wenn allerdings West- und Ostdeutschland gemeinsam als Gesamtdeutschland analysiert werden sollen, muss in der Regel die Überrepräsentation von ostdeutschen Befragten durch ein sogenanntes Designgewicht korrigiert werden.

Die ALLBUS wird von GESIS – Leibniz-Institut für Sozialwissenschaften verantwortet. Die Daten stehen über den GESIS-Datenbestandskatalog für den kostenfreien Download zur Verfügung, lediglich eine (ebenfalls kostenfreie) Registrierung ist erforderlich (www.gesis.org/allbus). Mit ALLBUS-Vollversion und ALLBUScompact lassen sich zwei Datenversionen unterscheiden. Beim ALLBUScompact handelt es sich um eine Einsteigerversion, die weniger demographische Merkmale enthält. ALLBUScompact eignet sich daher vor allem für Einsteiger in die Datenanalyse.

Was ist das Designgewicht?

Mit Designgewicht ist eine theoretisch-statistische Gewichtung aufgrund einer bekannten Auswahlwahrscheinlichkeit gemäß eines Stichprobenplans gemeint (Diekmann 2011, S. 427). In der ALLBUS sind Befragte aus den neuen Bundesländern überproportional vertreten, um gesonderte Analysen der ostdeutschen Teilstudie mit aussagekräftigen Fallzahlen zu gewährleisten. Bei einer Analyse für das gesamte Bundesgebiet würde ohne eine Gewichtung, bei der diese Überrepräsentierung korrigiert wird, die Auswertung allerdings verzerrt sein, da im Verhältnis mehr Ost- als Westdeutsche berücksichtigt werden. Diese Überrepräsentierung wird durch eine Gewichtungsvariable (das Designgewicht) korrigiert. Solche Gewichtungsvariablen sind in den jeweiligen Datensätzen bereits enthalten und müssen bei der statistischen Analyse berücksichtigt werden (siehe Terwey 2014 für eine ausführliche Erläuterung bei den ALLBUS-Daten).

Chapel Hill Expert Survey Der Chapel Hill Expert Survey (CHES) ist eine standardisierte Expertenbefragung zur Erfassung der nationalen Parteipositionen zu europäischer Integration, Ideologie und politischen Streitfragen in zahlreichen europäischen Ländern. Die erste Befragung wurde 1999 durchgeführt, weitere Befragungen folgten 2002, 2006, 2010 und 2014. Die Zahl der Untersuchungsländer stieg von 14 westeuropäischen Ländern im Jahr 1999 über 24 derzeitige oder potenzielle EU-Mitglieder im Jahr 2006 auf 31 Länder im Jahr 2014; gleichzeitig stieg auch die Zahl der betrachteten nationalen Parteien von 142 auf 268. Die Befragung 2014 umfasst alle EU-Mitgliedsstaaten sowie Parteien in Norwegen, der Schweiz und der Türkei. Der Datensatz und die Dokumentation ist frei im Internet verfügbar (http://chesdata.eu). Die Veröffentlichungen von Bakker et al. (2015) und Polk et al. (2017) bieten nützliche Informationen zum Design der Expertenbefragung und präsentieren empirische Informationen zu nationalen Parteipositionen im europäischen Vergleich.

DeutschlandTrend und Politbarometer Außerhalb von Wissenschaft und Forschung sind DeutschlandTrend und Politbarometer vermutlich die bekanntesten Erhebungen. Seit 1977 werden von der Forschungsgruppe Wahlen die Politbarometer für das ZDF durchgeführt, der DeutschlandTrend von Infratest dimap ist das Pendant zum Politbarometer für die ARD. Politbarometer und DeutschlandTrend sind keine akademischen Projekte; die Daten werden allerdings für die wissenschaftliche Lehre und Forschung zur Verfügung gestellt.

Beim Politbarometer und beim DeutschlandTrend werden die wahlberechtigten Bürger in telefonischen – meist monatlichen, in Wahlzeiten aber auch wöchentlichen – Erhebungen unter anderem zu ihrer Wahlabsicht und Parteipräferenz, zur Zufriedenheit mit der Leistung der Regierung und der Opposition, zur Demokratiezufriedenheit, zur Selbsteinschätzung auf einem Links-Rechts-Kontinuum oder auch zu den wichtigsten (wahrgenommen) Problemen in Deutschland befragt. Politbarometer und DeutschlandTrend zielen jedoch insbesondere darauf ab, das aktuelle (nicht immer nur) politische „Meinungsbild" zu erfassen. In einzelnen Erhebungen finden sich auch Fragen zu aktuellen Ereignissen (etwa zu Auslandseinsätzen der Bundeswehr, zum Bahn-Streik oder zur Kernenergie). Teilweise werden auch andere gesellschaftliche Fragestellungen in das Erhebungsprogramm aufgenommen (z. B. zur Fußball-Nationalmannschaft). Durch das stark variierende Frageprogramm und die im Vergleich zur ALLBUS nur knapp abgefragte Demographie, eignen sich beide Erhebungen nur eingeschränkt für komplexere wissenschaftliche Fragestellungen. Allerdings erlaubt gerade das Politbarometer, das seit 1977 erhoben wird, durchaus eine langfristige Analyse von Trends und Entwicklungen (z. B. zur Parteiidentifikation).

Umfangreiche Informationen zum Politbarometer finden sich auf der Homepage der Forschungsgruppe Wahlen (www.forschungsgruppe.de). Dort finden sich auch ein Glossar zur Umfrageforschung, Erläuterungen zur Methodik des Politbarometers sowie Antworten auf häufig gestellte Fragen (siehe aber auch Wüst 2003). Berichte zum DeutschlandTrend werden auch auf der Homepage von Infratest dimap veröffentlicht (www.infratest-dimap.de). Die Daten des Politbarometers (seit 1977) und des DeutschlandTrends (seit 2008) werden von GESIS aufbereitet und stehen über den GESIS-Datenbestandskatalog für den kostenfreien Download zur Verfügung.

Eurobarometer Das Eurobarometer ist das Umfrageinstrument der Europäischen Kommission (z. B. Schmitt 2003; Brinegar und Jolly 2004; Signorelli 2012). Bereits seit 1973 werden jeweils im Frühjahr und im Herbst etwa 1000 zufällig ausgewählte Bürger je EU-Mitgliedsland zu den unterschiedlichsten Themen – von Urlaubszielen über Umweltschutz bis hin zur Gesundheitsvorsorge – befragt. Die Europäische Kommission möchte von den Befragten aber auch wissen, wie sie über die Europäische Union denken. Deshalb sind Fragen zur EU selbst ein fester Bestandteil dieser Erhebungen: Wie bewerten die Bürger die Staatengemeinschaft? Gibt sie den Menschen ein Gefühl von Begeisterung oder von Misstrauen? Wünschen sich die Europäer eine schnellere oder langsamere Integration? Insbesondere die wissenschaftliche Forschung zu den Einstellungen der Bürger zur Staatengemeinschaft stützt sich zu einem großen Teil auf die Daten des Eurobarometers (z. B. Gabel 1998b; Kaina 2009; Braun und Tausendpfund 2014).

Die Bezeichnung „Eurobarometer“ ist etwas irreführend, da mittlerweile vier unterschiedliche Instrumente unter „Eurobarometer“ zusammengefasst werden: Das oben beschriebene Standard-EB (Standard Eurobarometer), das Special-EB (Eurobarometer Special Surveys), das Flash-EB (Flash Eurobarometer) und das EB-Qualitative (Qualitative studies). Beim Standard-Eurobarometer handelt es sich – wie weiter oben bereits berichtet – um eine halbjährliche persönliche Befragung von etwa 1000 Bürgern in jedem EU-Mitgliedsland. Das Special-EB ist eine – unregelmäßig durchgeführte – umfassende Befragung zu wechselnden Themen. Mit dem Flash-Eurobarometer sind telefonische Erhebungen zu konkreten Ereignissen gemeint (z. B. Referenden im Kontext des EU-Vertrags). Beim EB-Qualitative handelt es sich um ein qualitatives Erhebungsprogramm. Im Mittelpunkt der sozialwissenschaftlichen Forschung steht das Standard-EB, welches es erlaubt, die langfristige Entwicklung der Haltung der Bürger zur Staatengemeinschaft zu analysieren. Deshalb wird das Eurobarometer häufig mit dem Standard-EB gleichgesetzt.

! Auftraggeber des Eurobarometers ist – wie oben bereits erwähnt – die Europäische Kommission. Beim Eurobarometer handelt es sich damit nicht um ein akademisches, sondern um ein politisches Projekt. Als politischer Akteur sind die Ziele der Kommission nicht immer mit den Zielen der akademischen Forschung identisch. Dies betrifft zunächst einmal die Etablierung des Umfrageinstruments. Nach Pausch (2009, S. 541–542) diente die Etablierung einer europaweiten Meinungsumfrage dazu, „das elitengeleitete Projekt der europäischen Integration einer breiten Öffentlichkeit schmackhaft zu machen und die Entscheidungen der EG zu legitimieren“ sowie eine proeuropäische Haltung in der Bevölkerung zu fördern (ähnlich auch Immerfall und Sobisch 1997, S. 26). Da die Europäische Kommission als „Motor der europäischen Integration“ (Sturm 2010, S. 22) dem Integrationsprojekt allerdings nicht neutral gegenüber steht, ist sie nach Höpner und Jurczyk (2012, S. 327; 2015) daran interessiert, „den Integrationswillen der Bürgerinnen und Bürger so ausgeprägt wie nur möglich darstellen zu können“. Dieses Bestreben habe Konsequenzen für die Auswahl der Fragen, die Frageformulierungen und die Antwortkategorien. Nach einer Durchsicht von 93 Fragebögen äußern Höpner und Jurczyk (2012, S. 345) den *„Verdacht strategischer Manipulationen“* (Hervorhebung im Original), die allesamt darauf abzielen, „die Ergebnisse in eine pro-europäische, integrationsfreundliche Richtung zu lenken“. Dieser Kritik hält Bläser (2013, S. 354) entgegen, dass im Zeitraum 2005 bis 2012 mehr als 200 unterschiedliche Fragen zur „Europäischen Integration“ gestellt wurden, bei denen viele auch eine negative Polung aufweisen. Da das Eurobarometer großen Wert darauf lege, von der Öffentlichkeit als wissenschaftlich-methodisch ausgewiesenes Instrument ernst genommen zu werden, könne es nicht im Sinne der

Eurobarometer-Autoren sein, „falsche Umfrageergebnisse zu publizieren" (Bläser 2013, S. 355). Trotz der Diskussion um mögliche „Sonnenscheinfragen", wird das Eurobarometer auch künftig vielen wissenschaftlichen Studien als empirische Grundlage zur Beantwortung der jeweiligen Forschungsfragen dienen. Dies ist nicht überraschend, da sich das Eurobarometer erstens durch die große Anzahl der Länder, zweitens die Vielzahl an abgefragten Themen sowie drittens den langen Erhebungszeitraum für zahlreiche Fragestellungen eignet. Allerdings sollte sich der Forscher stets bewusst sein, dass es sich beim Eurobarometer nicht um ein akademisches Projekt handelt. Die Daten des Eurobarometers werden von der Europäischen Kommission für die sozialwissenschaftliche Forschung und Lehre zur Verfügung gestellt und bei GESIS im Datenarchiv für Sekundärauswertungen bereitgestellt (www.gesis.org/eurobarometer-data-service/home). Umfangreiche Informationen sowie Berichte zu den einzelnen Erhebungen finden sich auch auf der Homepage der Europäischen Kommission.

European Election Studies (EES) Die European Election Studies (ESS) sind ein akademisch initiiertes Projekt zur Analyse der Europawahlen, das von einem Zusammenschluss europäischer Wahlforscher – der European Election Study Group – getragen wird. Die EES umfassen dabei mehrere Teilstudien. Im Mittelpunkt stehen jeweils Wählerbefragungen, die 1979 und 1984 im Rahmen der Eurobarometer, seit 1989 im Rahmen eigener Wählerstudien durchgeführt werden. Darüber hinaus wurden für 1979 und seit 1999 Mediendaten gesammelt. Für die Wahljahre 1979, 1994 und 2009 liegen zudem Daten aus Kandidatenbefragungen vor. Seit 2004 gibt es darüber hinaus eine bis ins Jahr 1979 zurückreichende Sammlung und inhaltsanalytische Aufbereitung der Europawahlprogramme.

Zur Europawahl 2009 gelang es der European Election Study Group, eine umfangreiche Forschungsförderung für das Projekt „Providing an Infrastructure for Research on Electoral Democracy in the European Union" (PIREDEU) im siebten Forschungsrahmenprogramm der Europäischen Kommission einzuwerben. Die am Europäischen Hochschulinstitut in Florenz angesiedelte Studie umfasste die Befragung von Wählern und Kandidaten, Medieninhaltsanalysen, Wahlprogrammanalysen und die Sammlung relevanter Kontextdaten (unter anderem Wahlsysteminformationen und Wahlergebnisse). Das Hauptaugenmerk lag auf der inhaltlichen Abstimmung der einzelnen Studienteile. Schließlich erhöht sich durch gleiche oder verknüpfte Fragestellungen das (kombinierte) Analysepotenzial der Studienteile erheblich. Dies wird einerseits die inhaltlichen Fragestellungen homogenisieren und andererseits bestehende Forschungslücken schließen (Wüst und Tausendpfund 2009, S. 9). Auch für die Europawahl 2014 konnte eine Europawahlstudie realisiert werden, die sich aus fünf Komponenten zusammen-

setzt: Wählerbefragung, Medienanalyse, Wahlprogrammauswertung, Kandidatenbefragung sowie (erstmalig) eine Social Media Studie.

Informationen zu den European Election Studies sowie den einzelnen Komponenten finden sich auf der Projekthomepage (http://europeanelectionstudies.net). Die Daten sowie die Dokumentationen zu den Datensätzen werden über GESIS der interessierten Öffentlichkeit zur Verfügung gestellt.

European Social Survey (ESS) Der European Social Survey (ESS) ist eine europäische Studie der vergleichenden empirischen Sozialforschung, die seit 2002 alle zwei Jahre durchgeführt wird. Das Ziel des ESS besteht darin, die „Einstellungen, Wertorientierungen und Verhaltensmuster der Bevölkerungen in den europäischen Staaten zu beschreiben und zu erklären sowie Trends zu ermitteln" (Keil 2009, S. 431). An der ersten Erhebungswelle 2002/2003 haben sich 22 Länder beteiligt, bei der achten Welle im Jahr 2016 waren es 24 Länder. Allerdings haben nicht alle Länder an allen Erhebungen teilgenommen (Deutschland war bisher immer beteiligt).

Beim ESS wird ein in allen Ländern einheitlicher Fragebogen eingesetzt, der verschiedene Aspekte des politischen und gesellschaftlichen Zusammenlebens abdeckt. Der Fragebogen besteht aus sogenannten Kernmodulen, die in jeder Erhebung eingesetzt werden, und variablen Modulen, die es erlauben, spezifische Themen einzubinden. Die Kernmodule enthalten Fragen zu Mediennutzung, Politik, Vertrauen in Institutionen, Immigration, Angst vor Verbrechen, Gesundheit, Religion, Wohlbefinden, Wertorientierungen sowie zur Demographie. Die variablen Module befassten sich bisher unter anderem mit „Altersdiskriminierung und Einstellungen zum Wohlfahrtsstaat" (ESS 4), „Vertrauen in Strafjustiz und Polizei, Arbeit und Familie" (ESS 5) oder auch „Verständnis und Bewertungen der Demokratie" (ESS 6). Die variablen Module werden dabei jeweils über ein Wettbewerbsverfahren ausgewählt, an dem Wissenschaftler teilnehmen können (siehe für eine gelungene Darstellung z. B. Schnaudt et al. 2014; 2016).

Dem ESS wird eine außerordentlich hohe Datenqualität bescheinigt (Kohler 2008, S. 420). Danach hat der ESS im Vergleich zu Eurobarometer (EB), European Value Study (EVS) und dem International Social Survey Programme (ISSP) die beste Dokumentation, verwendet die besten Auswahlverfahren und zeigt die geringsten Abweichungen zwischen internen und externen Repräsentativitätskriterien (zur Stichprobenziehung des ESS siehe z. B. Häder et al. 2009). Der ESS ist die empirische Grundlage für zahlreiche nationale (z. B. van Deth 2004; Keil und van Deth 2012; Braun und Tausendpfund 2013) und internationale Veröffentlichungen (z. B. Meulemann 2008). Auf der Homepage zum ESS findet sich eine Bibliographie zu empirischen Studien, die mit Daten des ESS gearbeitet haben (www.europeansocialsurvey.org/bibliography).

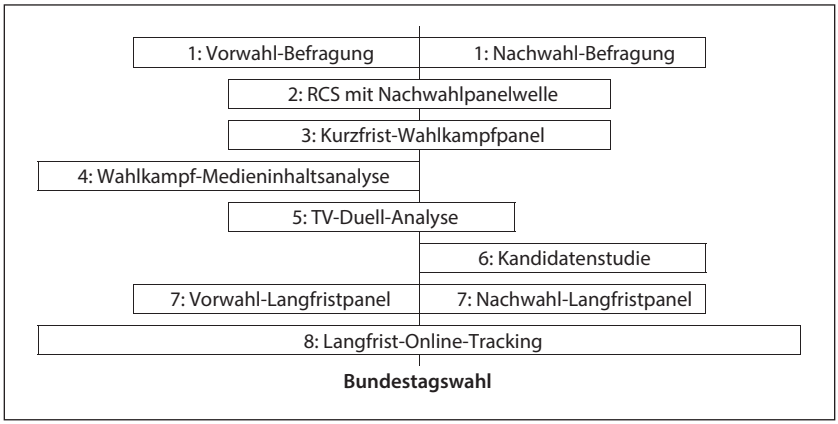
Die Daten des European Social Survey werden vom Norwegian Social Science Data Services (NSD) aufbereitet und interessierten Nutzern kostenlos über das Internet zur Verfügung gestellt (www.europeansocialsurvey.org). Neben der Möglichkeit zum kostenlosen Herunterladen der Daten stehen auch sämtliche Dokumente (z. B. Fragebögen, Richtlinien zur Stichprobenziehung) zur Verfügung. Der ESS ist damit eine der transparentesten Erhebungen weltweit.

German Longitudinal Election Study Wahlen sind nicht nur die zentralen Institutionen der repräsentativen Demokratie, sondern es handelt sich dabei auch – trotz rückläufiger Beteiligung – um die häufigste Form politischer Beteiligung (z. B. Gabriel 2012; Schmitt 2014). Angesichts dieser Bedeutung von Wahlen für die Demokratie ist es nicht überraschend, dass die Wahlforschung zu den theoretisch und methodisch am weitesten entwickelten Disziplinen der Sozialwissenschaften gehört. Allerdings existiert in Deutschland im Vergleich zu vielen anderen Ländern (z. B. USA, Großbritannien, Österreich) bisher keine nationale – auf Dauer finanzierte – wissenschaftliche Wahlstudie.[23] Die German Longitudinal Election Study (GLES) soll der Zwischenschritt auf dem Weg zu einer nationalen Wahlstudie sein.

Bei der GLES handelt es sich um ein von der Deutschen Forschungsgemeinschaft gefördertes Langfristvorhaben zur Analyse der Bundestagswahlen 2009, 2013 und 2017 (Schmitt-Beck et al. 2010, S. 142). Wie Abbildung 17 zeigt, umfasst das Forschungsprogramm der GLES dabei verschiedene, aufeinander abgestimmte Teilstudien. Das „Herzstück" der GLES-Erhebung ist eine umfangreiche persönlich-mündliche Befragung einer Zufallsstichprobe der Wahlberechtigten, die in eine Vorwahl- und Nachwahlwelle aufgeteilt ist. Diese Erhebungen umfassen eine große Anzahl an Merkmalen, um die Hintergründe der Wahlbeteiligung und Wahlentscheidung analysieren zu können (Komponente 1). Mehrere Komponenten im GLES-Forschungsprogramm zielen darauf ab, die kurzfristigen Dynamiken des Wahlprozesses zu analysieren. Dazu dient eine sogenannte Rolling-Cross-Section-Umfrage (Komponente 2), die es erlaubt, Veränderungen der öffentlichen Meinung als Reaktion auf Wahlkampagnen und Medienberichterstattung zu identifizieren (Hoops et al. 2013). Die Rolling-Cross-Section-Umfrage wird durch ein Wahlkampfpanel ergänzt, das durch eine Online-Erhebung realisiert wird (Komponente 3).

23 Es existiert aber für jede Bundestagswahl (mindestens) eine Wahlstudie, die im GESIS-Datenarchiv für Sekundäranalysen zur Verfügung steht. Allerdings handelt es sich um Einzelprojekte, sodass mangels langfristiger Kontinuität der Erhebungsinstrumente die kontinuierliche Beobachtung politischer Einstellungen und Verhaltensweisen nur schwierig zu realisieren ist.

Abbildung 17 GLES-Design im Überblick (vereinfachte Darstellung)



Quelle: http://gles.eu/wordpress/design/

Die Umfragenkomponenten werden durch Analysen des Wahlkontexts ergänzt. Eine Medieninhaltsanalyse der Berichterstattung konzentriert sich auf die wichtigsten Themen, Kandidatenprofile und mögliche Koalitionsszenarien (Komponente 4). Zudem umfasst das GLES-Design eine detaillierte Studie zu TV-Duellen (Komponente 5). Schließlich beinhaltet die GLES eine Kandidatenstudie, das heißt, alle Kandidaten der Bundestagswahl werden befragt (Komponente 6). Die Komponenten 7 und 8 zielen darauf ab, Langfristdynamiken zu beobachten und zu analysieren. Alle Umfragekomponenten sind durch einen identischen Kernfragenkatalog verbunden, der jeweils durch komponentenspezifische Fragen ergänzt wird. Dadurch wird das Analysepotenzial erheblich erweitert.

Die GLES ist kein Projekt eines Einzelforschers, sondern ein gemeinschaftliches Projekt der deutschen Wahlforschung. Die hierfür geschaffene organisatorische Basis ist die Deutsche Gesellschaft für Wahlforschung (DGfW). Informationen zur DGfW finden sich unter www.dgfw.eu; die GLES wird auf der Seite www.gles.eu vorgestellt. Die im Rahmen der GLES erhobenen Daten werden über GESIS kostenfrei zur Verfügung gestellt. Über das Forschungsdatenzentrum Wahlen von GESIS sind auch Datensätze zu früheren Bundestagswahlen sowie zu Landtagswahlen verfügbar. Auch die Datensätze der österreichischen nationalen Wahlstudie (www.autnes.at) sind bei GESIS über den Datenbestandskatalog erhältlich.

International Social Survey Programme Das International Social Survey Programme (ISSP) ist ein internationales Kooperationsprogramm, das 1984 von verschiedenen Forschungseinrichtungen aus Australien, Deutschland, Großbritannien und den USA gegründet wurde. Mittlerweile beteiligen sich 48 Länder am ISSP, das jährlich eine gemeinsame Umfrage zu sozialwissenschaftlich relevanten Themen durchführt. Jede Erhebung zeichnet sich durch einen thematischen Schwerpunkt aus, der nach einigen Jahren wiederholt wird (z. B. soziale Ungleichheit, Religion, Umwelt). Dadurch können Entwicklungen über die Zeit betrachtet werden.

Nach den ISSP-Working-Principles ist bei den jährlichen Erhebungen eine nationale repräsentative Zufallsstichprobe von 1400 Personen angestrebt (Minimum: 1000). Die einheitliche Befragungszeit ist mit 15 Minuten je Person relativ gering, sodass jeweils nur wenige Fragen gestellt werden können. Nach der Darstellung auf der GESIS-Projektseite sind vier Merkmale besonders zentral. Erstens ist die internationale Zusammenarbeit dauerhaft und kontinuierlich. Dadurch werde die international vergleichende Forschung zu einem festen Bestandteil der nationalen Forschungsprogramme der Mitgliedsländer. Zweitens basiere das Programm auf Freiwilligkeit und Kooperation. Wichtige Entscheidungen über Module oder Fragebögen, über Mitglieder oder deren Verpflichtungen werden gemeinsam durch Abstimmung in der ISSP-General-Assembly gefällt. Drittens finanziere jedes Mitgliedsland seine eigene Umfrage. Gemeinsame ISSP-Ressourcen, etwa zur Koordination des Programms, zur Entwicklung der englischsprachigen Basisfragebögen oder für methodische Forschung, existieren nicht. Viertens sei das ISSP eines der wenigen internationalen Umfrageprogramme, das study monitoring durchführt und Berichte zu den jährlichen Umfragen veröffentlicht.

Weiterführende Informationen zum ISSP finden sich online unter www.issp.org sowie bei GESIS (www.gesis.org/issp/home/). Die Daten (einschließlich Fragebögen) werden von GESIS archiviert und sind über den Datenbestandskatalog kostenlos verfügbar.

Mikrozensus Der Mikrozensus ist eine Haushaltsbefragung der amtlichen Statistik in Deutschland. Rund 830 000 Personen in etwa 370 000 privaten Haushalten und Gemeinschaftsunterkünften werden stellvertretend für die gesamte Bevölkerung zu ihren Lebensbedingungen befragt. Dies ist etwa ein Prozent der Bevölkerung, das nach einem festgelegten statistischen Zufallsverfahren ausgewählt wird. Die Befragung findet in der Regel mündlich statt und bei den meisten Fragen besteht Auskunftspflicht.

Der Mikrozensus stellt Daten zur Bevölkerungsstruktur sowie zur wirtschaftlichen und sozialen Lage der Bevölkerung bereit. Mit Informationen zu Familie und Lebenspartnerschaft, Arbeitsmarkt und Erwerbstätigkeit sowie Beruf und

Mikrodaten-Informationssystem

Das Mikrodaten-Informationssystem MISSY stellt datenrelevante Informationen, die für die Analyse von amtlichen Mikrodaten unabdingbar sind, in systematischer Weise online zur Verfügung. Dazu sammelt MISSY ein breites Spektrum relevanter Metadaten an zentraler Stelle. Diese reichen von globalen Informationen zu den Studien, beispielsweise dem Datenzugang oder Einzelheiten der Erhebungen in den teilnehmenden Ländern, bis hin zu Detailinformationen auf Variablenebene wie beispielsweise Häufigkeitsverteilungen, Filterangaben oder Fragetexte. Das Projekt MISSY wurde vom Bundesministerium für Bildung und Forschung (BMBF) finanziert und im German Microdata Lab der GESIS in Zusammenarbeit mit dem Bereich Wissenstechnologien für Sozialwissenschaften umgesetzt. Für weitere Informationen siehe www.gesis.org/missy.

Ausbildung hat sich der Mikrozensus zu einer wichtigen Datenquelle entwickelt. Genutzt werden die Statistiken von Verantwortlichen aus Parlamenten und Verwaltung, von der Wissenschaft wie auch der breiten Öffentlichkeit.

Seit dem Bundesstatistikgesetz von 1987 stehen anonymisierte Individualdaten der amtlichen Statistik durch die statistischen Ämter als sogenannte faktisch anonymisierte Daten der Wissenschaft zur Verfügung. Die Datenverfügbarkeit ist allerdings eingeschränkt (Schimpl-Neimanns und Weiss 2014). Die Mikrozensus-Scientific-Use-File ist eine faktisch anonymisierte 70-Prozent-Substichprobe des Original-Mikrozensus. Für die faktisch anonymisierte Substichprobe gilt, dass sie als systematische Zufallsauswahl aus dem Originalmaterial des Statistischen Bundesamtes gezogen wurde. Informationen zum Mikrozensus und zur Datenverfügbarkeit finden sich unter: www.forschungsdatenzentrum.de/bestand/mikrozensus/index.asp.

pairfam Das 2008 gestartete Beziehungs- und Familienpanel pairfam (Panel Analysis of Intimate Relationships and Family Dynamics) ist eine multidisziplinäre Längsschnittstudie zur Erforschung der partnerschaftlichen und familialen Lebensformen in Deutschland. Die jährlich erhobenen Befragungsdaten von über 12 000 bundesweit zufällig ausgewählten Personen der Geburtsjahrgänge 1971 bis 1973, 1981 bis 1983 und 1991 bis 1993 sowie von deren Partnern, Eltern und Kindern bieten ein einmaliges Analysepotenzial bezüglich der Entwicklung von Partnerschafts- und Generationenbeziehungen in unterschiedlichen Lebensphasen (Huinink et al. 2011).

Die dokumentierten Daten des Beziehungs- und Familienpanels stehen der Fachöffentlichkeit für wissenschaftliche Analysen zur Verfügung. Auf der Projektseite unter www.pairfam.de finden sich umfangreiche Informationen zu dieser soziologischen Studie sowie die Projektdaten.

PIAAC Bei dem Programme for the International Assessment of Adult Competencies (PIAAC) handelt es sich ähnlich wie bei PISA um eine von der Organisation für wirtschaftliche Zusammenarbeit und Entwicklung (OECD) geförderte und international vergleichende Studie zur Erfassung von Kompetenzen mit dem Fokus auf der erwachsenen Bevölkerung im Alter von 16 bis 65 Jahren. Dabei wurden 2011/2012 in 24 Ländern grundlegende Kompetenzen untersucht, die notwendig sind, um den alltäglichen und beruflichen Anforderungen erfolgreich begegnen zu können. Hierzu gehören Lesekompetenz, alltagsmathematische Kompetenz und technologiebasiertes Problemlösen. Ergänzend werden soziale und ökonomische Merkmale der befragten Personen erhoben (z. B. die individuelle Bildungs- und Berufsbiographie). Auf dem PIAAC-Webportal von GESIS – Leibniz-Institut für Sozialwissenschaften stehen Berichte, Dokumentationen und die Scientific-Use-File für Deutschland zum Herunterladen bereit: www.gesis.org/piaac/. Bei Rammstedt und Maehler (2014) findet sich eine Übersicht zur PIAAC-Studie.

PISA Sind die Jugendlichen auf die Herausforderungen der Zukunft gut vorbereitet? Können sie ihre Ideen und Vorstellungen effektiv analysieren, begründen und auch kommunizieren? Haben sie die notwendigen Kompetenzen für ein lebenslanges Lernen? Diese Fragen soll das Programme for International Student Assessment (PISA) beantworten. PISA ist eine vergleichende Studie, die Basiskompetenzen der nachwachsenden Generation erfassen soll und von der Organisation für wirtschaftliche Zusammenarbeit und Entwicklung (OECD) durchgeführt wird. Laut der OECD werden mit PISA die Basiskompetenzen erfasst, „die in modernen Gesellschaften für eine befriedigende Lebensführung in persönlicher und wirtschaftlicher Hinsicht sowie für eine aktive Teilnahme am gesellschaftlichen Leben notwendig sind“ (Baumert et al. 2001, S. 16; siehe zu Schulleistungsstudien auch Sälzer 2016).

Im Mittelpunkt der Schulleistungsstudie stehen die Kompetenzen der Schüler in Mathematik, Lesen und Naturwissenschaften. Erstmals wurde die PISA-Studie im Jahr 2000 in 32 Ländern (darunter 28 OECD-Mitgliedsstaaten) durchgeführt, danach erfolgten die Erhebungen in einem Dreijahreszyklus. In jedem Zyklus wird ein Kompetenzbereich gründlicher getestet, dem dann zwei Drittel der Testzeit zugeteilt werden, während in den beiden anderen Bereichen jeweils nur zusammenfassende Leistungsprofile ermittelt werden. Umfangreiche Informationen

zu PISA finden sich unter www.oecd.org/pisa/home/. Die Datensätze stehen ebenfalls auf der OECD-Seite zur Verfügung: www.oecd.org/pisa/pisaproducts/.

SHARE Wie altern Menschen in den Ländern der Europäischen Union? Diese Frage kann mit Daten des internationalen Umfrageprojekts „Survey of Health, Ageing and Retirement in Europe" (SHARE) beantwortet werden. 2004 wurde erstmals eine sogenannte repräsentative Befragung der Bevölkerung ab 50 Jahren in elf europäischen Ländern durchgeführt. Mittlerweile sind sechs Befragungswellen abgeschlossen und in 20 verschiedenen Ländern wurden mehr als 290 000 Interviews durchgeführt. Bei SHARE handelt es sich um eine Panelstudie. Das bedeutet, dass die gleichen Personen alle zwei Jahre befragt werden. Dadurch können intraindividuelle Veränderungen untersucht werden, beispielsweise Effekte von Reformen oder Lebensereignissen. Neben Fragen zu Familie, Freunde, Freizeit und Gesundheit werden bei SHARE auch körperliche und medizinische Tests durchgeführt (z. B. Greifkraft in den Händen oder Merkfähigkeit). Auf Grundlage der SHARE-Daten konnte beispielsweise ein positiver Effekt von sportlichen Betätigungen auf die Gesundheit nachgewiesen werden (de Souto Barreto et al. 2017). Weitere Informationen zu SHARE finden sich unter www.share-project.org.

Sozio-ökonomische Panel (SOEP) Das Sozio-ökonomische Panel (SOEP) ist eine Wiederholungsbefragung privater Haushalte in Deutschland, die im jährlichen Rhythmus seit 1984 bei denselben Personen und Familien in der Bundesrepublik Deutschland durchgeführt wird. Bereits im Juni 1990, also noch vor der Währungs-, Wirtschafts- und Sozialunion, wurde die Studie auf das Gebiet der ehemaligen DDR ausgeweitet. Zur breiteren Erfassung gesellschaftlichen Wandels wurde in den Jahren 1994/95 zudem eine „Zuwanderer-Stichprobe" realisiert (Schupp 2009, S. 352). Im Gegensatz zur ALLBUS werden beim SOEP jedes Jahr die gleichen Personen eines Haushalts befragt. Es handelt sich folglich um eine Paneleerhebung (zur Erinnerung: Beim ALLBUS werden alle zwei Jahre unterschiedliche Personen befragt. Deshalb handelt es sich um eine sogenannte Querschnittserhebung).

Das SOEP wurde gegründet, „um sozial- und wirtschaftswissenschaftliche Fragestellungen im Rahmen von Haushalten, Lebensläufen und lebenslaufbezogenem Verhalten analysieren zu können" (Wagner et al. 2008, S. 304). Die jährliche Erhebung besteht deshalb aus einem teils gleichen, teils wechselnden Fragepro-gramm. Zum Kern des Erhebungsprogramms gehören Fragen aus den folgenden Bereichen (Wagner et al. 2008, S. 305): Demographie und Wohnsituation, Persönlichkeitsmerkmale und Grundorientierungen (z. B. Präferenzen und Werte), Vorschul- und Schulbildung, berufliche Bildung und Weiterbildung, Qualifikation, Arbeitsmarkt- und Berufsmobilität, Einkommen, Vermögen und soziale Sicher-

heit, Gesundheit sowie Sorgen und Zufriedenheit (allgemeine Lebenszufriedenheit und Bereichszufriedenheiten).

Das SOEP wird vom Deutschen Institut für Wirtschaftsforschung (DIW) in Berlin verantwortet. Auf der Homepage (www.diw.de/de/soep) finden sich umfangreiche Informationen zur Geschichte, Methode und Forschungspraxis des SOEP. Die anonymisierten SOEP-Daten können für wissenschaftliche Forschungszwecke genutzt werden; die Weitergabe der Daten erfolgt über ein in der Abteilung SOEP eingerichtetes Forschungsdatenzentrum. Voraussetzung für die Weitergabe ist ein (kostenloser) Datenweitergabevertrag. Bedingt durch die Panel-Struktur ist die Analyse mit SOEP-Daten allerdings deutlich komplexer als etwa mit den Daten der ALLBUS. Deshalb sind fortgeschrittene statistische Kenntnisse etwa zur Paneldatenanalyse (z. B. Giesselmann und Windzio 2012; Brüderl 2010) sowie eines Statistikprogramms (SPSS, Stata) erforderlich, um mit den SOEP-Daten angemessen arbeiten zu können.

World Values Survey (WVS) Bei der World Values Survey (WVS) handelt es sich vermutlich um die derzeit umfangreichste und weiträumigste Datenerhebung. Die WVS geht auf die Werte-Wandel-Studie von Ronald Inglehart (1977, 1995) zurück und beinhaltet deshalb insbesondere Fragen zu Werten und Einstellungen. Die Erhebung deckt Themen wie traditionelle Werte, Einstellungen zu Familie, Arbeitsleben, Wirtschaft, Umwelt, Religion und Moral sowie Identität ab. Zusätzlich werden auch zentrale demographische Informationen wie Geschlecht, Alter und Schulbildung erfasst. In den teilnehmenden Ländern werden jeweils zwischen 1000 und 3000 Personen über 18 Jahren persönlich befragt (Keil 2009, S. 435).

Die World Values Surveys sind aus den European Values Surveys (EVS) hervorgegangen, die in die WVS integriert worden sind. Mittlerweile existieren sechs Erhebungen der WVS. Die erste Befragung wurde Anfang der 1980er Jahre durchgeführt. Die zweite Welle wurde Anfang der 1990er Jahre, die dritte Welle Mitte der 1990er Jahre und die vierte Welle Ende der 1990er Jahre realisiert. Die fünfte Welle folgte 2005 bis 2009, die sechste Welle 2010 bis 2014. Eine siebte Welle ist aktuell in Planung.

Informationen zur WVS finden sich online unter www.worldvaluessurvey.org. Dort finden sich auch die Datensätze, die nach (kostenloser) Registrierung für die Datenauswertung zur Verfügung stehen. Eng mit der WVS ist die EVS verbunden, eine europäische Studie über die Vorstellungen der Bürger zu Familie, Arbeit, Religion, Politik und Gesellschaft. Informationen zu der EVS finden sich unter www.europeanvaluesstudy.eu; die Daten werden von GESIS archiviert und können über den Datenbestandskatalog heruntergeladen werden.

8.3.2 Makroebene

Bei Makrodaten handelt es sich um (verdichtete) Informationen über eine konkrete Analyseeinheit (z. B. Arbeitslosenquote der lokalen, regionalen oder nationalen Ebene). Solche Daten sind häufig an Einrichtungen bzw. Institutionen gebunden (z. B. Eurostat, Statistisches Bundesamt), die die Daten entweder selbst erheben oder sie langfristig für die interessierte Öffentlichkeit bereitstellen.

Wikipedia ist keine Datengrundlage

Auch bei Wikipedia (www.wikipedia.org) finden sich zahlreiche Makrodaten. Die freie Enzyklopädie ist aber keine geeignete Quelle für eine wissenschaftliche Arbeit, da die intersubjektive Nachvollziehbarkeit der einzelnen Informationen nicht gewährleistet ist. Für wissenschaftliche Arbeiten sind stets die Original- bzw. Primärdaten zu verwenden. Dabei ist eine klare und eindeutige Quellenangabe obligatorisch, um die intersubjektive Nachvollziehbarkeit zu gewährleisten.

Neben diesen (staatlichen) Institutionen gibt es eine Reihe ambitionierter Forschungsprojekte, die ebenfalls Makrodaten für die wissenschaftliche Gemeinschaft zur Verfügung stellen. Wichtige Institutionen und Forschungsprojekte werden im Folgenden vorgestellt. Eine Übersicht wichtiger Makrodaten der Sozialwissenschaften ist auch auf der Homepage der Arbeitsstelle „Quantitative Methoden“ verfügbar: http://e.feu.de/makrodaten.

Bundesagentur für Arbeit Die Bundesagentur für Arbeit bietet Zeitreihen zur deutschen Arbeitsmarkt- und zur Beschäftigungsstatistik an (http://statistik.arbeitsagentur.de). Die Statistiken sind regional gegliedert (Bund, Länder, Kreise). Informativ sind auch die Grundlagen der Statistik, in der zentrale Begriffe wie „Beschäftigungsquote“, „Hilfequote“ oder auch „Saisonbereinigung“ dargestellt werden. Das Institut für Arbeitsmarkt- und Berufsforschung (IAB) ist die Forschungseinrichtung der Bundesagentur für Arbeit. Auf der Homepage des IAB (www.iab.de) finden sich umfangreiche Informationen zu aktuellen Befragungen sowie Publikationen.

CIA – The World Factbook Das CIA World Factbook (http://e.feu.de/cia) bündelt zahlreiche Makrodaten über einzelne Länder. Es handelt sich um ein kompaktes Online-Nachschlagewerk, das Informationen zu Geographie, Bevölkerung, Regierung, Wirtschaft und Energie sowie Kommunikation, Transport und Militär bereithält.

Comparative Manifesto Project (CMP) Das Comparative Manifesto Project (CMP) sammelt Wahlprogramme aus über 50 Ländern und wertet diese inhaltsanalytisch aus. Das CMP ist am Wissenschaftszentrum Berlin für Sozialforschung (WZB) angesiedelt und untersucht, ob, inwieweit und unter welchen Bedingungen Parteien programmatisch auf ökonomische, soziale, kulturelle und internationale Herausforderungen reagieren. Dazu werden die Wahlprogramme politischer Parteien in mehr als 50 Ländern für alle freien demokratischen Wahlen seit 1945 in quantitativen Inhaltsanalysen ausgewertet. Ziel des Projekts ist es, die politischen, wirtschaftlichen und gesellschaftlichen Positionen politischer Parteien im internationalen und im zeitlichen Vergleich zu messen. Bislang standen Herausforderungen an etablierte Demokratien der OECD-Welt einerseits und an die neuen Demokratien in Mittel- und Osteuropa andererseits im Fokus der Forschung. Die Datensätze sind über die Projekthomepage (http://manifestoproject.wzb.eu) frei verfügbar. Dort finden sich auch umfangreiche Informationen zur Codierung der Wahlprogramme.

Eurostat Das Statistische Amt der Europäischen Gemeinschaften (Eurostat) stellt Daten und Informationen der EU-Mitgliedstaaten zusammen und bereitet diese auf, sodass harmonisierte (vergleichbare) Daten auf europäischer Ebene zur Verfügung stehen. Die Homepage von Eurostat (http://epp.eurostat.ec.europa.eu) bietet einen direkten Zugriff auf eine große Auswahl von Daten und Indikatoren, unter anderem zu Bevölkerung, Bruttosozialprodukt, Arbeitslosenquote, Erwerbstätigenquote oder auch zur Inflation. Über eine Datenbank können individuelle Tabellen generiert und abgespeichert werden.

Freedom House Freedom House (https://freedomhouse.org) ist eine Forschungseinrichtung mit dem Ziel, die Freiheit weltweit zu fördern. Seit 1972 werden jährlich Berichte über den Stand der politischen Rechte (political rights) sowie der Bürgerrechte (civil liberties) in 193 Ländern bzw. Gebieten der Welt veröffentlicht. Freiheit, das heißt die Existenz politischer Rechte sowie die Gewährleistung bestimmter Bürgerrechte, wird über eine Checkliste, die 25 Fragen beinhaltet, auf einer Skala von 1 bis 7 gemessen (wobei 1 den höchsten und 7 den niedrigsten Freiheitsgrad anzeigt). Die Bewertung der beiden Kategorien erfolgt jeweils über Länderexperten. Der Freedom-House-Index misst explizit nicht die Qualität der

Demokratie, sondern den Grad politischer Rechte und bürgerlicher Freiheiten. Allerdings werden die Angaben von Freedom House durchaus als Indikatoren für die Demokratiequalität genutzt. Dies wird mittlerweile allerdings kritisch bewertet (siehe z. B. Müller und Pickel 2007; Lauth 2008, S. 37–38).

OECD Die Organisation for Economic Co-operation and Development (OECD) ist einer der größten und wichtigsten Produzenten internationaler Makrodaten. Die Organisation sammelt Daten, bereitet diese auf und veröffentlicht die Daten und Berichte (www.oecd.org). Im Datenzentrum der OECD (https://data.oecd.org/) wird eine sehr große Bandbreite an Indikatoren und Statistiken veröffentlicht, die als Excel-Daten abgespeichert werden können. Die OECD bietet darüber hinaus zahlreiche eigene Publikationen an, in denen die umfangreichen statistischen Informationen verbal aufbereitet sind (www.oecd-ilibrary.org).

Parliament and Government Composition Database (ParlGov) Die Parliament and Government Composition Database (ParlGov) ist ein Projekt von Holger Döring und Philip Manov (Döring 2013). ParlGov ist eine Dateninfrastruktur, die versucht, die Mängel zur Erhebung von Daten über politische Institutionen zu überwinden. Mit ParlGov stellen die Initiatoren Informationen über Parteipositionen, Wahlergebnisse und Regierungszusammensetzungen bereit. Derzeit enthält ParlGov Daten über Wahlen und Regierungen für alle EU-Staaten und die meisten OECD-Mitglieder für die gesamte Nachkriegszeit. Die Datenbank verbindet rund 1400 Parteien, 680 Wahlen mit 5800 Wahlergebnissen sowie 2300 Kabinette in 960 Regierungen. Die Informationen werden in Tabellen aufbereitet und stehen über die Homepage (www.parlgov.org) als Download zur Verfügung.

Polity IV Polity IV ist ein Angebot des Center for Systemic Peace, das jährliche Informationen über Regime- und Autoritätscharakteristika für 167 Staaten zur Verfügung stellt. Zur Bestimmung der Qualität eines Regimes werden drei Merkmale berücksichtigt, die in vier bis sieben Stufen differenziert werden: „Wettbewerbsgrad der politischen Partizipation", „Wettbewerbsgrad und die Offenheit der politischen Rekrutierung" sowie „Begrenzung der Exekutive" (Lauth 2008, S. 37). Die Daten sind im Internet verfügbar (www.systemicpeace.org/polity/polity4x.htm); derzeit wird das Polity Project weiterentwickelt (www.systemicpeace.org/polityproject.html).

Quality of Government Basic Dataset Das Quality of Government Institute der Universität von Göteborg stellt einen Datensatz bereit, der verschiedene Indikatoren der Qualität einer Regierung bündelt. Der QoG-Basic-Dataset beinhaltet keine eigenen erhobenen Daten, sondern setzt sich aus zahlreichen Daten ver-

schiedener Anbieter zusammen. Der umfangreiche Datensatz enthält Angaben von 193 Ländern und beinhaltet Informationen zu Korruption, Bürokratie, Wahlregeln, Fragmentierung oder auch Gleichberechtigung, Vertrauen und Zufriedenheit der Bürger mit der Demokratie. Auf der Homepage des Instituts www.qog.pol.gu.se findet sich eine umfangreiche Dokumentation zum Datensatz.

Statista Das kommerzielle Angebot Statista (http://de.statista.com) bietet – zum Teil auch kostenlos – Zugang zu umfangreichen Datensätzen von Markt- und Meinungsforschungsinstitutionen sowie aus Wirtschaft und amtlicher Statistik.

Statistisches Bundesamt und statistische Landesämter Das Statistische Bundesamt ist der führende Anbieter amtlicher statistischer Informationen in Deutschland (www.destatis.de). Die 2400 Beschäftigten an den Standorten Wiesbaden, Bonn und Berlin erheben, sammeln, systematisieren, analysieren und präsentieren statistische Informationen. Dabei deckt die Behörde eine große Bandbreite an Themen ab – von Daten zu Gebiet und Bevölkerung über einzelne Wirtschaftsbereiche bis hin zu Preisentwicklung, öffentlichen Finanzen und volkswirtschaftlichen Gesamtrechnungen. Über die GENESIS-Onlinedatenbank des Statistischen Bundesamts lassen sich die Informationen in Tabellen ausgeben und können gespeichert werden. Zusätzlich finden sich auf den Seiten des Statistischen Bundesamts auch zahlreiche Erläuterungen zu Statistiken und Begriffen. Informationen zum Zensus 2011, der Volkszählung der statistischen Ämter, finden sich auf einer eigenen Internetseite (www.zensus2011.de).

Der Präsident des Statistischen Bundesamts ist auch der Bundeswahlleiter und für die ordnungsgemäße Durchführung der Bundestagswahlen und der Europawahlen (in Deutschland) zuständig. Auf der Seite www.bundeswahlleiter.de finden sich daher auch die amtlichen Ergebnisse bisheriger Wahlen. Neben dem Statistischen Bundesamt gibt es in Deutschland statistische Landesämter, die spezifische Daten auf Landesebene sammeln, aufbereiten und der Öffentlichkeit zur Verfügung stellen (z. B. Wahlen, Erwerbstätigkeit). In größeren Gemeinden gibt es zudem eigene kommunale Ämter, die lokale Daten sammeln, aufbereiten und analysieren. Solche statistischen Daten können für die Forschung äußerst wertvoll sein. Einige Beispiele: Schäfer (2012) nutzt lokale Daten, um den Einfluss der sinkenden Wahlbeteiligung auf das Wahlergebnis zu untersuchen. Für die Analyse des Einflusses des lokalen Kontexts auf individuelle Einstellungen und Verhaltensweisen verwenden van Deth und Tausendpfund (2013) lokale Daten aus 28 hessischen Gemeinden.

Vereinte Nationen Auch die Vereinten Nationen bieten umfangreiche statistische Informationen, die von der United Nations Statistics Division aufbereitet und zur

Verfügung gestellt werden. In einer Meta-Datenbank (http://data.un.org) werden 34 unterschiedliche Datenbanken zu den verschiedensten Themenbereichen ausgewertet (z. B. Bildung, Gesundheit, Entwicklung). Bei den Vereinten Nationen finden sich auch Informationen zu den sogenannten Millennium-Entwicklungszielen (http://mdgs.un.org/unsd/mdg/). Die Daten, Indikatoren und Dokumente können auf der Homepage heruntergeladen werden. Eine weitere bekannte Indikatorenreihe ist der Human Development Index (HDI), der seit 1990 mit einer Maßzahl den Stand der menschlichen Entwicklung in den Ländern der Welt verdeutlichen will. Der Index setzt sich dabei aus drei Dimensionen zusammen: Lebenserwartung bei der Geburt, Bildungsniveau und das Pro-Kopf-Einkommen als Indikator für den Lebensstandard. Die Daten stehen auf der Homepage zum Download zur Verfügung (http://hdr.undp.org/en/data).

Weltbank Die Weltbank (www.worldbank.org/) wurde im Juli 1944 auf der Währungs- und Finanzkonferenz der Vereinten Nationen gegründet. Sie ist eine Sonderorganisation der Vereinten Nationen. Ursprünglich war ihr Ziel, nach dem Zweiten Weltkrieg den Wiederaufbau zu fördern und in Zusammenarbeit mit dem Internationalen Währungsfonds (IWF) stabile Währungen zu schaffen. Seit den 1960er Jahren ist es ihre Hauptaufgabe, die Armut in der Welt zu bekämpfen und die Lebensbedingungen der Menschen in den Entwicklungsländern zu verbessern. Die Weltbank sammelt, systematisiert und publiziert eine Vielzahl an Daten – angefangen vom Klimawandel über Armut bis hin zu sozialer Entwicklung. Darüber hinaus finden sich bei der Weltbank zahlreiche Publikationen (www.worldbank.org/reference/).

8.4 Recherche nach sozialwissenschaftlichen Datensätzen

Mittlerweile steht eine enorme Anzahl an sozialwissenschaftlichen Datensätzen für die interessierte akademische Öffentlichkeit zur Verfügung. Mit Blick auf die Verfügbarkeit von Datensätzen sieht die Welt der empirischen Sozialforschung heute sicherlich „besser“ aus als noch vor einem Jahrzehnt (zur Entwicklung der Datengrundlage siehe z. B. Mochmann 2014). Die grundsätzlich positive Entwicklung wird allerdings durch ein eher unübersichtliches Datenangebot getrübt. Im deutschsprachigen Raum ist sicherlich GESIS – Leibniz-Institut für Sozialwissenschaften die erste (institutionelle) Anlaufstelle für die Suche nach geeigneten Datensätzen (www.gesis.org). Mit der Registrierungsagentur für Sozial- und Wirtschaftsdaten, die von GESIS und der Deutschen Zentralbibliothek für Wirtschaftswissenschaften (ZBW) betrieben wird, steht eine weitere Plattform für die Recherche nach Sozial- und Wirtschaftsdaten bereit (www.da-ra.de). Schließlich

bietet auch der Rat für Sozial- und Wirtschaftsdaten (www.ratswd.de) ein umfangreiches Informationsangebot rund um Forschungsdaten. Die drei Institutionen – GESIS, ZBW und RatSWD – haben auch einen Wegweiser publiziert, der nützliche Tipps und Hinweise bei der Suche nach Forschungsdaten enthält.

Wie bei der Literaturrecherche gibt es auch bei der Recherche nach Datensätze nicht die „eine" erfolgversprechende Recherchestrategie. Vielmehr sollten verschiedene Suchstrategien für eine erfolgreiche Datensuche kombiniert werden. Mit Watteler (2017, S. 148–151) lassen sich dabei drei Suchstrategien unterscheiden: die institutionelle Suche, die Suche über Forscher und die Suche über Inhalte.

Drei allgemeine Suchstrategien Bei der institutionellen Suche sollten zunächst einmal Einrichtungen identifiziert werden, die sich mit dem Forschungsgegenstand beschäftigen. Mittlerweile existiert in Deutschland eine ausdifferenzierte Forschungslandschaft mit einer Vielzahl an Einrichtungen und Hochschulen. Wer sich beispielsweise inhaltlich mit dem Arbeitsmarkt beschäftigt, der wird sicherlich das Institut für Arbeitsmarkt- und Berufsforschung kennen. Das Deutsche Jugendinstitut in München bündelt beispielsweise zahlreiche Forschungsprojekte aus dem Bereich der Jugendforschung. Hinweise auf wirtschaftliche Datensätze finden sich unter anderem beim Zentrum für Europäische Wirtschaftsforschung (ZEW). Solche Institute können als Wegweiser auf der Suche nach geeigneten Datensätzen genutzt werden. Bei der Suche über Forscher sind Publikationen dieser Forscher interessant. Bei empirischen Studien sollte die Datenquelle sichtbar ge-

Wie finde ich relevante Forschungsdaten?

Antworten auf diese Fragen bietet die Ratgeberbroschüre „Auffinden-Zitieren-Dokumentieren", die GESIS – Leibniz-Institut für Sozialwissenschaften, ZBW – Leibniz Informationszentrum Wirtschaft und der Rat für Sozial- und Wirtschaftsdaten (RatSWD) veröffentlicht haben. In dieser Broschüre finden sich grundlegende Informationen zum Finden, Dokumentieren und Archivieren quantitativer Forschungsdaten in den Sozial- und Wirtschaftswissenschaften. Die Broschüre ist unter der Internetadresse www.auffinden-zitieren-dokumentieren.de verfügbar. Informationen zu Datensätzen bietet auch die Zeitschrift „Wirtschafts- und Sozialwissenschaften" (Schmollers Jahrbuch). Die Sektion „European Data Watch" enthält Darstellungen und Diskussionen von Datenquellen, die insbesondere für empirisch arbeitende Sozialwissenschaftlerinnen von Interesse sind. Eine Übersicht findet sich unter: www.ratswd.de/publikationen/european-data-watch.

kennzeichnet sein. Mit der Bezeichnung kann nun gezielt nach den Daten gesucht werden. Durch die Veröffentlichung von ZA-Nummer oder DOI-Angaben wird die Suche sehr vereinfacht, da beide Angaben der Identifikation von Datensätzen dienen. Bei der Suche nach Inhalten bietet sich direkt der GESIS-Datenbestandskatalog an. Über die erweiterte Suche können Daten gezielt nach einzelnen Suchbegriffen recherchiert, das Erhebungsjahr spezifiziert oder auch Forscher gesucht werden.

Watteler (2017) gibt weitere Tipps und Tricks zur effizienten Datenrecherche. Zudem finden sich bei den Institutionen auch Ansprechpartner, die bei der Suche und Auswahl von Datensätzen weiterhelfen können. Und sicherlich haben auch Betreuer von Haus- und Abschlussarbeiten wertvolle Hinweise, die bei der Suche berücksichtigt werden können. Hinweise zur Datenrecherche sind auch auf der Homepage der Arbeitsstelle „Quantitative Methoden" verfügbar: http://e.feu.de/datenrecherche.

Forschungsdesign 9

Das Forschungsdesign – häufig auch Untersuchungsanordnung genannt – umfasst im weiten Verständnis die Beantwortung mehrerer Fragen: Wann, wo, wie und wie oft müssen die empirischen Indikatoren an welchen Objekten erfasst werden, um die formulierten Hypothesen prüfen zu können (z. B. Westle 2009c, S. 133; ähnlich auch Schnell et al. 2013, S. 199; Stein 2014, S. 138)? In Anlehnung an Diekmann (2011, S. 194) lassen sich bei einem engeren Verständnis des Forschungsdesigns drei zentrale Entscheidungen unterscheiden: Erstens muss in Abhängigkeit von den formulierten Hypothesen die Untersuchungsebene festgelegt werden. Zweitens muss die Untersuchungsform geklärt werden. Drittens muss die Häufigkeit der Datenerhebung spezifiziert werden. Diese drei Entscheidungen werden teilweise nacheinander, teilweise aber auch zeitlich parallel und in einem wechselseitigen Zusammenhang getroffen. Die Festlegung des Forschungsdesigns hat dabei auch Konsequenzen für das Auswahlverfahren (Kapitel 10) und die Datenerhebung (Kapitel 11).

9.1 Untersuchungsebene

Die Formulierung von Hypothesen beinhaltet immer auch eine Entscheidung über die Untersuchungsebene. Mit der Individual-, der Kollektiv- und der Kontexthypothese werden drei Hypothesenarten unterschieden (siehe Kapitel 6). Mit der jeweiligen Hypothesenart ist eine spezifische Untersuchungsebene verbunden: Bei Individualhypothesen ist die Mikroebene die Untersuchungsebene. Beide Merkmale der Hypothese – also uV und aV – sind der Individualebene zuzuordnen. Bei Kollektivhypothesen ist die Makroebene die Untersuchungsebene. Häufig werden Untersuchungen auf der Makroebene auch Aggregatdatenanalysen genannt. Beide Merkmale einer Kollektivhypothese sind der Kollektiv- bzw. Aggregatebene zu-

© Springer Fachmedien Wiesbaden GmbH, ein Teil von Springer Nature 2018

M. Tausendpfund, *Quantitative Methoden in der Politikwissenschaft*,
Grundwissen Politik, https://doi.org/10.1007/978-3-658-20698-7_9

zuordnen. Bei Kontexthypothesen wird ein Einfluss des Kontexts auf individuelle Einstellungen und/oder Verhaltensweisen postuliert. Bei einer Kontexthypothese ist die abhängige Variable der Mikroebene (Individualebene) und die unabhängige Variable der Makroebene (Kontextebene) zuzuordnen. Die empirische Überprüfung von Kontexthypothesen ist besonders anspruchsvoll, da empirische Daten der Individual- und der Kontextebene erforderlich sind. In Kapitel 6 wurden die mit der jeweiligen Untersuchungsebene verbundene Forschungsstrategie – Individualdatenanalyse, Aggregatdatenanalyse und Kontextanalyse – dargestellt. Zum besseren Verständnis werden die einzelnen Forschungsstrategien in den folgenden Abschnitten anhand eines gemeinsamen Beispiels illustriert. Als Forschungsthema wird die Wahlbeteiligung bei Europawahlen gewählt. Die Forschungsfrage lautet: Welche Faktoren fördern die Beteiligung an den Europawahlen?

9.1.1 Individualdatenanalyse

Die Beteiligung an einer Wahl ist – auf den ersten Blick betrachtet – eine individuelle Entscheidung: Eine Person beteiligt sich an einer Wahl oder eine Person beteiligt sich nicht an einer Wahl. Die abhängige Variable stellt die individuelle Wahlbeteiligung dar. Die für eine empirische Untersuchung auf der Individualebene notwendige Information, ob sich eine Person an einer Wahl beteiligt, wird in der Regel durch eine Befragung gewonnen. In der Forschungspraxis lassen sich dabei zwei Fragevarianten unterscheiden. Die erste Variante wird Wahlbeteiligungsabsicht genannt. Dabei werden Personen vor einer Wahl gefragt, ob sie sich an der kommenden Wahl beteiligen werden. Die zweite Variante wird als berichtete Wahlbeteiligung bezeichnet. Dabei werden Personen nach der Wahl gefragt, ob sie sich an der vergangenen Wahl beteiligt haben. Bereits der Zeitpunkt der Befragung – vor oder nach der Wahl – hat empirische Konsequenzen. In der Regel ist der Anteil der Wähler bei der berichteten Wahlbeteiligung geringer als bei der Wahlbeteiligungsabsicht.[24]

24 Die Wahlforschung ist mit zahlreichen methodischen Herausforderungen konfrontiert. Bei beiden Varianten zur Erfassung der Wahlbeteiligung lässt sich ein Overreporting der Wahlbeteiligung feststellen (Zeglovits und Kritzinger 2014; Hanmer et al. 2014; Sciarini und Goldberg 2016). Mit Overreporting wird das Phänomen bezeichnet, das die durch Umfragen ermittelte Wahlbeteiligung regelmäßig höher ausfällt als die offizielle (amtliche) Wahlbeteiligung. Zur Erklärung lassen sich zwei Ansätze unterscheiden: Einerseits geben einige Befragte lediglich vor, zu wählen, obgleich sie dies tatsächlich nicht tun. Andererseits beteiligen sich Personen, die nicht zu erreichen sind oder ein Interview verweigern, seltener an Wahlen (Caballero 2014, S. 442–443).

Zusammenhänge auf Individualebene prüfen Bei einer Individualdatenanalyse werden jetzt – auf Basis formulierter und mit dem Forschungsstand verknüpfter Hypothesen – andere Individualmerkmale gesucht, die mit der abhängigen Variable (hier: die Wahlbeteiligung) in Beziehung stehen. Ein solches Individualmerkmal ist beispielsweise das politische Interesse. Folgende Hypothese wäre plausibel: Je stärker das politische Interesse, desto größer die Wahrscheinlichkeit der Wahlbeteiligung.[25]

Allerdings steht bei einer Individualdatenanalyse nicht das einzelne Individuum, sondern der vermutete Zusammenhang zwischen zwei Individualmerkmalen im Mittelpunkt. Auf Basis einer größeren Datenerhebung (bei Europawahlen: European Election Study) wird untersucht, ob sich der Zusammenhang zwischen den beiden Individualmerkmalen „Politisches Interesse" und „Wahlbeteiligung" bestätigt oder eben nicht. Nicht das einzelne Individuum wird betrachtet, sondern der allgemeine Zusammenhang zwischen zwei Merkmalen.[26] Sollte sich bei einer empirischen Untersuchung die Hypothese bestätigen, dann bedeutet dies, dass die Wahrscheinlichkeit einer Wahlbeteiligung bei stärkerem politischem Interesse höher ist als bei schwächeren politischen Interesse. Probabilistische Hypothesen können durch konträre Einzelfälle nicht widerlegt werden.

Mehrere Hypothesen testen Selbstverständlich ist das politische Interesse nur ein Merkmal, das einen Einfluss auf die Wahlbeteiligung entfaltet. Die Wahlforschung hat mittlerweile unzählige Individualmerkmale identifiziert, die einen (mehr oder weniger) starken Einfluss auf die individuelle Chance der Wahlbeteiligung haben: Parteiidentifikation, Wahlnorm, Bildung, Alter oder auch die Zufriedenheit mit der Demokratie (siehe z. B. Lamers und Roßteutscher 2014; Steinbrecher und Rattinger 2011; Caballero 2014). Entsprechend wird in einem Forschungsprojekt zur Wahlbeteiligung nicht nur ein Faktor untersucht bzw. eine Hypothese überprüft, sondern es werden mehrere bzw. möglichst alle relevanten Individualmerkmale der Wahlbeteiligung identifiziert und empirisch überprüft. Auf Basis statistischer Analyseverfahren (hier: logistische Regression) können schließlich Aussagen über die tatsächliche Relevanz und Effektstärke der einzel-

25 In einem Forschungsprojekt müsste die Hypothese selbstverständlich noch theoretisch erläutert werden. Warum sollte das politische Interesse die Wahrscheinlichkeit der Wahlbeteiligung fördern? Welche Argumente sprechen für eine positive Beziehung zwischen politischem Interesse und Wahlbeteiligung? Zudem müssten empirische Befunde dargestellt werden, die die Plausibilität der Hypothese unterstreichen.

26 Deshalb handelt es sich bei der Hypothese „Je stärker das politische Interesse, desto größer die Wahrscheinlichkeit der Wahlbeteiligung" auch um eine sogenannte probabilistische Hypothese.

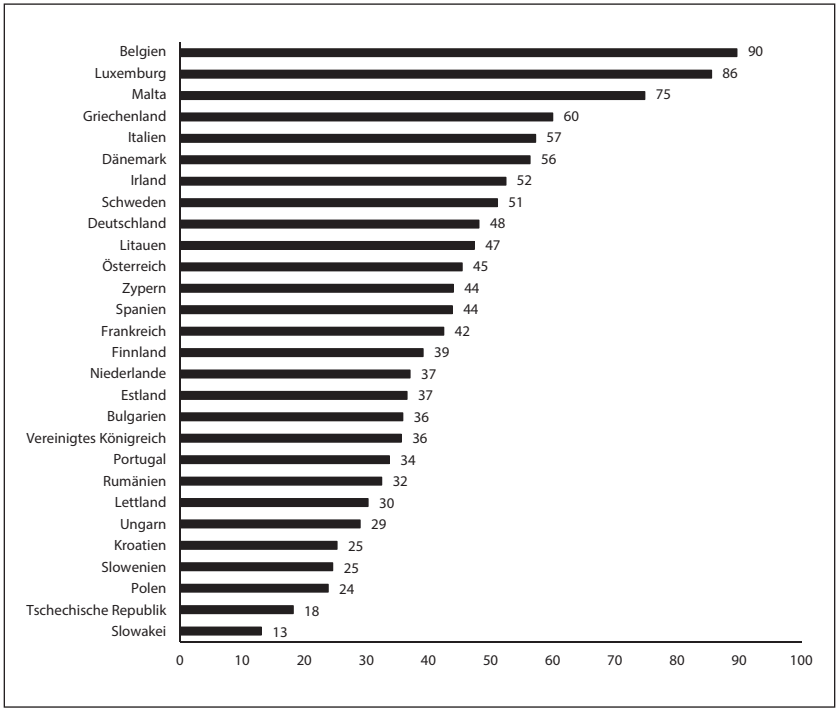
nen Merkmale getroffen werden (zur logistischen Regression siehe z. B. Behnke 2015; Best und Wolf 2010).

9.1.2 Aggregatdatenanalyse

Bei der Aggregatdatenanalyse werden Zusammenhänge auf der Kollektiv- bzw. Aggregatebene betrachtet. Mit anderen Worten: Während bei einer Individualdatenanalyse Unterschiede zwischen Individuen erklärt werden, werden bei einer Aggregatdatenanalyse Unterschiede zwischen Kollektiven bzw. Aggregaten erklärt.

Abbildung 18 zeigt die Wahlbeteiligung bei der Europawahl 2014 im Ländervergleich. Auf den ersten Blick sind erhebliche Unterschiede zwischen den einzel-

Abbildung 18 Wahlbeteiligung bei der Europawahl 2014 (Angaben in Prozent)



Quelle: Eigene Darstellung auf Basis der Angaben des Europäischen Parlaments (www.europarl.europa.eu/elections2014-results/de/turnout.html; 15. 11. 2017)

nen EU-Mitgliedsstaaten zu beobachten. In Belgien beteiligten sich fast 90 Prozent der Wahlberechtigten am Urnengang; in der Slowakei waren es gerade einmal 13 Prozent. Bei einer Aggregatdatenanalyse werden jetzt Merkmale gesucht, die die Unterschiede zwischen den Ländern erklären können.

Analog zur Individualdatenanalyse werden auch bei der Aggregatdatenanalyse auf Basis theoretischer Überlegungen Hypothesen formuliert, die bei einer empirischen Analyse geprüft werden. Dabei werden Merkmale der Aggregatebene mit anderen Merkmalen der Aggregatebene in Beziehung gesetzt. Diese Hypothesen werden Kollektivhypothesen genannt.

Mittlerweile liegen zur Erklärung der nationalen Unterschiede der Wahlbeteiligung bei Europawahlen umfangreiche Forschungsergebnisse vor (z. B. Franklin 2001; Schmitt 2005; zur Erklärung von Länderunterschieden bei nationalen Hauptwahlen siehe z. B. Freitag 1996; Geys 2006). Neben der Wahlpflicht können unter anderem der Wahltag, der Wahlzyklus oder auch die Möglichkeit der Briefwahl die Unterschiede zwischen den Ländern erklären. Mit Hilfe statistischer Analyseverfahren können auch bei der Aggregatdatenanalyse Aussagen über die tatsächliche Relevanz und Bedeutung der einzelnen Merkmale getroffen werden.

! Für viele Fragestellungen sind Aggregatdatenanalysen von großer Bedeutung. Bei historischen (wahlsoziologischen) Fragestellungen (z. B. Falter 1991; Debus 2016) sind Aggregatdatenanalysen zudem die einzige Forschungsstrategie, da Individualdaten nicht existieren. Aber auch für die Erklärung lokaler Unterschiede in der Wahlbeteiligung sind Aggregatdaten durchaus nützlich (z. B. Vetter 2008; Schäfer et al. 2012). Allerdings können sie keine Auskunft darüber geben, warum eine Person wählt und eine andere Person nicht wählt. Daher besteht bei Aggregatdatenanalysen immer auch die Gefahr eines ökologischen Fehlschlusses.

9.1.3 Kontextanalyse

Bei einer Kontextanalyse befindet sich die abhängige Variable (hier: Wahlbeteiligung) auf der Individualebene. Die unabhängigen Variablen verteilen sich allerdings auf die Individual- und die Kontextebene. Der Begriff „Kontext“ bezieht sich dabei auf das Umfeld des Befragten. Bei Wahlanalysen ist mit Kontext häufig das nationalstaatliche Umfeld gemeint. Allerdings kann sich der Kontext auch auf die Region (z. B. Wahlkreise) oder auch die Gemeindeebene beziehen.

Abbildung 19 verdeutlicht die Logik einer Kontextanalyse (häufig auch Mehrebenenanalyse bezeichnet). Pfeil 1 in Abbildung 19 postuliert einen Zusammenhang zwischen dem politischen Interesse und der Wahlbeteiligung. Dies entspricht einer klassischen Individualhypothese: Je stärker das politische Interesse,

desto größer die Wahrscheinlichkeit der Wahlbeteiligung. Mit der Kontextanalyse verbindet sich allerdings die (theoretische) Vorstellung, dass individuelle Einstellungen und Verhaltensweisen nicht ausschließlich auf andere Individualmerkmale zurückzuführen sind.

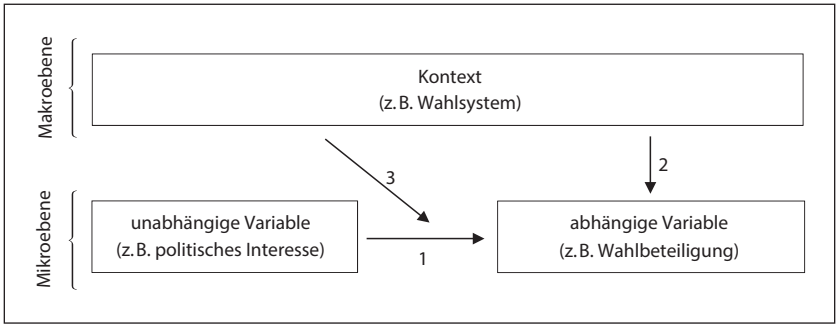
> *„Die Kontextanalyse ist jene Variante der empirischen Sozialforschung, in der systematisch auch die soziale und personale Umgebung in die Variablen aufgenommen wird, mit denen ein bestimmtes Verhalten erklärt werden soll." (Esser 1999a, S. 428)*

Was ist ein struktureller Effekt? Danach ist die Wahlbeteiligung einer Person nicht nur auf andere Individualmerkmale zurückzuführen (z. B. Parteiidentifikation, Wahlnorm), sondern auch von Merkmalen des Umfelds bzw. Kontexts abhängig. Pfeil 2 in Abbildung 19 stellt den Einfluss des Kontexts auf die individuelle Wahlbeteiligung dar. Ein Einfluss des Kontexts unabhängig von den Effekten der Individualmerkmale wird allgemein als struktureller Effekt bezeichnet.

> *„Ein Kontexteffekt (ein ‚struktureller Effekt' also) liegt dann vor, wenn sich bei Kontrolle der unabhängigen individuellen Variablen noch ein statistisch merklicher Effekt der Kontext-Variablen feststellen läßt." (Esser 1999a, S. 429)*

Darüber hinaus kann der Kontext bzw. ein Merkmal des Kontexts auch einen Effekt auf den Zusammenhang zwischen der unabhängigen und der abhängigen Variable auf der Individualebene haben. In Abbildung 19 wird dies durch Pfeil 3 gekennzeichnet. In der Sprache der empirischen Sozialforschung wird dies als Cross-Level-Interaktion bezeichnet.

Abbildung 19 Kontextanalyse am Beispiel der Wahlbeteiligung



Quelle: Eigene Darstellung

Sorgfältig durchgeführte Kontextanalysen sind äußerst anspruchsvoll. Zum einen sind empirische Informationen auf der Mikroebene (Individualebene) und der Makroebene (Kontextebene) erforderlich, zum anderen sind insbesondere die Kontexthypothesen theoretisch plausibel zu begründen (eine Einführung bieten z. B. Braun et al. 2010). Mittlerweile liegen Studien vor, die die individuellen und kontextuellen Faktoren der Wahlbeteiligung empirisch untersuchen (zur Europawahl siehe z. B. Clark 2014; zu nationalen Hauptwahlen siehe z. B. Hadjar und Beck 2010; Bühlmann und Freitag 2006).

9.1.4 Zusammenfassung

Bereits bei der Formulierung der Hypothesen ist eine Entscheidung des Forschungsdesigns gefallen. Dies verdeutlicht, dass die einzelnen Phasen eines Forschungsprozesses nicht unabhängig voneinander zu betrachten sind. Entscheidungen bei der Entwicklung der Forschungsfrage, der Konzeptspezifikation oder der Operationalisierung haben einen Einfluss auf das Forschungsdesign.

Die Festlegung der Untersuchungsebene dient daher auch der Plausibilitätsprüfung. Können abhängige und unabhängige Variable einer Hypothese explizit einer Untersuchungsebene zugeordnet werden? Welche Analysestrategie liegt den formulierten Hypothesen zugrunde? Können den Konzepten einer Hypothese beobachtbare Indikatoren zugeordnet werden? Diese Fragen sind nach der Festlegung der Untersuchungsebene zu beantworten.

9.2 Untersuchungsform

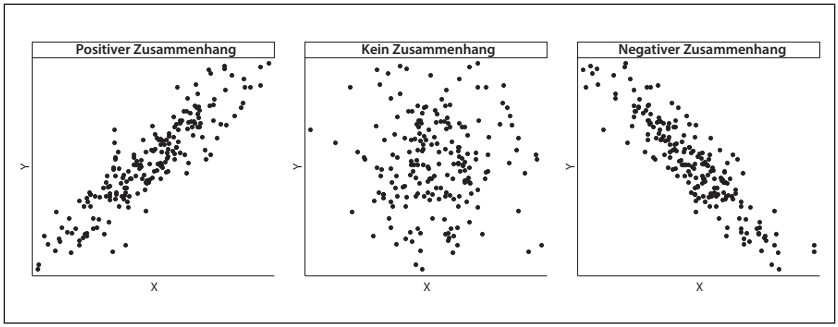
Im Rahmen des Forschungsdesigns muss festgelegt werden, mit welcher Untersuchungsform die aus den theoretischen Überlegungen hergeleiteten Hypothesen überprüft werden sollen. In den Sozialwissenschaften wird mit einer Vielzahl empirischer Untersuchungsformen gearbeitet. Jede dieser Untersuchungsformen umfasst bestimmte Vorgehensweisen, die festlegen, wie empirische Informationen gesammelt und wie die Hypothesen anschließend empirisch überprüft werden. Die Vielzahl der Untersuchungsformen lässt sich auf zwei Grundtypen zurückzuführen: Experimente und nicht-experimentelle Untersuchungsformen. Die nicht-experimentellen Untersuchungsformen werden in der Literatur auch als Ex-post-facto-Designs bezeichnet. Die beiden Grundtypen der Untersuchungsform werden in diesem Abschnitt vorgestellt. Vorab wird der Unterschied zwischen Korrelation und Kausalität erläutert, um die Vor- und Nachteile der beiden Grundtypen besser einordnen zu können.

9.2.1 Korrelation und Kausalität

In Hypothesen werden vermutete Zusammenhänge zwischen zwei Merkmalen postuliert. Typische Beispiele für Hypothesen sind etwa „Je höher die Bildung, desto größer das Einkommen“ oder „Je stärker das politische Interesse, desto größer die Wahrscheinlichkeit der Wahlbeteiligung“. Sozialwissenschaftler sind allerdings weniger an einfachen Zusammenhängen zwischen Merkmalen interessiert, sondern vor allem an Ursache-Wirkungs-Zusammenhängen: Ist eine höhere Bildung die „Ursache“ für ein größeres Einkommen? Ist ein stärkeres politisches Interesse die „Ursache“ für eine größere Wahrscheinlichkeit der Wahlbeteiligung? In Hypothesen sind die vermuteten kausalen Beziehungen zwar sprachlich angelegt (Wenn-dann bzw. Je-desto), aber der empirische Nachweis eines Zusammenhangs zwischen einer vermuteten Ursache (uV) und einer Wirkung (aV) ist kein ausreichender Beleg für einen kausalen Effekt einer unabhängigen Variable auf eine abhängige Variable (Bortz und Döring 2006, S. 11). Es ist zwischen Korrelation und Kausalität zu unterscheiden (Kühnel und Dingelstedt 2014).

Was ist eine Korrelation? Eine Korrelation beschreibt den Zusammenhang zwischen zwei Merkmalen, das heißt, ob die Ausprägung eines Merkmals mit der Ausprägung eines anderen Merkmals korrespondiert. Eine positive Korrelation bedeutet, dass hohe Werte eines Merkmals mit hohen Werten eines anderen Merkmals auftreten. Von einer negativen Korrelation wird gesprochen, wenn zu einem hohen Wert eines Merkmals tendenziell ein niedriger Wert des anderen Merkmals gehört. In Abbildung 20 sind verschiedene Arten von Korrelationen dargestellt.

Abbildung 20 Korrelationen



Quelle: Eigene Darstellung

Korrelation ist nicht Kausalität Der Korrelationskoeffizient ist ein Maß für den Grad des linearen Zusammenhangs zwischen zwei Merkmalen; er kann Werte zwischen –1 und +1 annehmen. Werte mit einem positiven Vorzeichen deuten auf einen positiven Zusammenhang und Werte mit einem negativen Vorzeichen auf einen negativen Zusammenhang hin. Eine Korrelation von +1 bzw. –1 bedeutet, dass zwei Merkmale perfekt korrelieren. In der empirischen Praxis sind solche perfekten Korrelationen selten. Werte ± 0 deuten auf keinen Zusammenhang zwischen den beiden Merkmalen hin. Der Korrelationskoeffizient gibt aber nur den mehr oder weniger starken Zusammenhang von zwei Merkmalen an und ist „ausschließlich ein Maß für Gleichklang in den Daten" (Krämer 1994, S. 144). Auf Basis des Korrelationskoeffizienten kann keine Aussage über den Ursache-Wirkungs-Zusammenhang getroffen werden. Eine unterhaltsame Möglichkeit, ein Gespür für Korrelationen zu entwickeln, bietet das Spiel „Guess the Correlation": http://guessthecorrelation.com.

Korrelation und Kausalität

Der Unterschied zwischen Korrelation (Zusammenhang) und Kausalität (Ursache-Wirkungs-Zusammenhang) lässt sich an einem klassischen Beispiel illustrieren. Empirisch lässt sich eine positive Korrelation zwischen der Anzahl der Feuerwehrleute bei Einsätzen und der Höhe des Sachschadens nachweisen: Je mehr Feuerwehrleute, desto höher der Sachschaden. Diese positive Korrelation ist aber keine kausale Beziehung zwischen der Anzahl der Feuerwehrkräfte und dem Sachschaden. Die Feuerwehrkräfte sind vermutlich nicht die Ursache für die Höhe des Sachschadens, sondern die Größe des Feuers. Bei einem größeren Feuer sind einerseits mehr Feuerwehrleute im Einsatz und andererseits ein höherer Schaden wahrscheinlich. Die Ursache für die Anzahl der Einsatzkräfte und die Höhe des Sachschadens ist die Größe des Feuers. Bei Krämer (2009) finden sich weitere unterhaltsame Beispiele, die den Unterschied zwischen Korrelation und Kausalität anschaulich erläutern (siehe auch Bauer et al. 2014).

In sozialwissenschaftlichen Studien sind korrelative Zusammenhänge und kausale Beziehungen allerdings nicht so einfach unterscheidbar wie im oben dargestellten Beispiel. In Anlehnung an Hill (1965) nennen Schnell et al. (2013, S. 203) unter anderem folgende Bedingungen, die erfüllt sein müssen, damit eine Beziehung

zwischen einer uV (Ursache) und einer aV (Wirkung) als kausal interpretiert werden kann:

- Die Ursache sollte der Wirkung zeitlich vorausgehen.
- Der Zusammenhang zwischen zwei Merkmalen sollte stark sein.
- Der Zusammenhang zwischen zwei Merkmalen sollte sich in verschiedenen Untersuchungskontexten an unterschiedlichen Populationen zeigen.
- Ein theoretisches Argument (ein plausibler Mechanismus), der dem Kausalzusammenhang zugrunde liegt, sollte genannt werden können.

Die Prüfung dieser Bedingungen und damit der Nachweis kausaler Beziehungen sind keineswegs trivial. Die beiden Grundtypen der Untersuchungsform – Experimente und Ex-post-facto-Designs – unterscheiden sich insbesondere im Grad der Zuverlässigkeit, kausale Beziehungen identifizieren zu können. Das Experiment ist die beste Methode für den Nachweis von Ursache-Wirkungs-Zusammenhängen (z. B. Aronson et al. 2004, S. 43; Sedlmeier und Renkewitz 2008, S. 125; Kühnel und Dingelstedt 2014, S. 1026). Deshalb gilt die experimentelle Methode als das „Untersuchungsdesign zur Analyse eines kausalen Zusammenhangs" (Behnke et al. 2010, S. 73) und als „Idealtyp empirischer Forschung" (Behnke et al. 2010, S. 52). Ex-post-facto-Designs eignen sich nur eingeschränkt für den Nachweis kausaler Beziehungen. Da bei vielen Fragestellungen allerdings keine Experimente möglich sind, stellen Ex-post-facto-Designs die dominierende Untersuchungsform in Soziologie, Politik- und Verwaltungswissenschaft dar.

9.2.2 Experiment

Ausgangspunkt eines jeden Experiments ist eine angenommene Ursache-Wirkungs-Beziehung zwischen einer unabhängigen Variable (uV) und einer abhängigen Variable (aV). Bei einem Experiment wird die unabhängige Variable (die Ursache) kontrolliert verändert, um anschließend mögliche Effekte auf die abhängige Variable (die Wirkung) zu beobachten. Diese kontrollierte Veränderung wird als Treatment oder Stimulus bezeichnet. Wenn sich nach der Änderung der uV eine Veränderung der aV beobachten lässt, dann ist diese Veränderung (wahrscheinlich) eine kausale Ursache der unabhängigen Variable.

Wie lässt sich die Grundidee experimenteller Forschung in der Praxis umsetzen? Ein Beispiel: In den Medien wird häufiger über einen Zusammenhang zwischen gewalttätigen Videospielen und aggressivem Verhalten spekuliert. Diese angenommene Beziehung lässt sich in eine sozialwissenschaftliche Hypothese übersetzen: „Je größer der Konsum von Gewaltspielen, desto stärker das aggres-

Was ist ein Treatment?

Der Begriff „Treatment“ oder „Stimulus“ ist weit gefasst: In Abhängigkeit von der Fragestellung kann es sich unter anderem um eine schriftliche Information, ein Video, Musik oder auch die Anwesenheit einer anderen Person handeln. Ein Treatment stellt in einem Experiment die unabhängige Variable dar. Einigen Personen wird das Treatment verabreicht (Experimentalgruppe), anderen Personen nicht (Kontrollgruppe). Nach der Verabreichung des Treatments wird die interessierende Reaktion (die aV) in beiden Gruppen erfasst und verglichen. Existiert ein Kausalzusammenhang zwischen der uV und der aV, dann werden sich die Reaktionen der beiden Gruppen unterscheiden.

sive Verhalten.“ Bei dieser Hypothese ist der Konsum von Gewaltspielen die unabhängige Variable, die einen Einfluss auf die abhängige Variable (hier: aggressives Verhalten) hat. Bei einem Experiment wird jetzt die uV manipuliert (hier: Konsum von Gewaltspielen). Die Teilnehmer an einem Experiment (die sogenannten Versuchspersonen bzw. Probanden) werden in ein Labor eingeladen. Ein Teil der Versuchspersonen (per Zufall bestimmt) wird dem Treatment ausgesetzt, das heißt, die Personen (die Experimentalgruppe) spielen für einen bestimmten Zeitraum im Labor ein gewalttätiges Computerspiel. Der andere Teil der Teilnehmenden (die Kontrollgruppe) spielt entweder gar kein Computerspiel oder eines mit nicht-gewalttätigem Inhalt. Danach wird bei allen Versuchspersonen das aggressive Verhalten (die aV) erfasst (z. B. durch Beobachtung oder durch Befragung). Sollte bei den Personen, die zuvor Gewaltspiele konsumiert haben, ein höheres aggressives Verhalten beobachtbar sein als bei den Personen, die keine Gewaltspiele gespielt haben, dann ist dieser Unterschied (wahrscheinlich) auf den Konsum der gewalttätigen Videospiele zurückzuführen.

Kontrolle durch Randomisierung Der Nachweis einer kausalen Beziehung ist nur möglich, wenn sich Experimental- und Kontrollgruppe nur im Hinblick auf das Treatment – hier: Konsum von Gewaltspielen – unterscheiden. Schließlich können auch andere Faktoren (z. B. Geschlecht, Alter, Bildung) einen Einfluss auf aggressives Verhalten haben. Möglicherweise sind männliche Jugendliche aggressiver als weibliche Jugendliche. Wenn sich nun der Geschlechteranteil zwischen Experimental- und Kontrollgruppe unterscheidet (z. B. mehr männliche Jugendliche in der Experimental- als in der Kontrollgruppe), dann sind Unterschiede zwischen Experimental- und Kontrollgruppe möglicherweise auf die ungleiche

Geschlechterverteilung und nicht etwa auf den Konsum von Gewaltspielen zurückzuführen. Deshalb müssen bei einem Experiment die Experimental- und die Kontrollgruppe sehr ähnlich sein und dürfen sich in nichts außer dem Treatment unterscheiden. Bei einfachen (und vor allem bekannten) Faktoren wie Geschlecht oder Alter ist das sicherlich unproblematisch. Allerdings existieren möglicherweise auch Faktoren, die einen Einfluss auf das aggressive Verhalten haben, die zum Zeitpunkt der Untersuchung überhaupt noch nicht bekannt sind. Wie lässt sich sicherstellen, dass es zwischen der Experimental- und der Kontrollgruppe keinen systematischen Unterschied gibt? Die Antwort für dieses schwierige Problem heißt Randomisierung (Fisher 1951).

Mit Randomisierung ist die zufällige Verteilung der Versuchspersonen auf Experimental- und Kontrollgruppe gemeint. Die Zuweisung in Experimental- oder Kontrollgruppe basiert also nicht auf Basis soziologischer, psychologischer oder biologischer Kriterien, sondern ausschließlich auf einem Zufallsprozess.[27] Diese Zufallsauswahl gewährleistet, dass sich die Verschiedenheiten der Teilnehmer in Geschlecht, Alter und anderen Faktoren gleichmäßig über die verschiedenen Bedingungen verteilen. Systematische Unterschiede in der Gruppenzusammensetzung sind damit unwahrscheinlich. Dadurch ist der Einfluss anderer (vor allem unbekannter) Faktoren auf die abhängige Variable in beiden Gruppen gleich und der in der Experimentalgruppe beobachtbare Effekt des Treatments auf die abhängige Variable ist mit großer Sicherheit auf den Stimulus zurückzuführen.

In der Literatur werden drei klassische experimentelle Designs unterschieden: Erstens die Nachher-Untersuchung mit Kontrollgruppe, zweitens die Vorher-Nachher-Untersuchung mit Kontrollgruppe und drittens das Solomon-Vier-Gruppen-Design (Behnke et al. 2010, S. 61–64; siehe auch Eifler 2014, S. 204). Bei

Was ist ein Experiment?

Nach Aronson et al. (2004, S. 43) ist das Experiment „eine Methode, in welcher der Forscher auf Zufallsbasis Teilnehmern Versuchsbedingungen zuteilt und sicherstellt, dass diese Situationen identisch sind außer der unabhängigen Variable (die Variable, von der angenommen wird, dass sie eine kausale Wirkung auf die Reaktionen der Menschen hat)“.

27 Die Randomisierung bei einem Experiment darf nicht mit der Zufallsauswahl bei Stichproben verwechselt werden. Die Randomisierung bei Experimenten dient der Kontrolle, die Zufallsauswahl bei Stichproben der Gewährleistung von „Repräsentativität“ (Behnke et al. 2010, S. 77–78; siehe auch Kapitel 10).

allen drei Formen ist die zufällige Verteilung der Personen in Experimental- und Kontrollgruppe obligatorisch.

Nachher-Untersuchung Das oben beschriebene Beispiel zur Untersuchung eines möglichen Zusammenhangs zwischen dem Konsum von Gewaltspielen und aggressivem Verhalten entspricht einer Nachher-Untersuchung mit Kontrollgruppe (siehe Abbildung 21). Die Teilnehmer werden zufällig der Experimental- oder der Kontrollgruppe zugewiesen. Während die Teilnehmer in der Experimentalgruppe mit Gewaltspielen konfrontiert werden, erhalten die Personen in der Kontrollgruppe nicht dieses Treatment. Die Verabreichung des Treatments wird in Abbildung 21 mit einem X angezeigt. Einige Zeit nach dem Konsum von Gewaltspielen wird das aggressive Verhalten erfasst. In der empirischen Praxis füllen die Personen in Experimental- und Kontrollgruppe meist einen entsprechenden Fragebogen aus. Alternativ könnte auch das Verhalten der Personen beobachtet werden. In Abbildung 21 ist die Erfassung der abhängigen Variablen (hier: aggressives Verhalten) mit einem O gekennzeichnet.

Abbildung 21 Nachher-Untersuchung mit Kontrollgruppe

Randomisierung	Gruppe	Zeit1	Zeit2
Teilnehmer werden zufällig einer der zwei Gruppen zugewiesen	Experimentalgruppe	X	O
	Kontrollgruppe		O

Anmerkung: O = Erhebung der aV. X = Treatment.

Quelle: Eigene Darstellung

Bei der Analyse wird untersucht, ob sich das aggressive Verhalten in der Experimentalgruppe von dem in der Kontrollgruppe unterscheidet. Hierzu werden die Mittelwerte der Experimental- und der Kontrollgruppe verglichen. Statistische Verfahren erlauben es zu prüfen, ob ein vorhandener Unterschied eine zufällige Schwankung darstellt oder systematisch ist.

Bei der einfachen Nachher-Untersuchung mit Kontrollgruppe werden die beobachteten Unterschiede der abhängigen Variable auf das Treatment zurückgeführt. Diese Schlussfolgerung ist zulässig, wenn für den Unterschied zwischen Experimental- und Kontrollgruppe keine andere Ursache außer dem Treatment vorliegt. Die Randomisierung gewährleistet die zufällige Verteilung der Personen in Kontroll- und Experimentalgruppe. Dadurch sollten sich Experimental- und Kontrollgruppe nur zufällig unterscheiden.

> *„Wenn die Gruppengröße jedoch klein ist, dann können diese zufällig auftretenden Abweichungen nicht mehr vernachlässigbar gering sein und die beobachteten Unterschiede zwischen Experimentalgruppe und Kontrollgruppe sind möglicherweise darauf zurückzuführen, dass beide Gruppen vor der Verabreichung des Stimulus verschieden waren." (Behnke et al. 2010, S. 62)*

Im gewählten Beispiel könnten bereits vor dem Konsum von Gewaltspielen in der Experimentalgruppe mehr aggressive Personen gewesen sein. Der möglicherweise beobachtete Effekt wäre dann nicht etwa durch das Treatment verursacht worden, sondern lediglich durch die ungleiche Zusammensetzung der beiden Gruppen. Um solche Verzerrungen überprüfen zu können, werden die Teilnehmer bereits vor dem Treatment nach ihrem aggressiven Verhalten befragt. Diese Vorgehensweise entspricht der Vorher-Nachher-Untersuchung mit Kontrollgruppe.

Vorher-Nachher-Untersuchung In Abbildung 22 ist der klassische Versuchsplan mit Vorher-Nachher-Untersuchung und einer Kontrollgruppe dargestellt. Erneut werden die Teilnehmer zufällig der Experimental- oder der Kontrollgruppe zugewiesen. Bei dieser Vorgehensweise wird vor der Verabreichung des Treatments bei allen Teilnehmern das aggressive Verhalten erfasst. Bei einer erfolgreichen Randomisierung sollte sich das durchschnittliche Aggressivitätsniveau nicht zwischen Experimental- und Kontrollgruppe unterscheiden. Etwas technischer formuliert: O_1 sollte mit O_2 vergleichbar sein.

Nach der Vorher-Messung wird der Experimentalgruppe das Treatment verabreicht, während die Teilnehmer der Kontrollgruppe kein Treatment erhalten. Danach wird erneut das aggressive Verhalten der Teilnehmer erfasst. Bei den Teilnehmenden der Kontrollgruppe sollte kein Unterschied der abhängigen Variable feststellbar sein, das heißt O_2 sollte mit O_4 vergleichbar sein. In der Experimentalgruppe müsste bei einer korrekten Hypothese das aggressive Verhalten O_3 (deutlich) stärker ausgeprägt sein als bei O_1. Die Differenz zwischen O_1 und O_3 kann

Abbildung 22 Vorher-Nachher-Untersuchung mit Kontrollgruppe

Randomisierung	Gruppe	Zeit1	Zeit2	Zeit3
Teilnehmer werden zufällig einer der zwei Gruppen zugewiesen	Experimentalgruppe	O_1	X	O_3
	Kontrollgruppe	O_2		O_4

Anmerkung: O = Erhebung der aV. X = Treatment.

Quelle: Eigene Darstellung

dann als kausaler Effekt angesehen und auf die Verabreichung des Treatments zurückgeführt werden.

Die Vorher-Nachher-Untersuchung mit Kontrollgruppe schließt viele Alternativerklärungen aus. Allerdings könnte bei dem gewählten Design das aggressive Verhalten nicht allein durch das Treatment verursacht worden sein, sondern durch eine Wechselwirkung zwischen Treatment und der vorherigen Messung. Erst durch die Kombination von Messung und Konsum von Gewaltspielen ist ein Anstieg des aggressiven Verhaltens der Teilnehmer zu beobachten. Zum Ausschluss einer solchen Alternativerklärung sind zusätzlich zum Design der Vorher-Nachher-Untersuchung mit Kontrollgruppe zwei weitere Gruppen erforderlich. Bei beiden zusätzlichen Gruppen wird auf eine Vorher-Messung verzichtet und nur eine dieser zusätzlichen Gruppen wird dem Treatment ausgesetzt. Eine solche Versuchsanordnung wird als Solomon-Vier-Gruppen-Design bezeichnet (z. B. Behnke et al. 2010, S. 63; Schnell et al. 2013, S. 215; Eifler 2014, S. 203).

Solomon-Vier-Gruppen-Design In Abbildung 23 ist das nach Richard L. Solomon (1949) benannte Solomon-Vier-Gruppen-Design dargestellt. Alle Teilnehmer werden per Zufall einer der vier Gruppen zugewiesen. In beiden Experimentalgruppen werden die Versuchspersonen dem Treatment ausgesetzt. Dabei wird in der ersten Experimentalgruppe die abhängige Variable vor und nach der Verabreichung des Treatments erhoben, in der zweiten Experimentalgruppe nur nach dem Konsum der Gewaltspiele. In den beiden Kontrollgruppen werden keine gewaltsamen Videospiele verabreicht. In der ersten Kontrollgruppe wird das aggressive Verhalten zweimal erhoben, in der zweiten Kontrollgruppe nur einmal. Eine Wechselwirkung zwischen Vorher-Messung und Treatment wäre dann gegeben, wenn sich O_3 und O_5 unterscheiden.

Abbildung 23 Solomon-Vier-Gruppen-Design

Randomisierung	Gruppe	Zeit1	Zeit2	Zeit3
Teilnehmer werden zufällig einer der vier Gruppen zugewiesen	Experimentalgruppe 1	O_1	X_1	O_3
	Kontrollgruppe 1	O_2		O_4
	Experimentalgruppe 2		X_2	O_5
	Kontrollgruppe 2			O_6

Anmerkung: O = Erhebung der aV. X = Treatment.

Quelle: Eigene Darstellung

Interne und externe Validität Ziel eines Experiments ist es, möglichst alle Alternativerklärungen für einen Zusammenhang zwischen einer unabhängigen Variable und einer abhängigen Variable auszuschließen.[28] In einem Experiment soll die mögliche Veränderung der aV ausschließlich auf die Manipulation der uV zurückzuführen sein. Wenn für eine eingetretene Veränderung nur eine bestimmte Ursache (die uV) verantwortlich sein kann, dann wird ein Experiment als intern valide bezeichnet.

> *„Als intern valide gilt eine experimentelle Untersuchung, wenn ein signifikanter Unterschied zwischen Kontrollbedingung und Versuchsbedingung tatsächlich auf die experimentelle Manipulation der unabhängigen Variable (und nichts anderes) zurückzuführen ist." (Faas und Huber 2010, S. 726)*

Eine hohe interne Validität ist meist mit künstlichen Rahmenbedingungen und realitätsfremden Situationen verbunden. Klassische „Laborexperimente" werden in eigens vorbereiteten Räumlichkeiten durchgeführt, die Teilnehmer werden zufällig verschiedenen Gruppen zugewiesen, es werden mehr oder weniger künstliche Treatments „verabreicht" und alle Reaktionen der Teilnehmer werden genau protokolliert. Die experimentelle Methode wird gelegentlich genau für diese „Künstlichkeit", welche die interne Validität sicherstellen soll, kritisiert. Danach können experimentelle Befunde nicht ohne weiteres auf das tatsächliche Leben übertragen werden. Die Übertragbarkeit experimenteller Befunde auf die Realität wird als externe Validität bezeichnet.

> *Externe Validität ist das „Ausmaß, in welchem die Ergebnisse einer Studie auf andere Situationen oder Menschen verallgemeinert bzw. generalisiert werden können". (Aronson et al. 2004, S. 48)*

Zwei Arten der Generalisierbarkeit Mit Aronson et al. (2004, S. 48) lassen sich zwei Arten der Generalisierbarkeit unterscheiden. Erstens die Generalisierbarkeit von einer künstlichen auf natürliche Situationen und zweitens die Generalisierbarkeit von den Versuchspersonen auf alle Menschen. Für beide Herausforderungen gibt es durchaus plausible Lösungsstrategien (Aronson et al. 2004,

---

28 Die wichtigsten Störfaktoren hat bereits Campbell (1957) zusammengestellt: Unklare zeitliche Abfolge, Einflüsse einer selektiven Auswahl, Einflüsse der Zeit, Einflüsse der Reifung, Regression, systematische Ausfälle von Personen, Einflüsse der Messung und Effekte veränderter Messinstrumente. Eine knappe Darstellung dieser Störfaktoren bieten Eifler (2014, S. 199) und Behnke et al. (2010, S. 65–67). Einzelne Störfaktoren können auch zusammenwirken und zusätzliche Probleme bereiten (siehe z. B. Shadish et al. 2002; Schnell et al. 2013, S. 209).

S. 48–51), die aber jeweils die interne Validität eines Experiments ungünstig beeinflussen.

> *„Unglücklicherweise stehen interne und externe Validität eines Experiments oft in einem Trade-off-Verhältnis zueinander. Die Erhöhung der internen Validität ist nur durch Verzicht auf externe Validität zu erreichen. Umgekehrt ist die externe Validität desto höher, je ‚natürlicher' die Experimentalsituation ist. Damit schwinden aber die Kontrollmöglichkeiten und die interne Validität nimmt ab. Der Forscher steht also vor der Entscheidung, in welchem Maße er auf Validität der einen Art verzichten möchte, um mehr Validität der anderen Art zu sichern." (Behnke et al. 2010, S. 68)*

Bei der Bewertung von Experimenten wird der internen Validität eine größere Bedeutung zugesprochen als der externen Validität (Eifler 2014, S. 204). Schließlich ist das Experiment die einzige Untersuchungsform, die kausale Mechanismen identifizieren kann. Deshalb wird bei Experimenten meist eine hohe interne Validität angestrebt, die häufig zu Lasten der externen Validität geht (siehe zur internen und externen Validität auch McDermott 2011).

Unterschiedliche Bedeutung des Begriffs „Validität"

Der Begriff der Validität wird in den Sozialwissenschaften in unterschiedlichen Zusammenhängen verwendet. Validität bezeichnet in der Phase der Operationalisierung die Gültigkeit eines Messinstruments (siehe Kapitel 7). Ein Messinstrument ist valide (gültig), wenn es tatsächlich das misst, was es messen sollte. Im Kontext von Experimenten wird zwischen der internen und externen Validität differenziert. Die interne Validität bezieht sich auf die Gültigkeit der Schlussfolgerung, dass die unabhängige Variable (und nur diese) die abhängige Variable beeinflusst. Die externe Validität bezieht sich auf die Generalisierbarkeit der Befunde eines Experiments auf andere Situationen oder Menschen. Die verschiedenen Formen der Validität dürfen nicht miteinander verwechselt werden.

In den vorangestellten Abschnitten wurde das klassische Experiment dargestellt, das sich durch eine hohe interne Validität und eine geringe externe Validität auszeichnet. Das klassische Experiment wird auch als Laborexperiment oder einfach als Experiment bezeichnet. Die Vorher-Nachher-Untersuchung und das Solomon-Vier-Gruppen-Design sind Beispiele für das klassische Experiment. Ein

solches Experiment ist – wie in der Definition bereits beschrieben – durch zwei Merkmale gekennzeichnet (siehe auch Huber 2013, S. 67):

- Die Manipulation einer uV durch den Forscher
- Die zufällige Verteilung auf Experimental- und Kontrollgruppe (Randomisierung)

Das Ziel eines Experiments ist die Prüfung kausaler Beziehungen zwischen einer unabhängigen und einer abhängigen Variable. Ein erfolgreich durchgeführtes Experiment erlaubt eine möglichst sichere Kausalinterpretation zwischen einer uV und einer aV. Angesichts der künstlichen Untersuchungssituation sind mit Blick auf die externe Validität Abstriche zu machen. Neben dem (klassischen) Experiment werden auch Feld- und Quasi-Experimente zu den experimentellen Untersuchungsformen gezählt. Beide Varianten, die im Vergleich zum (echten) Experiment eine geringere interne Validität aufweisen, werden im Folgenden knapp dargestellt.[29]

Feldexperiment Das klassische Experiment wird unter möglichst kontrollierten Bedingungen durchgeführt. Zu diesen Bedingungen gehören die Räume, die der Forscher entsprechend vorbereitet. Ein Feldexperiment findet dagegen in der natürlichen Umwelt der Versuchspersonen statt, also zum Beispiel in einer Schule oder in einem Betrieb. In den eigenen Räumen – im Labor – hat der Forscher zwar eine bessere Kontrolle über den Versuchsablauf (interne Validität), aber die Generalisierbarkeit (externe Validität) ist eingeschränkt. Bei einem Feldexperiment ist die Kontrolle dagegen geringer, aber die Nähe zur natürlichen Umgebung erleichtert die Übertragbarkeit auf andere Situationen. Dies begünstigt die externe Validität.

29 Die Bezeichnung „Experiment“ zeichnet sich durch eine inflationäre Verwendung aus. Neben Feld- und Quasi-Experimenten werden in der Literatur unter anderem auch Survey- bzw. Umfrage-Experimente (Faas 2009, S. 82; Faas und Huber 2010, S. 728), Online-Experimente (Faas 2009, S. 82–83) oder auch natürliche Experimente (Bauer 2015) unterschieden. Die einzelnen Formen orientieren sich mehr oder weniger am „echten“ Experiment. Allerdings sind in der Regel zentrale Merkmale eines Experiments nicht erfüllt (z. B. keine Randomisierung, keine Kontrollgruppe), sodass jeweils im Einzelfall zu entscheiden ist, ob es sich tatsächlich noch um ein experimentelles Design handelt. Psychologen legen bei dieser Bewertung strengere Maßstäbe an als Soziologen oder Politikwissenschaftler.

Klassiker der Sozialwissenschaften: Die verlorenen Briefe

Stellen Sie sich vor, Sie finden beim Einkaufen in der Innenstadt – in der Nähe eines Briefkastens – einen frankierten Brief auf der Straße, der an die „Freunde der Nazi-Partei" adressiert wurde. Würden Sie ihn einwerfen? Was, wenn der identische Brief an die „Freunde der Kommunistischen Partei", an die „Gesellschaft für medizinische Forschung" oder einfach an einen Walter Carnap adressiert wurde? Diese Frage mussten in den 1960er Jahren Personen in der amerikanischen Stadt Connecticut beantworten. Dort hatten Stanley Milgram, Leon Mann und Susan Harter (1965) rund 400 Briefe „verloren", die sich nur im Empfänger unterschieden. 72 Prozent der Briefe an die Gesellschaft für medizinische Forschung und 71 Prozent der Post an Walter Carnap kamen an; dagegen jeweils nur 25 Prozent der Briefe an die „Freunde der Nazi-Partei" und die „Freunde der Kommunistischen Partei". Der unterschiedliche Rücklauf zeigte, dass sich mit der Methode der „verlorenen Briefe" unauffällig die Haltung der Menschen zu bestimmten Organisationen und damit zu bestimmten Themen erfassen lassen. Die Passanten waren – ungefragt – Teilnehmer eines Feldexperiments geworden, bei dem die unabhängige Variable systematisch variiert wurde (hier: Empfänger der Briefe). „The focus of the technique is not on the individual reaction to the lost letters but, rather, on the rate of response for a particular organization relative to other organizations that serve as controls" (Milgram et al. 1965, S. 437).

Quasi-Experiment Im Vergleich zu einem klassischen Labor-Experiment ist bei einem Quasi-Experiment keine Randomisierung möglich, das heißt, die teilnehmenden Personen können nicht zufällig in eine Experimental- und eine Kontrollgruppe verteilt werden (Eifler 2014, S. 203). Quasi-experimentelle Untersuchungsformen sind in der empirischen Sozialforschung häufig zu beobachten, zum Beispiel in der Bildungs- oder Organisationsforschung.

Ein Beispiel für ein mögliches Quasi-Experiment ist die Einführung von Ganztagsschulen, die den Schulerfolg begünstigen sollen. In einer Studie könnten die Leistungen von Schülern, die eine Ganztagsschule besucht haben, mit den Leistungen von Jugendlichen verglichen werden, die keine Ganztagsschule besucht haben. Die Zuweisung, ob ein Kind eine Ganztagsschule besucht hat oder nicht, liegt allerdings nicht in der Kontrolle des Forschers. Mögliche Leistungsunterschiede können auf den Schulbesuch zurückzuführen sein. Allerdings sind auch

alternative Erklärungen denkbar. Möglicherweise werden insbesondere Schulen mit leistungsstärkeren Jugendlichen zu Ganztagsschulen umgewandelt. Der kausale Rückschluss, ob tatsächlich von der uV ein Effekt auf die aV ausgeht, kann nicht abschließend geklärt werden (siehe zur Quasi-Experimenten ausführlich Huber 2013, S. 195–202).

Stärken und Schwächen Wie Tabelle 12 zeigt, haben alle drei Arten von Experimenten – Labor-, Feld- und Quasi-Experimente – ihre spezifischen Stärken und Schwächen. Dabei lässt sich das Spannungsverhältnis zwischen interner und externer Validität nicht immer auflösen. Wer vollständige Kontrolle über Versuchspersonen und Bedingungen benötigt, muss in der Regel auf die höhere externe Validität von Feldexperimenten verzichten. Wer untersuchen möchte, ob die im künstlichen Labor nachgewiesenen kausalen Beziehungen auch jenseits des Labors existieren, muss bei der Prüfung auf eine perfekte Kontrolle verzichten. Nach Faas und Huber (2010, S. 729) sollte sich die Forschung daher nicht „mit der einmaligen Untersuchung eines Sachverhalts" zufriedengeben, sondern verschiedene Methoden anwenden und die Ergebnisse, die mit einer Methode gewonnen wurden, mit einer anderen Methode erneut testen. Die Wiederholung (Replikation) einer Studie mit verschiedenen Populationen und/oder verschiedenen Methoden stärkt das Vertrauen in die Zuverlässigkeit der empirischen Befunde.

Das Experiment gilt allgemein als Idealtyp empirischer Forschung, um kausale Zusammenhänge entdecken bzw. prüfen zu können (Behnke et al. 2010, S. 52). In den Naturwissenschaften und in der Psychologie stellt das Experiment auch die Standard-Untersuchungsform dar. In der Soziologie und in der Politikwissenschaft ist in jüngerer Zeit zwar eine stärkere Aufmerksamkeit gegenüber experimentellen Untersuchungsformen festzustellen (z. B. Faas und Huber 2010; Druckman et al. 2011; Morton und Tucker 2014; Keuschnigg und Wolbring 2015; Kubbe 2016), aber mit Faas (2009, S. 75) lässt sich argumentieren, dass Experimente in diesen Disziplinen „niemals ein Allheilmittel sein können".

Tabelle 12 Vergleich verschiedener Arten von Experimenten

Merkmal	Labor-Experimente	Feld-Experimente	Quasi-Experimente
Manipulation der uV	ja	ja	ja
Randomisierung	ja	eingeschränkt	nein
Interne Validität	hoch	eingeschränkt	eingeschränkt
Externe Validität	eingeschränkt	hoch	hoch

Diese Prognose lässt sich in zweifacher Weise begründen: Erstens setzt die Durchführung eines Experiments eine manipulierbare unabhängige Variable voraus. Der Forscher verändert oder setzt in kontrollierter Art und Weise die unabhängige Variable, um anschließend mögliche Effekte auf die abhängige Variable beobachten zu können. Zweitens werden bei einem echten Experiment die Versuchspersonen zufällig auf Experimental- und Kontrollgruppe verteilt (Randomisierung). Dadurch sind systematische Unterschiede der Gruppen (z. B. Alter, Bildung, Geschlecht) unwahrscheinlich. Beide Bedingungen eines Experiments lassen sich bei vielen soziologischen und politikwissenschaftlichen Fragestellungen allerdings nicht realisieren.

Probleme der Manipulation Bei vielen sozialwissenschaftlichen Studien ist eine Manipulation der unabhängigen Variable nicht möglich. Bestimmte Merkmale (etwa Alter und Geschlecht) entziehen sich einer experimentellen Variation. Bei anderen Merkmalen sprechen ethische Gründe gegen eine entsprechende experimentelle Variation. Bei einer Analyse des kausalen Zusammenhangs zwischen Bildung und Einkommen müsste beispielsweise die Bildung experimentell variiert werden. Bei einem Experiment zur Kausalanalyse der Arbeitslosigkeit auf die Wahlbeteiligung müsste etwa die Arbeitslosigkeit variiert werden. Solche Einschränkungen existieren nicht nur bei Individualmerkmalen, sondern auch bei Merkmalen des Kontexts:

> *„The really big social science variables – culture, economic development, ethnic heterogeneity – probably could not be manipulated even if political scientists were permitted to try." (Green und Gerber 2003, S. 102)*

Probleme der Randomisierung Auch die zufällige Zuweisung in Experimental- oder Kontrollgruppe (Randomisierung) lässt sich bei vielen sozialwissenschaftlichen Fragestellungen nicht realisieren. Wer sich für den Einfluss der Ehe auf die Lebenszufriedenheit interessiert, kann Personen nicht zufällig den beiden Bedingungen „verheiratet" und „nicht verheiratet" zuweisen. Wer den Effekt der Parteiidentifikation untersuchen möchte, kann die Teilnehmer nicht zufällig in die Gruppen „Parteiidentifikation vorhanden" und „keine Parteiidentifikation" einteilen. Wer den kausalen Effekt der Bildung auf das Einkommen untersuchen möchte, kann nicht zufällig ausgewählte Personen jahrelang von jeglicher Schulbildung ausschließen.

> *„The most widely cited drawback, and the one that warrants most of our attention, is the inability to manipulate key political variables of interest. It is difficult to imagine how one could randomly assign presidential and parliamentary regimes for the purpose of evaluat-*

ing their relative strengths and weaknesses. Surely, world leaders cannot be persuaded to allow political scientists to randomize their foreign policies, systems of patronage, or prospects for retaining power.“ (Green und Gerber 2003, S. 102)

Bei vielen sozialwissenschaftlichen Fragestellungen sind die Anforderungen, welche für ein Experiment erforderlich sind, nicht erfüllt bzw. nicht erfüllbar. Für solche Fragestellungen bietet sich entsprechend eine nicht-experimentelle Untersuchungsform an. Solche nicht-experimentellen Untersuchungsformen werden Ex-post-facto-Designs genannt. Der zentrale Unterschied ist die stark eingeschränkte Interpretation einer möglichen kausalen Beziehung zwischen uV und aV.

9.2.3 Ex-post-facto-Design

Bei einem Ex-post-facto-Design wird kein Treatment verabreicht und keine Randomisierung durchgeführt. Bei der einfachsten – und häufigsten – Variante werden die abhängigen und unabhängigen Variablen zu einem Zeitpunkt erfasst. Bei einer Befragung im Rahmen der ALLBUS werden die Personen in einem Interview unter anderem gefragt,

- ob sie verheiratet sind und wie zufrieden sie mit ihrem Leben sind,
- wie stark sie sich für Politik interessieren und ob sie am Sonntag wählen gehen
- und welchen Bildungsabschluss sie haben und wie hoch das monatliche Nettoeinkommen ist.

Die verschiedenen Merkmale – Familienstand, Lebenszufriedenheit, Politikinteresse, Wahlabsicht, Bildungsniveau und Nettoeinkommen – werden mit verschiedenen Fragen einmalig erfasst. Je nach Forschungsfrage werden im Nachhinein (ex post facto) die Personen unterschieden, die verheiratet sind oder nicht, die sich für Politik interessieren oder nicht und die einen niedrigen oder hohen Schulabschluss haben. Auf Basis der Angaben werden die Personen „theoretisch“ in Experimental- und Kontrollgruppe eingeteilt, um die Hypothesen zu überprüfen. Mit Blick auf die oben dargestellten Merkmale wären folgende Hypothesen denkbar:

- Wenn eine Person verheiratet ist, dann wird sie mit ihrem Leben wahrscheinlich zufriedener sein als eine unverheiratete Person.
- Wenn sich eine Person für Politik interessiert, dann wird sie sich eher an Wahlen beteiligen als eine politisch uninteressierte Person.

- Wenn eine Person einen höheren Bildungsabschluss hat, dann wird ihr Nettoeinkommen wahrscheinlich höher sein als bei einem niedrigen Bildungsabschluss.

Probleme des Ex-post-facto-Designs Bei einem Ex-post-facto Design werden uV und aV zeitgleich erhoben und im Anschluss findet eine (theoretische) Trennung in unabhängige und abhängige Variable statt. Ex-post-facto-Designs sind nach Stein (2014, S. 141) zwar die verbreitetste Untersuchungsanordnung in den Sozialwissenschaften, aber hinsichtlich der Interpretation einer kausalen Beziehung zwischen einer uV und einer aV besonders problembehaftet. In der Literatur werden drei zentrale, aber unvermeidliche Probleme des Ex-post-facto-Designs genannt (z. B. Westle 2009c, S. 143–146; Behnke et al. 2010, S. 73–79; Schnell et al. 2013, S. 223; Stein 2014, S. 142).

1) Das Problem der kausalen Reihenfolge der Variablen.
2) Das Problem der Kontrolle von Drittvariablen.
3) Das Problem der Varianz der unabhängigen Variablen.

Problem der Kausalität Das offensichtlichste Problem des Ex-post-facto-Designs ist die kausale Reihenfolge der Variablen. In einem einfachen Ex-post-facto-Design werden uV und aV zeitgleich erfasst. Die Personen werden an einem Zeitpunkt zu ihrem Familienstand und ihrer Lebenszufriedenheit oder zu ihrem politischen Interesse und ihrer Wahlbeteiligung gefragt. Sicherlich kann der Familienstand die Lebenszufriedenheit beeinflussen, sodass verheiratete Menschen zufriedener sind als unverheiratete Menschen. Allerdings könnte es auch sein, dass eher zufriedene Menschen heiraten. Verheiratete Menschen könnten schon vor der Eheschließung zufriedener sein als unverheiratete Personen. Das Beispiel lässt sich auch auf den Zusammenhang zwischen politischem Interesse und Wahlbeteiligung übertragen. Sicherlich gibt es viele Argumente für einen „kausalen“

Kausalität und Ex-post-facto-Design

Bei einem Ex-post-facto-Design kann festgestellt werden, ob Merkmal A und Merkmal B besonders häufig gemeinsam auftreten (korrelieren). Es kann aber nicht abschließend geklärt werden, welches Merkmal die Ursache und welches Merkmal die Wirkung ist. Das Problem der kausalen Interpretation ist bei einem Ex-post-facto-Design grundsätzlich nicht lösbar.

Effekt des politischen Interesses auf die Wahlbeteiligung. Könnte es nicht aber auch sein, dass Menschen nach dem Urnengang politisch interessierter sind, da sie ihre Stimme abgegeben haben und wissen möchten, ob die gewählte Partei jetzt erfolgreich war oder nicht?

Es existieren Strategien, um das Problem der Kausalität etwas abzumildern. Bei der Analyse von Zusammenhängen zwischen Merkmalen, bei denen eine zeitliche Abfolge offensichtlich ist, ist das Problem der kausalen Interpretation etwas geringer. Zum Beispiel ist die Schulbildung – in der Regel – dem Nettoeinkommen zeitlich vorgelagert. Auch Faktoren wie Geschlecht und Alter sind Merkmalen wie soziales Vertrauen, Demokratiezufriedenheit oder politisches Interesse vorgelagert. Ein Einfluss des Vertrauens auf das Geschlecht oder das Alter ist aus logischen Gründen auszuschließen. Allerdings sind bei solchen Merkmalen die theoretischen Argumente für einen möglichen Einfluss nicht unmittelbar zugänglich. Warum sollte das Geschlecht einen Einfluss auf das soziale Vertrauen haben? Warum sollte sich das Lebensalter auf die Wahlbeteiligung auswirken? Nicht nur, aber insbesondere bei sogenannten demographischen Merkmalen sind theoretische Argumente anzuführen, die einen möglichen Zusammenhang zwischen dem demographischen Merkmal und der abhängigen Variablen begründen. Esser (1987, S. 238) spricht beim Alter beispielsweise von einer gewissen Beliebigkeit, da in vielen Studien unklar bleibt, was mit dem Merkmal „Alter“ überhaupt erfasst werden soll – zunehmende geistige Inflexibilität, Änderung von Handlungsopportunitäten im Lebenslauf oder Kohorteneffekte? Auch die Befragung der gleichen Personen zu mehreren Zeitpunkten kann das Problem der kausalen Interpretation etwas verringern. Es handelt sich dann um ein sogenanntes Paneldesign (siehe Abschnitt 9.3).

Grundsätzlich kann bei einem Ex-post-facto-Design die Interpretation eines kausalen Zusammenhangs von zwei Merkmalen ausschließlich über theoretische Überlegungen erfolgen, die einen Kausalitätsmechanismus plausibel erscheinen lassen. Nur auf Basis theoretischer Argumente kann eine beobachtbare Korrelation als kausale Beziehung interpretiert werden.

> *„Kein noch so ausgeklügeltes Verfahren kann daher bei einem Ex-post-facto-Design eine kausale Inferenz begründen. Die Statistik kann die Arbeit des Formulierens einer Theorie im besten Fall unterstützen, aber niemals ersetzen.“ (Behnke et al. 2010, S. 78)*

Der Schluss von Korrelation auf Kausalität ist einer der häufigsten methodischen Fehler in den Sozialwissenschaften. Es wird leicht vergessen, dass eine Korrelation es nicht erlaubt, kausale Folgerungen zu ziehen. In Tageszeitungen, aber auch in wissenschaftlichen Studien, werden Korrelation und Kausalität mitunter gleichgesetzt. Deshalb: Selbst wenn es sehr plausibel erscheint, dass ein Merkmal ein

anderes Merkmal bedingt, stellen Sie sich immer die Frage, ob es nicht auch Alternativerklärungen geben könnte (eine Sammlung von vermeintlichen Kausalitätsbeziehungen findet sich bei Aronson et al. 2004, S. 44–45; Krämer 2009; Bauer et al. 2014).

Problem von Drittvariablen Bei einem Experiment wird durch Randomisierung sichergestellt, dass Experimental- und Kontrollgruppe vor der Verabreichung des Treatments vergleichbar sind. Dadurch ist der Einfluss anderer (vor allem unbekannter) Faktoren auf die abhängige Variable in beiden Gruppen gleich und der Effekt des Treatments in der Experimentalgruppe ist mit großer Sicherheit auf den Stimulus zurückzuführen. Bei einem Ex-post-facto-Design gibt es keine Randomisierung und deshalb ist ein Zusammenhang von zwei Merkmalen möglicherweise auf andere Faktoren – sogenannte Drittvariablen – zurückzuführen.

Weiter oben wurde die Hypothese „Je mehr Feuerwehrleute im Einsatz, desto höher der Sachschaden" dargestellt. Bei einem Experiment müsste die Anzahl der Feuerwehrleute experimentell variiert werden. In einer Bedingung löschen wenige Feuerwehrleute das Feuer, in einer anderen Bedingung sind es viele Einsatzkräfte. Alle anderen Faktoren – unter anderem die Qualifikation der Feuerwehrleute, deren Ausrüstung und insbesondere die Größe des Feuers – wären in beiden Bedingungen völlig identisch. Unter solchen Kontrollbedingungen ist die formulierte Hypothese nicht sonderlich plausibel. Warum sollte die Anzahl der Feuerwehrleute die Ursache für die Höhe des Sachschadens sein? Der entscheidende „dritte" Faktor im Beispiel ist die Größe des Feuers. In der Realität sind bei einem größeren Feuer zum einen mehr Personen im Einsatz und zum anderen ein größerer Sachschaden wahrscheinlich. Die Größe des Feuers beeinflusst die Anzahl der Einsatzkräfte und die Höhe des Sachschadens. Bei Berücksichtigung der Grö-

Scheinkorrelation und Scheinkausalität

Mit Scheinkorrelation wird der Zusammenhang von zwei Merkmalen bezeichnet, der durch einen dritten Faktor verursacht wird. Scheinkorrelation ist ein weitverbreiteter Begriff in der empirischen Sozialforschung. Behnke et al. (2010, S. 76) ist allerdings zuzustimmen, dass die Bezeichnung durchaus irreführend ist, da die Korrelation zwischen den zwei Merkmalen durchaus existiert. Eine Scheinkorrelation bedeutet vielmehr, dass es zwischen den Merkmalen keine kausale Beziehung gibt. Deshalb ist die Bezeichnung „Scheinkausalität" angemessener (Behnke et al. 2010, S. 76; siehe auch Westle 2009c, S. 145).

ße des Feuers verschwindet die Korrelation zwischen Anzahl der Einsatzkräfte und der Höhe des Sachschadens.

Ein möglicher Zusammenhang zwischen Familienstand und Lebenszufriedenheit oder politischem Interesse und Wahlbeteiligung könnte auch durch andere Faktoren beeinflusst sein. Möglicherweise leben verheiratete Menschen gesünder und die Gesundheit beeinflusst die Lebenszufriedenheit. Das politische Interesse könnte das politische Wissen fördern und Personen mit größerem politischen Wissen beteiligen sich eher an Wahlen als Personen mit geringerem Wissen über Politik. Bei beiden Beispielen sind Drittvariablen denkbar, die den Zusammenhang zwischen den zwei Merkmalen verursachen.

Mit sogenannten multivariaten Analyseverfahren (z. B. multiple Regression) können solche Drittvariablen relativ einfach identifiziert werden. Vor der sogenannten statistischen Drittvariablenkontrolle müssen allerdings zwei Bedingungen erfüllt sein, die mit Statistik überhaupt nichts zu tun haben: Erstens müssen die Merkmale bekannt sein, die möglicherweise einen Einfluss auf die abhängige Variable ausüben. Zweitens müssen empirische Informationen dieser Merkmale vorliegen, um eine entsprechende „Kontrolle" überhaupt durchführen zu können.

> *„Dies ist der entscheidende Unterschied zum experimentellen Design. Durch die zeitliche Anordnung der einzelnen Schritte und durch die Randomisierung ist beim experimentellen Design gewährleistet, dass alle potenziellen Drittvariablen kontrolliert werden, unabhängig davon, ob sie dem Forscher bekannt sind oder nicht." (Behnke et al. 2010, S. 77)*

Bei einem Ex-post-facto-Design kann der Einfluss von möglichen Drittvariablen auf die abhängige Variable nicht konstant gehalten werden. Bereits bei der Datenerhebung müssen potenzielle Drittvariablen erfasst werden, um diese bei der Datenanalyse berücksichtigten zu können. Da aber niemals alle bekannten und unbekannten Drittvariablen bei einer Datenerhebung berücksichtigt werden können, kann bei einem Ex-post-facto-Design niemals grundsätzlich ausgeschlossen werden, dass eine nachgewiesene Korrelation nur eine Scheinkausalität darstellt.

Problem der Varianz Im Vergleich zu den Problemen der Kausalität und der Drittvariablenkontrolle stellt die Varianz der unabhängigen Variable ein kleineres – und in der empirischen Praxis – durchaus lösbares Problem dar. Was ist mit dem Problem der Varianz der uV gemeint?

> *„Varianz auf der unabhängigen Variablen bedeutet, dass sich die Befragten in den Ausprägungen dieser Variable unterscheiden." (Schnell et al. 2013, S. 223)*

Soll beispielsweise der Zusammenhang zwischen Bildung und Einkommen untersucht werden, dann muss man sicherstellen, dass sich das Bildungsniveau der Befragten unterscheidet. Sollten nur Daten von Personen mit Hochschulreife vorliegen, dann kann die Relation zwischen Bildung und Einkommen nicht untersucht werden. Soll der Zusammenhang zwischen politischem Interesse und Wahlbeteiligung geprüft werden, dann müssen Personen mit schwachem, mittlerem und starkem politischen Interesse interviewt werden, um die Beziehung prüfen zu können. Mit anderen Worten: Bei einem Ex-post-facto-Design ist die Variation der uV zu gewährleisten.

Bei vielen – insbesondere auf der Individualebene angesiedelten – Fragestellungen sollte die Varianz der unabhängigen Variablen ausreichend sein. Problematisch kann es bei eher seltenen Merkmalen werden (z. B. Arbeitslosigkeit). Bei Aggregatdatenanalysen ist die Varianz der uV teilweise schwieriger zu gewährleisten. Eine Analyse des Einflusses von Regierungssystemen (parlamentarisch, präsidentiell oder semi-präsidentiell) auf die nationale wirtschaftliche Entwicklung in der EU ist beispielsweise schwierig zu realisieren, da es in der EU mit Zypern nur ein präsidentielles System gibt (Döring und Hönnige 2008, S. 455). Bei einer Analyse des Einflusses der Wahlpflicht auf die Wahlbeteiligung sind (möglichst ähnliche) Länder mit Wahlpflicht und ohne Wahlpflicht auszuwählen.

9.2.4 Zusammenfassung

In der quantitativ orientierten Sozialforschung lassen sich mit Experimenten und dem Ex-post-facto-Design zwei Untersuchungsformen unterscheiden. Nur das Experiment ermöglicht die Analyse eines kausalen Zusammenhangs zwischen zwei Merkmalen. Da sich aber viele soziologische und politikwissenschaftliche Fragestellungen aus ethischen oder praktischen Gründen nicht mit Experimenten bearbeiten lassen, sind in den Sozialwissenschaften Ex-post-facto-Designs die dominierende Untersuchungsform. Im Gegensatz zur experimentellen Methode besteht bei einem Ex-post-facto-Design allerdings nicht die Möglichkeit, den vermuteten (kausalen) Effekt einer uV auf eine aV eindeutig zu bestimmen.

Ex-post-facto-Designs sind durch drei zentrale Probleme charakterisiert: Erstens die kausale Reihenfolge der Variablen, zweitens die Kontrolle von Drittvariablen und drittens die Varianz der unabhängigen Variablen. Insbesondere die beiden erstgenannten Probleme erlauben es nicht, einen gefundenen Zusammenhang von zwei Merkmalen als kausalen Effekt einer uV auf eine aV zu interpretieren.

! Für verlässliche Schlussfolgerungen bei Ex-post-facto-Designs sind theoretische Überlegungen von zentraler Bedeutung. Theoretisch relevante Drittvariablen, die einen Zusammenhang zwischen uV und aV beeinflussen können, müssen operationalisiert und erfasst werden.

> *„Die Qualität einer Ex-post-facto-Untersuchung lässt sich somit vor allem an der Kontrolle von Drittvariablen erkennen, wobei keineswegs allein deren Anzahl entscheidend ist, sondern auch ihre theoretische Integration." (Schnell et al. 2013, S. 228)*

In einem Ex-post-facto-Design sollten nur Merkmale berücksichtigt werden, bei denen es theoretische Argumente für einen entsprechenden Zusammenhang gibt. Die Identifikation möglicher relevanter Merkmale setzt allerdings die intensive Auseinandersetzung mit dem existierenden Forschungsstand voraus. Wer den Forschungsstand nicht kennt, der wird vermutlich unwichtige Merkmale berücksichtigen und zentrale Faktoren außer Acht lassen.

9.3 Häufigkeit der Datenerhebung

Der dritte Aspekt des Forschungsdesigns betrifft die Häufigkeit der Datenerhebung. Mit dem Querschnittdesign und dem Längsschnittdesign werden zwei Grundtypen unterschieden. Beim Querschnittdesign findet eine einmalige, beim Längsschnittdesign eine mehrfache Datenerhebung statt. Beim Längsschnittdesign lassen sich mit dem Trend-, Kohorten- und Paneldesign drei verschiedene Varianten unterscheiden. Die Ausführungen zur Häufigkeit der Datenerhebung beziehen sich dabei auf das Ex-post-facto-Design.

9.3.1 Querschnittdesign

Bei der einfachsten – und häufigsten – Variante des Ex-post-facto-Designs werden aV und uV zu einem Zeitpunkt erfasst. Das Querschnittdesign erlaubt eine aktuelle Bestandsaufnahme (Stein 2014, S. 142). Wie zufrieden sind die Bürger (zum Zeitpunkt der Erhebung) mit den Leistungen der Bundesregierung? Wie stark unterscheidet sich (zum Zeitpunkt der Erhebung) das Einkommen zwischen Männern und Frauen? Wie stark interessieren sich die Personen (zum Zeitpunkt der Erhebung) für Politik?

Solche Querschnittdesigns können und werden selbstverständlich auch für die Analyse von Zusammenhängen genutzt. Welche Faktoren beeinflussen die Bewertung der Bundesregierung? Welche Merkmale können die geschlechtsspezifischen

Was ist ein Querschnittdesign?

Bei einem Querschnittdesign werden die relevanten Merkmale einmalig und zu einem Zeitpunkt bzw. innerhalb eines kurzen Zeitraums erfasst.

Unterschiede beim Einkommen erklären? Warum unterscheidet sich das allgemeine Politikinteresse in Abhängigkeit vom Bildungsniveau? Ein Querschnittdesign erlaubt durchaus verlässliche Aussagen zu einem Thema und ermöglicht die Prüfung von Zusammenhängen. Allerdings dürfen mögliche Zusammenhänge zwischen einer uV und einer aV nicht als kausale Effekte interpretiert werden: Verlässliche Aussagen über kausale Beziehungen zwischen den Merkmalen sind bei einem Querschnittdesign grundsätzlich nicht möglich (Stein 2014, S. 142).

Zentral: Theoretische Argumente Deshalb ist die Qualität einer Querschnittuntersuchung in erster Linie von der theoretischen Argumentation abhängig. Welche theoretischen Argumente präsentiert der Autor für einen möglichen Zusammenhang zwischen uV und aV? Werden mögliche kausale Mechanismen zwischen uV und aV präsentiert? Werden mögliche Alternativerklärungen genannt? Grundsätzlich sind empirische Ergebnisse anzuzweifeln und insbesondere bei Schlussfolgerungen sowie möglichen Empfehlungen ist (mehr) Zurückhaltung angebracht.

In den Sozialwissenschaften dominieren Querschnittdesigns. Im Vergleich zu einem Längsschnittdesign sind die erforderlichen Ressourcen relativ gering und eine empirische Überprüfung der theoretisch formulierten Hypothesen ist bereits nach einer Datenerhebung grundsätzlich möglich.

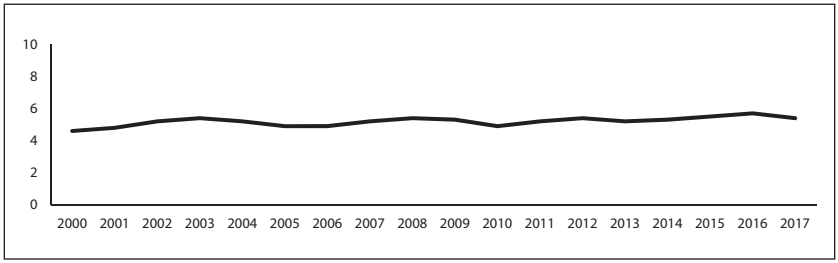
9.3.2 Längsschnittdesign

Bei einem Längsschnittdesign sind im Gegensatz zu einem Querschnittdesign wiederholte Erhebungen zu mehreren Zeitpunkten vorgesehen. Der kleine, aber entscheidende Unterschied bei den verschiedenen Varianten des Längsschnittdesigns: Beim Trend- und Kohortendesign werden die gleichen Merkmale zu mehreren Zeitpunkten an verschiedenen Untersuchungseinheiten erfasst. Bei einem Paneldesign werden die gleichen Merkmale zu mehreren Zeitpunkten an den gleichen Untersuchungseinheiten erhoben. Mit Untersuchungseinheiten sind in der Regel Personen gemeint, aber es können auch Makromerkmale wiederholt erfasst werden (z. B. jährliche Erhebung der Arbeitslosenquote oder der Inflationsrate).

9.3.2.1 Trenddesign

Bei einem Trenddesign werden die interessierenden Merkmale zu mehreren Zeitpunkten, aber an verschiedenen Untersuchungseinheiten erfasst. In jedem Semester könnte beispielsweise bei den Studierenden des Moduls M1 „Quantitative Methoden der Sozialwissenschaften" die allgemeine Zufriedenheit mit der Arbeit der Bundesregierung auf einer Skala von 0 bis 10 erfasst werden. Mit den individuellen Angaben können dann jedes Semester statistische Kennziffern (z. B. Modus, Median, Mittelwert) berechnet werden. Diese aggregierten Werte könnten im Längsschnitt verglichen werden. Hat die Zufriedenheit bei den Studierenden über die Zeit zu- oder abgenommen? Abbildung 24 zeigt auf Basis fiktiver Daten eine mögliche Entwicklung der Zufriedenheit mit der Bundesregierung.

Abbildung 24 Trenddesign mit fiktiven Daten



Quelle: Eigene Darstellung

Bei einem Trenddesign werden jeweils andere Personen befragt. Dies ist bei dem obigen Beispiel auch nachvollziehbar, weil jeweils die Studierenden des Moduls M1 „Quantitative Methoden der Sozialwissenschaften" befragt werden. In der Regel werden in jedem Semester andere Personen dieses Modul belegen. Bei sozialwissenschaftlichen Erhebungen wie ALLBUS oder ESS werden Aussagen über die Bevölkerung (die sogenannte Grundgesamtheit) angestrebt. Wir wollen beispielsweise wissen, wie zufrieden die Bevölkerung in Deutschland mit der Arbeit der Bundesregierung ist. Allerdings werden nicht etwa alle Bürger befragt, sondern nur eine Stichprobe (siehe Kapitel 10). Bei einem Trenddesign sind bei jeder Datenerhebung neue Stichproben erforderlich. Die einzelnen Stichproben setzen sich dabei aus verschiedenen Personen zusammen. Für jede Stichprobe werden unterschiedliche Personen befragt. Eine Trenderhebung entspricht damit mehreren Querschnitterhebungen zum gleichen Thema. Bei jeder Erhebung können

Mittelwerte oder Prozentangaben berechnet werden. Diese Kennwerte sind dann über die Zeit vergleichbar. Ein Trenddesign basiert in der Regel auf Stichproben. Dadurch sind Veränderungen über die Zeit (z. B. ein Anstieg der Zufriedenheit) möglicherweise auf Stichprobenfehler zurückzuführen. Solche Stichprobenfehler sind bei der Analyse immer zu berücksichtigen.

Bedingungen für einen Vergleich Neben unvermeidlichen Stichprobenfehlern formuliert Westle (2009c, S. 147) vier Bedingungen, die erfüllt sein müssen, um die Entwicklung bestimmter Merkmale über die Zeit vergleichen zu können. Erstens müssen die Merkmale bei den verschiedenen Erhebungen jeweils identisch bzw. äquivalent erfasst werden. Bei einer Befragung müssen beispielsweise jeweils die gleichen Fragen gestellt werden. Zweitens muss der „Typus" der Erhebungseinheiten jeweils identisch sein. Falls bei einer Erhebung beispielsweise die wahlberechtigte Bevölkerung interviewt wird, und bei einer anderen Erhebung die Bevölkerung ab 15 Jahren, dann wäre ein Vergleich nicht sinnvoll. Drittens sollten die Erhebungseinheiten jeweils mit demselben Auswahlverfahren ermittelt werden (siehe Kapitel 10). Viertens sollte die eigentliche Erhebung jeweils das gleiche Institut durchführen. Insbesondere die dritte und vierte Bedingung werden in der Praxis – insbesondere bei längeren Trenddesigns – nicht immer erfüllt. Dafür gibt es in der Regel auch gute und nachvollziehbare Gründe (z. B. Unzufriedenheit mit einem Institut), allerdings haben solche Veränderungen immer auch Konsequenzen für die Analyse im Zeitverlauf.

9.3.2.2 Kohortendesign

Ein Kohortendesign ist ein Spezialfall eines Trenddesigns. Was ist eine Kohorte?

> *„Eine Kohorte in den Sozialwissenschaften ist eine Bevölkerungsgruppe, die durch ein zeitlich gemeinsames, längerfristig prägendes Startereignis definiert wird." (Diekmann 2011, S. 318)*

Klassische „Startereignisse" sind Geburt, Eheschließung oder auch der Berufseintritt. Dabei werden Kohorten auf der Basis bestimmter individueller Ereignisse abgegrenzt (z. B. Geburt, Heirat, Berufseintritt). Es werden beispielsweise Personen befragt, die 1955, 1965 oder 1975 geheiratet haben. Alle Personen, die 1955 geheiratet haben, bilden entsprechend eine Kohorte.

Eine Kohorte ist zunächst einmal eine Bevölkerungsgruppe, ohne dass diese Personen in irgendeiner Art und Weise eine Gemeinschaft bilden müssen. Allerdings weisen insbesondere Geburtskohorten eine wichtige Gemeinsamkeit auf:

„Die Akteure einer bestimmten Geburtskohorte sind alle den Einflüssen einer bestimmten Epoche in der für das spätere Leben so wichtigen Phase der frühkindlichen Sozialisation ausgesetzt." (Esser 1999b, S. 268)

Kohorte und Generation Eine sozialwissenschaftliche Frage ist, ob die spezifischen historischen Umstände, in denen eine Kohorte aufgewachsen ist, Auswirkungen auf das spätere Leben haben. Wird diese Frage bejaht, dann werden in der Tradition von Karl Mannheim (1928) solche Kohorten als Generation bezeichnet. Danach kann von einer Generation gesprochen werden, wenn es in der jeweiligen Jugendzeit typische, abgrenzbare und nachhaltige Lebensumstände gegeben hat. Die in der Jugendzeit gemachten Erfahrungen – so die Annahme – bleiben auch im weiteren Lebensverlauf stabil und haben einen prägenden Einfluss auf Einstellungen und Verhaltensweisen (z. B. Esser 1999b, S. 269; Roßteutscher 2004, S. 189). Es könnte beispielsweise eine Vorkriegsgeneration, eine Kriegsgeneration und eine Nachkriegsgeneration unterschieden werden. In den jeweiligen Zeiträumen haben sich die Lebensbedingungen der einzelnen Kohorten unterschieden und diese Unterschiede waren so nachhaltig, dass sich diese Sozialisationserfahrungen auch Jahrzehnte später noch auf Einstellungen und Verhaltensweisen der Menschen auswirken (siehe ausführlicher z. B. Esser 1999b, S. 262–273; Engelhardt 2000, S. 535–539). Da allerdings nicht jede Geburtskohorte eine Generation ist, handelt es sich bei den Begriffen „Kohorte" und „Generation" nicht um Synonyme.

„In gewisser Weise ist der Kohortenbegriff eher bezogen auf statistische Aggregate, während der Generationenbegriff stärker auf eine gemeinsame Mentalität, ein ‚Generationenbewusstsein', abstellt." (Lois 2013, S. 18)

Heiratskohorten

Bei Diekmann (2011, S. 321) findet sich ein Beispiel für eine Kohortenstudie. Es werden Heiratskohorten der Jahre 1955, 1965 und 1975 verglichen, also Personen, die 1955, 1965 oder 1975 geheiratet haben. In allen drei Kohorten war das Scheidungsrisiko zwei bis fünf Jahre nach der Eheschließung am höchsten, allerdings unterscheidet sich der Anteil der geschiedenen Ehen nach einer Ehedauer von bis zu 9,5 Jahren erheblich zwischen den einzelnen Kohorten. In der Heiratskohorte 1955 waren nach einer Ehedauer von 9,5 Jahren 6,5 Prozent der Ehen geschieden, bei der Kohorte 1965 waren es zehn Prozent und bei der Kohorte 1975 waren es bereits 15 Prozent.

Voraussetzungen für Kohorteneffekte Die Annahme von Generationen- bzw. Kohorteneffekten ist an zwei Voraussetzungen geknüpft:[30] Erstens muss es eine Phase im Leben eines Menschen geben, in „der die für das spätere Leben zentralen, weitgehend stabilen Grundorientierungen erworben werden" (Arzheimer 2006, S. 321). Dies wird als formative Phase bezeichnet (Esser 1999b, S. 268). Zweitens muss es möglich sein, abgrenzbare Zeiträume zu definieren, in denen die Einstellungen der Menschen nachhaltig geprägt wurden. In der Literatur finden sich nicht nur verschiedene Konzepte (politischer) Generationen (z. B. Fogt 1982; Roßteutscher 2004; Kaspar und Falter 2007), sondern auch unterschiedliche Auffassungen zur zeitlichen Eingrenzung dieser sogenannten formativen Phase (für eine Übersicht siehe z. B. Tausendpfund 2013b, S. 99–113).

Kohorten-, Lebenszyklus- und Periodeneffekte Bei Kohortenuntersuchungen sind allerdings nicht nur Kohorteneffekte, sondern auch Lebenszyklus- und Periodeneffekte zu unterscheiden. Schließlich sind in der Empirie alle drei Effekte zu beobachten. Mit Kohorteneffekten sind systematische Unterschiede zwischen den Kohorten gemeint. Personen der Nachkriegsgeneration könnten beispielsweise gegenüber der Politik grundsätzlich positiver eingestellt sein als Angehörige der Kriegsgeneration, da sie in einer Phase des wirtschaftlichen Aufschwungs und des Friedens aufgewachsen sind. Lebenszykluseffekte – häufig auch Alterseffekte genannt – sind systematische Zusammenhänge zwischen den interessierenden Merkmalen (hier: Bewertung der Politik) und der seit dem Startereignis (hier: Geburt) verstrichenen Zeit (hier: das Alter der Person). Die Bewertung der Politik könnte auch Schwankungen im Lebensverlauf unterliegen. Möglicherweise sind jüngere und ältere Menschen gegenüber der Politik skeptischer eingestellt als Personen mittleren Alters. Bei Periodeneffekten handelt es sich um direkte Auswirkungen besonderer Ereignisse zu einem bestimmten Zeitpunkt. Periodeneffekte sind unmittelbare Anpassungen an aktuelle Ereignisse und Vorkommnisse, die alle Menschen unabhängig von Alter und Kohortenzugehörigkeit in gleicher Weise erfassen. Ein solches Ereignis könnte in Deutschland – mit Blick auf die Bewertung der Politik – die deutsche Wiedervereinigung gewesen sein.

Bei einer Untersuchung müssen jetzt Alters-, Perioden- und Kohorteneffekte analytisch und empirisch getrennt werden. Solche Analysen werden in der Literatur als APK-Analysen bezeichnet, da sie Auskunft über die relativen Effekte von Alter (A), Periode (P) und Kohortenzugehörigkeit (K) auf die jeweils interessie-

30 In der Literatur wird zwar häufig zwischen „Kohorte" und „Generation" differenziert, aber die Bezeichnungen „Generationeneffekt" bzw. „Kohorteneffekt" werden häufig synonym verwendet.

Klassiker der Sozialwissenschaften: The Silent Revolution

Die Wertewandeltheorie von Inglehart (1977, 1995, 2007) ist sicherlich der bekannteste sozialwissenschaftliche Ansatz, der auf solchen Kohorten- bzw. Generationeneffekten basiert. Danach hat in den westlichen Industrienationen ein intergenerationaler Wandel von materialistischen zu postmaterialistischen Werten stattgefunden. Materialistische Bedürfnisse wie ökonomische Sicherheit verlieren an Bedeutung, während postmaterialistische Anliegen wie Selbstverwirklichung wichtiger werden. Warum? Die Theorie des Wertewandels basiert auf zwei wesentlichen Annahmen: Der Mangel- und der Sozialisationshypothese. Die Mangelhypothese besagt, dass eine Person denjenigen Dingen den meisten Wert zumisst, die relativ knapp sind. Danach müssen erst zentrale Bedürfnisse erfüllt sein (z. B. Essen und Trinken), ehe ranghöhere Bedürfnisse (z. B. Selbstverwirklichung) an Bedeutung gewinnen. Die Sozialisationshypothese besagt, dass die grundlegenden Wertvorstellungen eines Menschen die Bedingungen widerspiegeln, die in seiner Jugendzeit vorherrschend waren. Wächst eine Person in einer Phase des wirtschaftlichen Wohlstands heran, dann wird sie wahrscheinlich postmaterialistische Wertorientierungen herausbilden, die auch gegen ökonomische Krisen im Erwachsenenalter resistent sind. Die lange Phase des wirtschaftlichen Erfolgs und der Sicherheit nach dem Zweiten Weltkrieg bildet gewissermaßen den Nährboden für die Ausbreitung postmaterialistischer Wertorientierungen.

rende aV geben (siehe z. B. Klein 2009 für die Wahl der Grünen; Arzheimer 2006 für die Wahlbeteiligung).

9.3.2.3 Paneldesign

Bei einer einmaligen Erhebung der relevanten Merkmale ist die Aussagekraft von Ex-post-facto-Designs erheblich eingeschränkt. Insbesondere können keine Aussagen über kausale Beziehungen zwischen den Merkmalen getroffen werden. Durch eine wiederholte Erhebung der Merkmale bei den gleichen Untersuchungseinheiten kann dieses Problem aber zumindest abgeschwächt werden. Ein solches Design wird allgemein als Panel bezeichnet (Schupp 2014).

Bei einem Panel werden – analog zu einem Trenddesign – die relevanten Merkmale mehrmals erhoben. Bei einem Trenddesign erfolgt die Erhebung al-

Was ist ein Panel?

Als Panel werden Untersuchungsanordnungen bezeichnet, die an denselben Untersuchungseinheiten (in der Regel Personen) dieselben Merkmale (mit derselben Operationalisierung) zu verschiedenen Zeitpunkten erheben (Schnell et al. 2013, S. 228). Die einzelnen Datenerhebungen werden auch als Erhebungswellen oder einfach als Wellen bezeichnet.

lerdings an verschiedenen, bei einem Panel an gleichen Untersuchungseinheiten. Da bei einem Trenddesign die Erhebungen an verschiedenen Objekten vorgenommen werden, können nur Veränderungen in der Gesamtheit der Gruppe festgestellt werden. Bei einem Paneldesign können durch die mehrfache Befragung einer Person auch Veränderungen bei einer Person beobachtet werden. Diese zwei Arten der beobachtbaren Veränderung werden als interindividuell und intraindividuell bezeichnet.

Intraindividuell bezeichnet die Veränderung einer Person auf einem Merkmal zwischen den Zeitpunkten der Messung. Diese Veränderung wird auch interne Fluktuation oder „turnover" genannt. Eine interindividuelle Veränderung bezieht sich auf die Gesamtheit der beobachteten Personen. Eine solche Veränderung wird als Nettoveränderung oder auch als „net change" bezeichnet (Schnell et al. 2013, S. 229–230).

In Deutschland ist das sozio-ökonomische Panel (SOEP) die bekannteste Panelerhebung. Seit 1984 werden jährlich die gleichen Personen in Haushalten in der Bundesrepublik Deutschland befragt. Dadurch können langfristige soziale und gesellschaftliche Trends besonders gut verfolgt werden. Mit dem Bildungspanel (www.neps-data.de) sowie dem Beziehungs- und Familienpanel pairfam (www.pairfam.de) werden aktuell noch weitere größere Panelerhebungen durchgeführt, die sich spezifischeren Fragestellungen widmen. Auch im Kontext der nationalen Wahlstudie (www.gles.eu) werden Panelerhebungen durchgeführt (siehe Abschnitt 8.3.1).

! Panelerhebungen sind organisatorisch aufwändig und kostenintensiv. Das Problem der kausalen Reihenfolge der interessierenden Merkmale kann bei einem Paneldesign dafür deutlich besser bearbeitet werden als bei einem Querschnitt- oder Trenddesign. Warum? Bei einem Paneldesign stehen die relevanten Informationen (z. B. Lebenszufriedenheit) vor und nach dem interessierenden Ereignis (z. B. Heirat) zur Verfügung. In einem klassischen Querschnittdesign (zeitgleiche Erhebung der Merkmale) kann zwar der Zusammenhang zwischen

Lebenszufriedenheit und Familienstand untersucht werden, aber es kann nicht geklärt werden, ob die Lebenszufriedenheit den Familienstand oder der Familienstand die Lebenszufriedenheit beeinflusst. Bei einem Paneldesign verfügen wir allerdings über Angaben zur Lebenszufriedenheit vor und nach einer Veränderung des Familienstands. Dadurch kann die kausale Reihenfolge der Merkmale besser bearbeitet werden. Zur angemessenen Bearbeitung solcher Fragestellungen sind in den Sozialwissenschaften allerdings auch fortgeschrittene statistische Verfahren erforderlich, die den Vorher-Nachher-Vergleich auf individueller Ebene umsetzen und das Potenzial der Daten nutzen (z. B. Brüderl 2010; Giesselmann und Windzio 2012, 2014).

Bei einem Paneldesign werden die gleichen Untersuchungseinheiten mehrmals mit den gleichen Erhebungsinstrumenten konfrontiert. Beim SOEP werden beispielsweise die gleichen Personen jährlich mit einem weitgehend identischen Fragebogen konfrontiert. Dabei sind zwei typische Probleme von Paneldesigns charakteristisch: Panelmortalität und Paneleffekte.

Panelmortalität Mit Panelmortalität ist der Anteil der Personen gemeint, die im Verlauf eines Panels ausfallen. Panelmortalität hat verschiedene Ursachen: Personen können sterben, umziehen oder einfach im Urlaub sein. Der häufigste Grund für eine hohe Panelmortalität ist allerdings der Verlust der Teilnahmemoti-

Klassiker der Sozialwissenschaften: The People's Choice

Bei einem Panel handelt es sich nicht etwa um eine neue Methode, sondern um ein Verfahren mit einer langen Tradition. Eine Forschergruppe um Paul F. Lazarsfeld (1968) hat bereits von Mai bis November 1940 in Erie County eine Panelstudie durchgeführt, um die Wahlabsicht der Bürger bei den anstehenden Präsidentschaftswahlen zu erfassen. Dazu wurde eine sieben Wellen umfassende Wiederholungsbefragung durchgeführt, wobei die Personen sechsmal vor der Wahl und einmal nach der Wahl befragt wurden. Auf Basis dieser empirischen Daten konnten unterschiedliche Fragestellungen bearbeitet werden: Wann stand die Wahlentscheidung fest? Welchen Einfluss hat der Wahlkampf auf die Wahlentscheidung? Wie unterscheiden sich die Wähler von Demokraten und Republikaner? Die Studie von Lazarsfeld et al. (1968) gilt heute als Klassiker der Wahlsoziologie. Eine aktuelle Wahlstudie mit einem Paneldesign hat Plischke (2014) vorgelegt. Er untersucht den Zeitpunkt der Wahlentscheidung und die zugrundeliegenden Entscheidungsprozesse bei Bundestagswahlen.

vation. Die Teilnahme an einem Panel ist freiwillig und häufig verlieren die Befragten über die einzelnen Wellen das Interesse an der Erhebung. Das Ausmaß der Panelmortalität variiert dabei in Abhängigkeit von der Anzahl der Wellen, dem Zeitabstand zwischen den Wellen und der sogenannten Panelpflege. Mit Panelpflege sind alle Maßnahmen gemeint, um die Panelmortalität möglichst gering zu halten (z. B. Dankschreiben, Geschenke, Recherchieren der neuen Adresse). Panelmortalität ist vor allem deshalb problematisch, weil die Ausfälle nicht zufällig erfolgen. Die ausgefallenen Personen unterscheiden sich häufig von den im Panel verbliebenen Personen (siehe zur Teilnahmeverweigerung in Panelstudien z. B. Haunberger 2011).

Paneleffekte Das zweite Problem von Panelstudien sind Paneleffekte. Damit sind Veränderungen der Teilnehmer durch die wiederholte Befragung gemeint. Diese können entstehen, wenn Personen durch die Teilnahme an der Befragung sich verstärkt mit Themen der Erhebung auseinandersetzen und in Folge dessen eine Haltung zu dieser Thematik entwickeln, verändern oder festigen (Schnell et al. 2013, S. 233; Schupp 2014).

Zur Verringerung typischer Panel-Probleme wurden mit dem alternierenden Panel, dem rotierenden Panel und dem geteilten Panel verschiedene Panel-Varianten entwickelt. Bei einem alternierenden Panel werden die Personen in zwei Grup-

Panel Conditioning

Eine jüngere Studie zu Panel Conditioning hat Bergmann (2015) vorgelegt. Panel Conditioning ist der Effekt wiederholter Befragungen auf die Einstellungen der Befragten durch die wiederholte Beantwortung identischer Fragen im Rahmen einer mehrwelligen Umfrage. Bergmann hat in seiner Studie einerseits einen theoretischen Ansatz entwickelt, der die bestehenden Ansätze zur Erklärung von Panel Conditioning-Effekten integriert, und andererseits eine Möglichkeit der Messung von Panel Conditioning vorgeschlagen. Vor diesem Hintergrund präsentiert Bergmann seine Befunde: Urteile gegenüber wiederholt abgefragten Objekten können von den Personen tatsächlich erheblich schneller geäußert werden als bei ihrer ersten Konfrontation. Zudem deuten die Ergebnisse darauf hin, dass sich die Personen aufgrund der wiederholten Konfrontation nicht nur intensiver mit ihren Einstellungen auseinandersetzen, sondern sich darüber hinaus auch aktiv um weitergehende Informationen bemühen (Bergmann 2015, S. 291).

pen eingeteilt, die dann abwechselnd befragt werden. Dadurch wird einerseits der Gesamtaufwand der Erhebungen verringert und andererseits lassen sich möglicherweise Ausfälle aufgrund von Belastungseffekten verringern. Beim rotierenden Panel wird zu jedem Erhebungszeitpunkt eine Teilgruppe durch eine neue Gruppe ersetzt. Beim geteilten Panel werden zwei Gruppen gebildet. Eine Gruppe „durchläuft" eine klassische Panelerhebung und eine zweite Gruppe wird als rotierendes Panel oder als wiederholte Querschnittuntersuchung in die Erhebung aufgenommen (siehe ausführlich Schnell et al. 2013, S. 233–235).

9.3.3 Zusammenfassung

Bei Ex-post-facto-Designs lassen sich mit dem Querschnitt- und dem Längsschnittdesign zwei Grundvarianten der Häufigkeit der Datenerhebung unterscheiden. Bei einem Querschnittdesign findet eine einmalige und bei einem Längsschnittdesign eine mehrmalige Datenerhebung statt. Bei einem Längsschnittdesign lassen sich mit dem Trend-, Kohorten- und Paneldesign verschiedene Varianten unterscheiden, die sich auf die konkrete Datenerhebung beziehen. Bei einem Trend- und Kohortendesign werden die Merkmale bei verschiedenen, bei einem Paneldesign bei den gleichen Untersuchungseinheiten erhoben.

> *„Zwischen den drei Designtypen existiert eine Informationshierarchie. Panelerhebungen sind informativer als Trenderhebungen und diese informativer als Querschnitterhebungen. Panelerhebungen können auch über aggregierte Trends und Trenderhebungen über Querschnitte informieren, während die umgekehrte Relation nicht gilt." (Diekmann 2011, S. 306)*

In Tabelle 13 sind die zentralen Unterschiede zwischen Querschnitt-, Trend- und Paneldesigns zusammengefasst. Ein Querschnittdesign ist durch eine einmalige Datenerhebung charakterisiert. Diese Variante erfordert vergleichsweise wenig Ressourcen und ist für viele Fragestellungen auch völlig ausreichend. Bei einem Trenddesign werden die Merkmale mehrmals erfasst (z. B. jährlich), aber bei unterschiedlichen Untersuchungseinheiten. Die Allgemeine Bevölkerungsumfrage der Sozialwissenschaften (ALLBUS) oder der European Social Survey (ESS) sind typische Beispiele für solche Trenderhebungen. Zu jedem Zeitpunkt kann etwa das durchschnittliche Nettoeinkommen oder die durchschnittliche Zufriedenheit mit der Bundesregierung erfasst und dargestellt werden. Dadurch lassen sich gesellschaftliche Veränderungen nachverfolgen. Bei einem Paneldesign werden die Merkmale mehrmals bei den gleichen Untersuchungseinheiten erfasst. Das Sozioökonomische Panel (SOEP) ist die bekannteste Panelerhebung in Deutschland.

Tabelle 13 Vergleich von Querschnitt-, Trend- und Paneldesign

Merkmal	Querschnittdesign	Trenddesign	Paneldesign
Häufigkeit der Datenerhebung	einmal	mehrmals	mehrmals
Gleiche Untersuchungseinheiten	./.	nein	ja

Ein Panel ermöglicht die Analyse intraindividueller Veränderungen, z. B. kann die individuelle Veränderung des Nettoeinkommens untersucht werden. Paneldesigns sind allerdings auch mit spezifischen Problemen behaftet (z. B. Panelmortalität, Paneleffekte), die bei der Analyse zu berücksichtigen sind.

Auswahlverfahren

10

Über wen oder was sollen Aussagen gemacht werden? Diese Frage steht bei der Festlegung des Auswahlverfahrens im Mittelpunkt. Eine erste Entscheidung bei der Konkretisierung des Auswahlverfahrens ist die Festlegung, ob alle Untersuchungsobjekte (z. B. alle Personen), über die eine Aussage gemacht werden soll, untersucht werden oder nur eine Teilmenge dieser Elemente. Aus zeitlichen, finanziellen und forschungspraktischen Gründen dominieren in den Sozialwissenschaften Teilerhebungen. Bei den Teilerhebungen lassen sich zufällige und nichtzufällige Auswahlverfahren unterscheiden. In diesem Kapitel werden wichtige sozialwissenschaftliche Auswahlverfahren vorgestellt, Vor- und Nachteile dieser Verfahren skizziert und die forschungspraktische Umsetzung dargestellt.

10.1 Grundgesamtheit

Quantitativ arbeitende Sozialforscher versuchen möglichst allgemeingültige Aussagen zu treffen. Deshalb beziehen sich Hypothesen wie „Je höher die Bildung, desto größer das Einkommen“ oder „Je stärker das politische Interesse, desto wahrscheinlicher die Wahlteilnahme“ häufig nicht auf einen konkreten geographischen Raum oder einen spezifischen Zeitpunkt. In einem sozialwissenschaftlichen Projekt kann allerdings nicht „die Welt in ihrer Gesamtheit“ (Behnke et al. 2010, S. 139) untersucht werden. Deshalb müssen Sozialwissenschaftler festlegen, welchen Ausschnitt der Wirklichkeit sie untersuchen möchten. Dieser Ausschnitt der Wirklichkeit wird Grundgesamtheit genannt.

> *„Unter Grundgesamtheit ist diejenige Menge von Individuen, Fällen, Ereignissen zu verstehen, auf die sich die Aussagen der Untersuchung beziehen sollen und die im Hinblick auf die Fragestellung und die Operationalisierung vorher eindeutig abgegrenzt werden muss.“ (Kromrey et al. 2016, S. 256)*

© Springer Fachmedien Wiesbaden GmbH, ein Teil von Springer Nature 2018

M. Tausendpfund, *Quantitative Methoden in der Politikwissenschaft*,
Grundwissen Politik, https://doi.org/10.1007/978-3-658-20698-7_10

Festlegung der Grundgesamtheit Die Grundgesamtheit einer Untersuchung muss möglichst exakt beschrieben werden. Diese Beschreibung besteht aus einem räumlichen, einem sachlichen und einem zeitlichen Definitionsteil (von der Heyde 2014a, S. 25). Der räumliche Definitionsteil bezieht sich auf das räumliche Gebiet der Grundgesamtheit. Bei einer Untersuchung zur Bundestagswahl wäre beispielsweise das Gebiet der Bundesrepublik Deutschland und bei einer Studie zur lokalen Wahlbeteiligung das jeweilige Gemeindegebiet das zugrundeliegende räumliche Gebiet. Der sachliche Definitionsteil bezieht sich bei einer Studie zur Wahlbeteiligung (in der Regel) auf die wahlberechtigte Bevölkerung. Die wahlberechtigte Bevölkerung unterscheidet sich in Abhängigkeit von der betrachteten Wahl; bei lokalen Wahlen sind beispielsweise auch EU-Ausländer wahlberechtigt. Dies ist bei der Definition der Grundgesamtheit jeweils zu berücksichtigen. Der zeitliche Definitionsteil konkretisiert unter anderem den Erhebungszeitraum. Dies stellt bei einer Befragung eine besondere Herausforderung dar, da sich die Bevölkerung ständig verändert. Bei einer Wahlstudie könnten dies beispielsweise alle Personen sein, die zu einem bestimmten Stichtag wahlberechtigt sind und in Privathaushalten leben. Formal ist der räumliche, sachliche und zeitliche Definitionsteil meist klar abgrenzbar, die Umsetzung in der Praxis kann allerdings durchaus eine Herausforderung sein.

Die Festlegung der Grundgesamtheit ist immer abhängig von der konkreten Forschungsfrage. Einige Beispiele: Bei einer Befragung der Studierenden der FernUniversität in Hagen bilden alle (zu einem bestimmten Zeitpunkt eingeschriebenen) Studierenden der FernUniversität die Grundgesamtheit. Bei einer

Beispiel: Grundgesamtheit bei der ALLBUS 2014

In den meisten empirischen Studien wird die konkrete Definition der Grundgesamtheit nur knapp dargestellt (z. B. die wahlberechtigte Bevölkerung). Für einen ersten Überblick ist dies auch meist ausreichend. Präzisere Definitionen finden sich in sogenannten Technical Reports. Im Bericht zur ALLBUS 2014 wird die Grundgesamtheit wie folgt beschrieben (Wasmer et al. 2017, S. 42): „Die Grundgesamtheit des ALLBUS 2014 besteht aus allen in der Bundesrepublik Deutschland in Privathaushalten wohnenden Personen (Deutsche und Ausländer), die vor dem 1. 1. 1996 geboren wurden. Ausländische Personen wurden befragt, wenn das Interview in Deutsch durchgeführt werden konnte. Ausländer, deren Deutschkenntnisse für eine Befragung mit einem deutschen Fragebogen nicht ausreichten, sind als Ausfälle aufgrund ‚mangelnder Sprachkenntnisse' verzeichnet."

Untersuchung der Zufriedenheit der Mitarbeiter einer Firma bilden alle Mitarbeiter die Grundgesamtheit und bei einer Analyse von Wahlprogrammen setzt sich die Grundgesamtheit aus allen Wahlprogrammen zusammen.

10.2 Vollerhebung

Bei einer Vollerhebung werden empirische Informationen aller Untersuchungsobjekte der Grundgesamtheit erhoben. Bei einer Befragung der Studierenden der FernUniversität in Hagen werden beispielsweise alle (aktuell) eingeschriebenen Studierenden befragt. Insbesondere bei einer überschaubaren Grundgesamtheit bietet sich häufig eine Vollerhebung an.

Vorteil einer Vollerhebung Eine solche Vollerhebung hat den Vorteil, dass „die gesamte Verteilung der Merkmale ihrer Elemente bekannt ist" (Schnell et al. 2013, S. 258). Beschreibende Lagemaße – z. B. Modus, Median, Mittelwert – können für die Grundgesamtheit berechnet werden.[31] Bei einer Teilerhebung müssen diese Merkmale einer Grundgesamtheit geschätzt werden. Solche Schätzungen sind immer mit einem „Fehler" behaftet, das heißt, der geschätzte Wert weicht mehr oder weniger vom „wahren Wert" ab. Basiert die Teilerhebung auf einem zufälligen Auswahlverfahren dann kann die Abweichung zwar berechnet werden, aber bei einer Vollerhebung existiert die Abweichung vom „wahren Wert" überhaupt nicht.

Nachteile einer Vollerhebung Vollerhebungen haben allerdings mehrere Nachteile. Erstens die Kosten. Je größer die Grundgesamtheit, desto höher die Kosten. Die Erhebungskosten sind bei einer Mitarbeiterbefragung eines multinationalen Konzerns mit etwa 60 000 Angestellten selbstverständlich höher als bei einer mittelständischen Firma mit 200 Mitarbeitern. Zweitens die erforderliche Zeit. Eine Befragung von 60 000 Personen dauert erheblich länger als eine Erhebung von 200 Angestellten. Bei einer Volkszählung dauert es – je nach Größe des Landes – mehrere Jahre, ehe die (ersten) Ergebnisse vorliegen. Die Ergebnisse haben dann häufig eine geringe Aktualität. Drittens kann eine Vollerhebung auch destruktiv sein. Bei einer vollständigen Qualitätskontrolle einer Whiskyproduktion, könnte die entsprechende Destillerie zeitgleich Insolvenz anmelden. Viertens kann eine Teilerhebung präzisere Ergebnisse ermöglichen als eine Vollerhebung. Bei einer

31 Besteht die Grundgesamtheit aus Menschen (z. B. Studierende der FernUniversität in Hagen), dann beteiligen sich in der Regel nicht alle Personen an einer entsprechenden Befragung. Diese Nicht-Beteiligung wird allgemein als Nonresponse bezeichnet und in Abschnitt 11.1.7 dargestellt.

Teilerhebung kann möglicherweise besonders geschultes Personal eingesetzt werden. Dadurch werden Erfassungsfehler verringert (Schnell et al. 2013, S. 259; eine Übersicht der Vor- und Nachteile bietet auch Kromrey et al. 2016, S. 254–255). Angesichts dieser Nachteile von Vollerhebungen ist es wenig überraschend, dass in den Sozialwissenschaften Teilerhebungen dominieren.[32]

10.3 Teilerhebung

Bei einer Teilerhebung werden nur bei einer Teilmenge der Grundgesamtheit die empirischen Informationen erhoben. In der Literatur finden sich folgende begriffliche Unterscheidungen: Grundgesamtheit, Auswahlgesamtheit und Stichprobe (z. B. Diekmann 2011, S. 376–380; Schnell et al. 2013, S. 261–263; Kromrey et al. 2016, S. 255–263).

Die Grundgesamtheit – häufig auch als angestrebte Grundgesamtheit bezeichnet – umfasst alle Elemente, über die Aussagen beabsichtigt sind. Die Auswahlgesamtheit umfasst alle Elemente, die eine Chance haben, in die Stichprobe zu gelangen. Im Idealfall ist die Auswahlgesamtheit mit der Grundgesamtheit deckungsgleich. In der Forschungspraxis sind mit Overcoverage und Undercoverage allerdings zwei Probleme zu unterscheiden.

Overcoverage und Undercoverage Overcoverage bezeichnet die Menge an Untersuchungsobjekten, die theoretisch in die Stichprobe gelangen kann, aber eigentlich gar nicht zur Grundgesamtheit gehört. Bei einer Wahlstudie sind dies beispielsweise Personen, die noch nicht volljährig und deshalb auch nicht wahlberechtigt sind. Mit Undercoverage sind Elemente gemeint, die eigentlich eine Chance haben sollten, in die Stichprobe zu gelangen, aber faktisch nicht in der Auswahlgesamtheit repräsentiert sind. Bei einer Online-Erhebung sind das beispielsweise Personen, die zwar zur Grundgesamtheit gehören, aber keinen Internetanschluss haben. Wenn in einer Auswahlgesamtheit Untersuchungsobjekte vorhanden sind, die nicht zur Grundgesamtheit gehören, dann wird von Overcoverage gesprochen. Wenn in der Auswahlgesamtheit Elemente der Grundgesamtheit nicht berücksichtigt wurden, dann spricht man von Undercoverage.

32 Deshalb müssen sich Studierende der Sozialwissenschaften auch mit der vielfach unbeliebten Inferenzstatistik auseinandersetzen (häufig auch als schließende oder induktive Statistik bezeichnet). Die Inferenzstatistik stellt „Instrumente“ bereit, die es erlauben, Rückschlüsse von der Stichprobe auf die Grundgesamtheit zu ziehen. Bei Behnke (2005; siehe auch Broscheid und Gschwend 2005) finden sich Argumente für die Anwendung inferenzstatistischer Methoden auch bei Vollerhebungen.

Auswahl- und Erhebungseinheiten Nach der Durchführung eines Auswahlverfahrens liegt eine Stichprobe vor. Eine Stichprobe ist eine Auswahl von Elementen aus der Auswahlgesamtheit.[33] Dabei wird zwischen Auswahl- und Erhebungseinheiten differenziert. Auswahleinheiten sind Einheiten, auf die sich das Auswahlverfahren (Auswahlplan) konkret bezieht. Erhebungseinheiten sind Einheiten, bei denen die Informationen konkret erhoben werden.

Ein Beispiel: Bei einer Studie zu den mathematischen Kompetenzen der Schüler in Nordrhein-Westfalen bilden zunächst alle Schulen die Grundgesamtheit. Aus dieser Grundgesamtheit werden Schulen ausgewählt. In diesem Fall stellen die Schulen die Auswahleinheiten dar, die mathematischen Kompetenzen werden allerdings mittels eines standardisierten Tests bei den Schülern erfasst. Die Schüler sind entsprechend die Erhebungseinheiten. In diesem Beispiel unterscheiden sich Auswahl- und Erhebungseinheiten. Zunächst werden Schulen bzw. Schulklassen ausgewählt, die konkreten Informationen dann aber bei den Schülern erhoben. Falls allerdings eine Namensliste aller Schülerinnen der Grundgesamtheit existiert, dann könnte diese Liste für die Stichprobenziehung genutzt werden. In diesem Fall wären Auswahl- und Erhebungseinheiten identisch. In der Forschungspraxis liegen solche Listen allerdings selten vor oder die Institutionen, die die Informationen haben, stellen diese aus Datenschutzgründen nicht zur Verfügung.

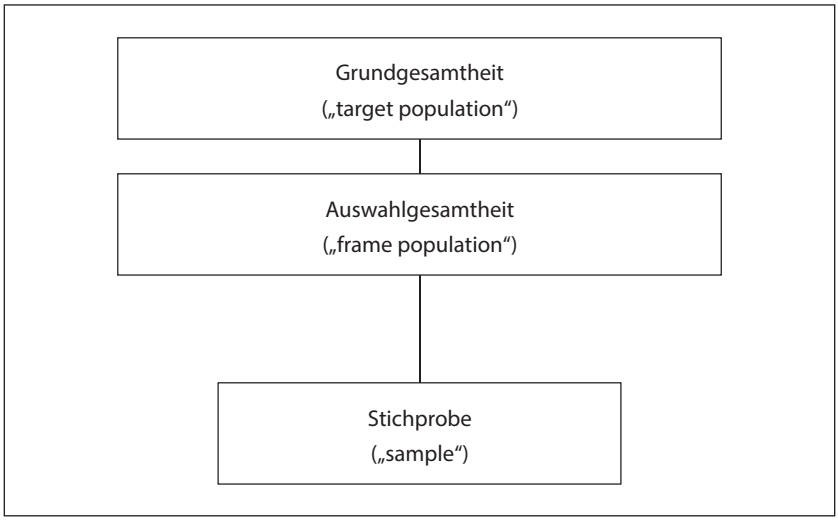
In Abbildung 25 ist die Beziehung zwischen Grundgesamtheit, Auswahlgesamtheit und Stichprobe dargestellt. Bei einer Untersuchung ist zunächst die Grundgesamtheit präzise festzulegen. Diese Festlegung besteht aus einem räumlichen, einem sachlichen und einem zeitlichen Definitionsteil. Bei einer lokalen Studie zum „Leben in der Gemeinde" könnten alle wahlberechtigten Bürger dieser Kommune zu einem bestimmten Stichtag die Grundgesamtheit bilden. Bei einer solchen Festlegung der Grundgesamtheit wären allerdings alle nicht-wahlberechtigten Bürger – insbesondere Kinder und Jugendliche sowie Nicht-EU-Bürger – aus der Untersuchung ausgeschlossen.

Im nächsten Schritt muss die Auswahlgesamtheit konstruiert werden, aus der faktisch die Stichprobe gezogen wird. Dazu könnte das lokale Einwohnermelderegister genutzt werden.[34] Dieses Register wird allerdings Fehler enthalten. Es wird

33 In der Fachliteratur besteht kein Konsens, ob nur dann von einer Stichprobe gesprochen werden darf, wenn die Auswahl auf einem Zufallsprozess basiert oder es sich immer um Stichproben handelt, wenn nicht alle Elemente betrachtet werden (Häder 2015, S. 142). An dieser Stelle wird der Auffassung Häders (2015, S. 142) gefolgt, demnach „es sich bei jeder Auswahl von Elementen aus einer Grundgesamtheit auch um eine Stichprobe handelt".

34 In Deutschland gelten die Daten der Einwohnermeldeämter als bestmögliche Auswahlgrundlage für Stichproben der allgemeinen Bevölkerung. Allerdings stehen Einwohnermelderegister aus Datenschutzgründen für viele Untersuchungen nicht zur Verfügung (siehe z. B. Albers 1997; von der Heyde 2014b). Siehe auch Abschnitt 10.4.3.

Abbildung 25 Grund- und Auswahlgesamtheit sowie Stichprobe



Quelle: Eigene Darstellung

Personen enthalten, die umgezogen oder verstorben sind (Overcoverage). Andererseits leben in der Gemeinde möglicherweise Personen, die zwar wahlberechtigt, aber aus technischen Gründen nicht im Register erfasst sind (Undercoverage). Bei der Stichprobenziehung werden direkt Personen ausgewählt, die befragt werden sollen. Daher sind in diesem Beispiel Auswahl- und Erhebungseinheiten deckungsgleich.

Zufällige und nichtzufällige Auswahlverfahren Bei Teilerhebungen lassen sich zwei Varianten unterscheiden: Zufällige und nichtzufällige Auswahlverfahren. Die Entscheidung für ein zufälliges oder nichtzufälliges Auswahlverfahren ist in erster Linie von der konkreten Forschungsfrage abhängig. Häufig sollen die auf Basis einer Stichprobe ermittelten Befunde allerdings auf die Grundgesamtheit übertragen werden. Ein Beispiel stellt die klassische Sonntagsfrage dar: „Welche Partei würden Sie wählen, wenn am kommenden Sonntag Bundestagswahl wäre?“ Nicht die Stichprobenergebnisse, sondern der Schluss von der Stichprobe auf die Grundgesamtheit aller wahlberechtigten Bürger ist bei dieser Frage interessant. Der Schluss von einer Stichprobe auf eine Grundgesamtheit setzt zwingend ein zufälliges Auswahlverfahren voraus. Nur bei einer Zufallsstichprobe kann innerhalb gewisser Fehlertoleranzen von den Ergebnissen einer Stichprobe auf die

Grundgesamtheit geschlossen werden. Deshalb kommt den zufälligen Auswahlverfahren in der quantitativ orientierten Sozialforschung eine besondere Bedeutung zu.

10.3.1 Zufällige Auswahlverfahren

Bei den zufälligen Auswahlverfahren lassen sich drei klassische Varianten unterscheiden: Die einfache Zufallsstichprobe, die geschichtete Zufallsstichprobe und die Klumpenstichprobe. In der Forschungspraxis werden die einzelnen Verfahren teilweise kombiniert. Verfahren, die Zufallsauswahlen in mehreren Stufen vornehmen, werden als mehrstufige Auswahlverfahren oder als „komplexe Stichproben" (Kohler und Kreuter 2017, S. 225) bezeichnet.

Einfache Zufallsstichprobe Die einfache Zufallsstichprobe gilt als die „Mutter" aller Stichprobenverfahren. Alle komplizierten Zufallsverfahren basieren auf der einfachen Zufallsstichprobe. Die Idee der einfachen Zufallsstichprobe basiert auf einem einfachen Prinzip. Jedes Element der Grundgesamtheit hat die gleiche Chance in die Stichprobe zu gelangen. Die Chance ist für jedes Element der Grundgesamtheit identisch und größer als 0.

- **Definition: Einfache Zufallsstichprobe**
  „Wird aus einer Grundgesamtheit mit N Elementen eine Stichprobe mit n Elementen so gezogen, dass jede mögliche Stichprobe mit n Elementen dieselbe Chance zur Realisierung besitzt, dann wird das Auswahlverfahren als ‚einfache Zufallsauswahl' (simple random sampling) und die resultierende Stichprobe als ‚einfache Zufallsstichprobe' (simple random sample, bzw. SRS) bezeichnet" (Schnell et al. 2013, S. 263). Dabei bezeichnet N die Anzahl der Elemente der Grundgesamtheit und n die Anzahl der Elemente der Stichprobe.

Die einfache Zufallsstichprobe ist das einfachste Auswahlverfahren und wird häufig auch als Urnenmodell bezeichnet. Auf Basis der einfachen Zufallsstichprobe werden oftmals zentrale Konzepte der induktiven Statistik wie Standardfehler, Konfidenzintervalle und Signifikanztests erläutert.

Zur Veranschaulichung der einfachen Zufallsstichprobe bietet sich folgendes Beispiel an: Bei einer Untersuchung der Studierenden der FernUniversität in Hagen bilden alle aktuell eingeschriebenen Studierenden die Grundgesamtheit. Aus dieser Grundgesamtheit wird zufällig eine Stichprobe gezogen (z. B. 1000 Studierende). Jeder Studierende hat bei der Stichprobenziehung die gleiche Chance in die Stichprobe zu gelangen. Für die empirische Analyse (z. B. Zusammenhänge

Gütesiegel ohne Wert: Repräsentativität

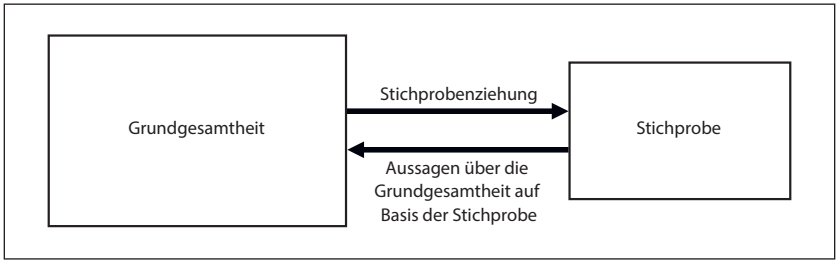
Nach Kruskal und Mosteller (1979a, b, c) gibt es mindestens neun unterschiedliche Bedeutungen des Begriffs „Repräsentativität". In den Medien wird der Begriff der Repräsentativität gelegentlich als Gütesiegel für eine Stichprobe verwendet. Häufig ist von repräsentativen Stichproben oder einem repräsentativen Querschnitt die Rede. Im Alltag ist damit ein „verkleinertes Abbild der Bevölkerung" gemeint. In der Fachliteratur (z. B. Diekmann 2011, S. 430–432; Schnell et al. 2013, S. 296–298) wird der Begriff der Repräsentativität sehr zurückhaltend verwendet. Wie Diekmann (2011, S. 430) überzeugend argumentiert, kann eine Stichprobe niemals sämtliche Merkmalsverteilungen der Grundgesamtheit „repräsentieren". Bei Repräsentativität handelt es sich auch nicht um einen Fachbegriff, sondern eher um eine Metapher. Da die Bezeichnung „Repräsentativität" weder rechtlich geschützt noch an bestimmte Kriterien gebunden ist, eignet sich dieser Begriff nicht als Gütesiegel für Stichproben (Lippe und Kladroba 2002; Gabler und Quatember 2012; Quatember 2015, S. 151–181). Für die Bewertung der Güte einer Stichprobe sind detaillierte Informationen (z. B. Art der Stichprobenziehung, Ausschöpfungsquote, Angaben über Ausfälle) erforderlich.

bestimmter Merkmale) wird die Stichprobe mit den 1000 Studierenden genutzt. Die Zufallsstichprobe steht gewissermaßen stellvertretend für die Grundgesamtheit. Allerdings sind Sozialwissenschaftlerinnen weniger an Aussagen über die Stichprobe interessiert, sondern vielmehr an Aussagen über die Grundgesamtheit.[35] Die Inferenzstatistik bietet entsprechende Verfahren, die es erlauben, die Aussagen von der Stichprobe auf die Grundgesamtheit zu übertragen (innerhalb gewisser Fehlertoleranzen). Allerdings setzen diese Verfahren zwingend eine Zufallsauswahl voraus (Diekmann 2011, S. 380). Deshalb stehen viele Sozialwissenschaftler sogenannten Quoten-Verfahren skeptisch gegenüber (siehe Abschnitt 10.3.2.2).

35 Die bekanntesten Beispiele für deutschlandweite Studien sind sicherlich der Deutschland-Trend (ARD) und das Politbarometer (ZDF). In beiden Erhebungen werden auf Basis von Stichproben Aussagen über die wahlberechtigte Bevölkerung in Deutschland getroffen. In beiden Erhebungen gehören Bürgerinnen und Bürger unter 18 Jahren nicht zur Grundgesamtheit.

Abbildung 26 stellt die Logik der einfachen Zufallsstichprobe dar. Aus einer Grundgesamtheit wird eine Stichprobe gezogen. Jedes Element der Grundgesamtheit hat die gleiche Chance in die Stichprobe zu gelangen. Die Stichprobe bildet die Grundlage für die empirische Analyse. Auf Basis inferenzstatistischer Berechnungen können die Befunde der Stichprobe innerhalb bestimmter Fehlergrenzen auf die Grundgesamtheit übertragen werden.

Abbildung 26 Einfache Zufallsstichprobe



Quelle: Eigene Darstellung

Die einfache Zufallsstichprobe setzt eine Liste aller Elemente der Grundgesamtheit voraus. Bei bestimmten Fragestellungen ist eine Liste sicherlich ohne größere Probleme verfügbar (z. B. Mitarbeiterbefragung in einer Firma, Kundenbefragung), bei vielen sozialwissenschaftlichen Fragestellungen existiert eine solche Liste allerdings nicht (z. B. Befragung der wahlberechtigten Bürger in der Bundesrepublik Deutschland). Deshalb spielt in der empirischen Praxis die einfache Zufallsauswahl eine untergeordnete Rolle. Meist werden mehrstufige Auswahlverfahren verwendet.

Geschichtete Zufallsstichprobe Bei einer geschichteten Zufallsstichprobe wird die Grundgesamtheit zunächst in Untergruppen (Schichten) aufgeteilt und dann für jede dieser Gruppen eine Zufallsstichprobe gezogen. Solche Schichtenstichproben werden angewendet, wenn bei einer Untersuchung sichergestellt werden soll, dass aus jeder Schicht ausreichend Beobachtungen vorhanden sind. Die Logik der Schichtenstichprobe lässt sich an unserem Beispiel der Befragung der Studierenden illustrieren. Die Zahl der Studierenden variiert erheblich zwischen den Fakultäten der FernUniversität in Hagen. Die Fakultäten bilden jeweils eine Schicht. Innerhalb der Schicht werden dann zufällig die Studierenden für die Erhebung ausgewählt. Ein anderes Beispiel: Bei einer Erhebung der Europäischen

Union bilden die Mitgliedstaaten jeweils eine Schicht. Innerhalb der Schicht werden dann zufällig die einzelnen Elemente ausgewählt.

- **Definition: Geschichtete Zufallsstichprobe**
  „Geschichtete Zufallsstichproben werden gezogen, indem die Elemente der Grundgesamtheit so in Gruppen (‚Schichten', ‚strata') eingeteilt werden, dass jedes Element der Grundgesamtheit zu einer – und nur zu einer – Schicht gehört und dann einfache Zufallsstichproben aus jeder Schicht gezogen werden" (Schnell et al. 2013, S. 269).

Bei geschichteten Zufallsstichproben lassen sich zwei Varianten unterscheiden: Die proportional geschichtete Stichprobe und die disproportional geschichtete Stichprobe. Bei der proportional geschichteten Stichprobe wird bei der Stichprobenziehung der Anteil der Elemente einer Schicht an der Grundgesamtheit berücksichtigt. Bei einer Stichprobenziehung der Studierenden der FernUniversität in Hagen hätten Studierende der Fakultät für Kultur- und Sozialwissenschaften höhere Auswahlchancen als Studierende der Fakultät für Rechtswissenschaften, da der Anteil der Studierenden der Fakultät für Kultur- und Sozialwissenschaften in der Grundgesamtheit höher ist. Entspricht die Fallzahl der Schichten nicht den jeweiligen Anteilen in der Grundgesamtheit, dann handelt es sich um eine disproportional geschichtete Stichprobe.

Klumpenstichprobe Die einfache Zufallsauswahl setzt eine Liste der Elemente der Grundgesamtheit voraus. Bei einer Befragung der Studierenden der FernUniversität in Hagen wäre dies beispielsweise eine Namensliste (vorzugsweise mit Postadresse oder gültiger E-Mail-Adresse). Auf Basis dieser Namensliste kann dann eine Zufallsauswahl realisiert werden. In vielen Fällen existiert eine solche Namensliste nicht. Bei einer geplanten Befragung der Schüler der achten Klassen in einer Gemeinde verfügt die Stadtverwaltung in der Regel nicht über eine Namensliste der betreffenden Schüler. Allerdings liegt der Verwaltung eine Liste der weiterführenden Schulen vor. Von beispielsweise 20 Schulen werden sieben Schulen zufällig ausgewählt. In diesen sieben Schulen werden dann alle Schülerinnen der achten Klassen befragt.

- **Definition: Klumpenstichprobe**
  „Als ‚Klumpenstichprobe' (‚cluster sample') wird eine einfache Zufallsstichprobe dann bezeichnet, wenn die Auswahlregeln nicht auf die Elemente der Grundgesamtheit, sondern auf zusammengefasste Elemente (‚Klumpen', Cluster) angewendet werden und jeweils die Daten aller Elemente eines Clusters erhoben werden" (Schnell et al. 2013, S. 270).

Die PISA-Studie basiert auf einer solchen Klumpenstichprobe. Es werden nicht die einzelnen Schüler, sondern die Schulen ausgewählt. Innerhalb der Schulen werden dann alle 15-jährigen Schüler befragt.[36] Bei einer solchen Stichprobenziehung stellen die Schulen die Einheiten dar, auf die sich der Auswahlplan konkret bezieht (Auswahleinheiten). Die Schüler sind die Erhebungseinheiten, bei denen die Informationen konkret erhoben werden.

Bei solchen Klumpenstichproben sind „Verzerrungen", sogenannte Klumpeneffekte, häufig unvermeidlich. Die Elemente eines „Clusters" (hier: Schüler in den ausgewählten Schulen) sind ähnlicher als die Elemente zwischen den Klumpen. Klumpenstichproben sind ungenauer als einfache Zufallsstichproben, wenn sich die Elemente eines Clusters stark ähneln und die Cluster sich stark voneinander unterscheiden (Beispiel: private versus staatliche Schulen). „Verzerrungen" auf Basis der Stichprobenziehung werden allgemein als Design-Effekt bezeichnet. Solche Design-Effekte lassen sich allerdings mit komplexen mathematischen Verfahren berechnen bzw. korrigieren.

Mehrstufige Auswahlverfahren In der Praxis führen Kostenerwägungen und Beschränkungen bei der Durchführbarkeit zu anderen Stichprobenziehungen als die einfache Zufallsauswahl (Kohler und Kreuter 2017, S. 225). Für viele interessante Grundgesamtheiten (z. B. die wahlberechtigte Bevölkerung einer Bundestagswahl) steht auch keine (zentrale) Liste aller Elemente dieser Grundgesamtheit zur Verfügung. Deshalb basieren die meisten sozialwissenschaftlichen Erhebungen (z. B. ALLBUS, ESS) auch nicht auf einer einfachen Zufallsstichprobe, sondern auf einem mehrstufigen Auswahlverfahren. In Abschnitt 10.4 werden typische Auswahlverfahren in der Forschungspraxis dargestellt. Zuvor werden allerdings die nichtzufälligen Auswahlverfahren präsentiert.

36 Um die Vergleichbarkeit der Ergebnisse in den verschiedenen Ländern zu gewährleisten, wird die PISA-Studie in vergleichbaren Zielpopulationen durchgeführt. Vor dem Hintergrund unterschiedlicher schulorganisatorischer Regelungen, beispielsweise Einschulungsalter oder Struktur des Bildungssystems, sind Definitionen der Zielpopulationen, die sich auf bestimmte Klassenstufen beziehen, für internationale Vergleiche ungeeignet. Deshalb wird die Zielpopulation von PISA durch das Lebensalter bestimmt. Bei der PISA-Studie werden Schüler untersucht, die zum Beginn des Testzeitraums zwischen 15 Jahren/drei Monaten und 16 Jahren/zwei Monaten alt waren – unabhängig von der besuchten Jahrgangsstufe oder Art der Bildungseinrichtung (Baumert et al. 2001).

10.3.2 Nichtzufällige Auswahlverfahren

Die nichtzufälligen Auswahlverfahren lassen sich in zwei Gruppen unterteilen: „willkürliche Auswahlen“ und „bewusste Auswahlen“. Die willkürliche Auswahl wird als „Auswahl aufs Geratewohl“ (Löffler und von der Heyde 2014, S. 20) bezeichnet. Für wissenschaftliche Zwecke ist dieses Verfahren völlig ungeeignet. Bei einer bewussten Auswahl werden die Fälle auf Basis theoretischer Überlegungen ausgewählt. Es existieren Kriterien der Fallauswahl und diese werden explizit offengelegt. Die bewusste Auswahl basiert allerdings nicht auf einem Zufallsmechanismus. Inferenzstatistische Verfahren können nicht angewendet werden. Deshalb sind – je nach Fragestellung – auch keine Rückschlüsse auf eine Grundgesamtheit möglich.

10.3.2.1 Willkürliche Auswahlen

! Mit willkürlichen Auswahlen werden Leserinnen von Tageszeitungen oder Zuschauer von „politischen“ Gesprächsrunden tagtäglich konfrontiert. Das typische Beispiel sind Straßeninterviews, bei denen willkürlich irgendwelche Personen befragt werden. Die Aussagen dieser Personen werden dann in Tageszeitungen veröffentlicht oder per „Einspieler“ in einer Fernseh- oder Hörfunksendung präsentiert. Häufig werden solche Befragungen auch mit folgenden Worten eingeleitet: „Wir haben einmal Personen in der Innenstadt befragt.“ Solche Befragungen dienen bestenfalls der Illustration oder der Unterhaltung. Für wissenschaftliche Zwecke sind solche Befragungen völlig wertlos. Weder ist die Grundgesamtheit sinnvoll definiert, noch ist vor der Stichprobenziehung für jedes Element der Grundgesamtheit die Auswahlwahrscheinlichkeit sinnvoll angebbar (Schnell et al. 2013, S. 289; siehe auch Akremi 2014, S. 272–273).

- **Definition: Willkürliche Auswahl**
  „Willkürliche Auswahlen (‚Auswahlen aufs Geratewohl‘) sind Auswahlen, bei denen die Entscheidung über die Aufnahme eines Elements der Grundgesamtheit in die Stichprobe unkontrolliert durch einen Auswahlplan nur im Ermessen des Auszuwählenden liegt“ (Schnell et al. 2013, S. 289). Wird eine Stichprobe nach dem Prinzip der Verfügbarkeit zusammengestellt, so wird diese Stichprobe in der englischen Fachliteratur als „Convenience Sample“ bezeichnet.

Eine willkürliche Auswahl erlaubt keinen Rückschluss auf eine Grundgesamtheit, da die Grundgesamtheit nicht festgelegt ist. Es werden willkürlich Passanten befragt und es völlig unklar, wen diese Passanten repräsentieren sollen. Solche Be-

fragungen sind im besten Fall unterhaltsam, aber niemals für wissenschaftliche Aussagen geeignet.

Auswahl der Versuchspersonen bei Experimenten

Die Teilnehmer von Experimenten (z. B. in der Psychologie) werden praktisch niemals per Zufallsauswahl ausgewählt. Im Gegenteil: Es handelt sich in der Regel um höchst selektive Stichproben (z. B. Henrich et al. 2010; Kubbe 2016, S. 116–118). Die meisten Probanden bei (psychologischen) Experimenten sind Psychologiestudierende der ersten Fachsemester, die per Studienordnung verpflichtet sind, an Experimenten teilzunehmen. Was können wir aus solchen Experimenten dann aber überhaupt lernen? In Experimenten wird nicht etwa der Anteil der CDU-Wähler bei der nächsten Bundestagswahl ermittelt oder die Zufriedenheit der Bürger mit der Europäischen Union untersucht. Experimente zielen darauf ab, kausale Mechanismen zu entdecken. Einigen Personen wird dazu ein Treatment verabreicht, anderen Personen nicht (siehe Kapitel 9). Dabei werden die Teilnehmer eines Experiments zwar nicht per Zufallsauswahl ausgewählt, aber die Verteilung in Experimental- oder der Kontrollgruppe erfolgt zufällig (Prinzip der Randomisierung). Nach der Verabreichung des Treatments an die Teilnehmer der Experimentalgruppe werden die Reaktionen der Experimental- und der Kontrollgruppe verglichen. Zeigen die Teilnehmenden der Experimentalgruppe eine andere Reaktion als die Personen der Kontrollgruppe, dann wird dies auf das Treatment zurückgeführt. Experimente müssen zu unterschiedlichen Zeiten, an unterschiedlichen Orten und mit verschiedenen Personen wiederholt werden, um allgemeingültige Aussagen machen zu können (siehe für zahlreiche Vorschläge z. B. Henrich et al. 2010). Mit der Meta-Analyse existiert ein statistisches Verfahren, mit dem die Ergebnisse von mehreren Experimenten „gemittelt“ werden können (zur Methode siehe z. B. Borenstein et al. 2009; für eine Anwendung siehe z. B. Hyde 2005).

10.3.2.2 Bewusste Auswahlen

Im Gegensatz zur willkürlichen Auswahl werden bei der bewussten Auswahl vor der Auswahl systematische Regeln formuliert, die jedoch nicht auf einem Zufallsprozess, sondern auf inhaltlichen Kriterien beruhen. Bewusste Auswahlen sind

zwar „nicht geeignet, um Befunde auf die Grundgesamtheit zu verallgemeinern, können jedoch unter bestimmten Bedingungen zur Theoriebildung oder Falsifikation beitragen“ (Westle 2009b, S. 170). In der Fachliteratur werden verschiedene Varianten der bewussten Auswahl unterschieden (für eine Übersicht siehe z. B. Westle 2009b, S. 170–174; Schnell et al. 2013, S. 290–296; Löffler und von der Heyde 2014, S. 20). An dieser Stelle werden die Auswahl entscheidender Fälle und die Quotenauswahl sowie das Most Similar Cases Design (MSCD) und Most Different Cases Design (MDCD) vorgestellt.

Auswahl entscheidender Fälle Die Logik der Auswahl entscheidender Fälle basiert auf den Grundannahmen des kritischen Rationalismus (Popper 1982 [1934]). Danach sind Aussagen niemals verifizierbar, weil Wahrheit prinzipiell nicht er-

Klassiker der Sozialwissenschaften: Zur Soziologie des Parteiwesens

Robert Michels (1989 [1911]) Studie „Zur Soziologie des Parteiwesens in der modernen Demokratie. Untersuchungen über die oligarchischen Tendenzen“ gilt als Klassiker der Parteiensoziologie (für eine Würdigung siehe z. B. Walter und Klatt 2014). Michels untersucht in seinem Werk den Zusammenhang zwischen Demokratie, Parteien und Eliten. Seine zentrale These besagt, dass jede Organisation oligarchische Tendenzen aufweist (Michels 1989 [1911], S. 12). Die empirische Untersuchung basiert allerdings nicht etwa auf einer Zufallsstichprobe aller Organisationen, sondern auf einem Einzelfall – der SPD im Kaiserreich. Trotz dieser bewussten Auswahl kann Michels seine zentrale These über den Einzelfall hinaus generalisieren. Schließlich hat er nicht irgendeine Organisation ausgewählt – beispielsweise eine Organisation, bei der oligarchische Tendenzen zu erwarten waren (z. B. Militär, Kirche). Im Gegenteil: Die sozialdemokratische Partei galt im Kaiserreich als Negation oligarchischer Tendenzen. Die Beobachtung oligarchischer Tendenzen sogar bei der SPD galt lange Zeit als Beleg für die Existenz oligarchischer Tendenzen in jeder menschlichen Organisation. Mit anderen Worten: Michels hat seine These „Wer Organisation sagt, sagt Tendenz zur Oligarchie“ am „unwahrscheinlichsten“ Fall nachweisen können. Deshalb wird der Zusammenhang zwischen Organisation und Oligarchie häufig auch als „Ehernes Gesetz der Oligarchie“ oder „Eisernes Gesetzes der Oligarchie“ bezeichnet. Erst die Studie „Union Democracy“ (Lipset et al. 1956) hat die allgemeingültige Aussage von Michels falsifiziert.

reichbar ist. Allerdings sind Aussagen falsifizierbar. Es kann bei einer empirischen Untersuchung geprüft werden, ob eine Aussage zutrifft oder nicht. Nach der Formulierung einer empirisch überprüfbaren Hypothese wird bei der Auswahl entscheidender Fälle nach Untersuchungseinheiten „gesucht", die der Hypothese widersprechen (Westle 2009b, S. 171). Die formulierte Hypothese muss sich folglich unter den „unwahrscheinlichsten" Bedingungen bestätigen, um vorläufig akzeptiert zu werden.

> *„Das wissenschaftliche Vorgehen sollte nicht darauf gerichtet sein, die eigenen Vermutungen zu bestätigen. Vielmehr sollte man sich die größte Mühe geben, Tatsachen zu finden, die im Widerspruch zu den eigenen Vermutungen stehen, da man nur so eine Chance hat, sich der Wahrheit anzunähern." (Behnke und Behnke 2006, S. 91–92)*

Werden solche Gegenbeispiele entdeckt, dann muss die Hypothese umformuliert und/oder präzisiert werden. Die neuformulierte Hypothese wird dann erneut einer empirischen Prüfung unterzogen.

In der Forschungspraxis ist allerdings eine stärkere Fokussierung auf die Hypothesenbestätigung zu beobachten. Sozialwissenschaftler möchten ihre Hypothesen „bestätigen" (verifizieren) anstatt sie zu „widerlegen" (falsifizieren). Dies ist mit Blick auf den kritischen Rationalismus problematisch und wird mittlerweile auch kritisch diskutiert (Lehrer et al. 2007).

Quotenauswahl Die Quotenauswahl – häufig auch Quotenverfahren oder Quotenmethode genannt – ist das am häufigsten nicht-zufällige Auswahlverfahren in der Marktforschung (Meier und Hansen 2014, S. 197). Das Institut für Demoskopie in Allensbach nutzt die Quotenauswahl auch in der Wahlforschung (für eine ausführliche Darstellung siehe Noelle-Neumann und Petersen 2005, S. 255–281). In der Methodenliteratur gibt es eine leidenschaftliche Debatte, ob auf Basis der Quotenauswahl sogenannte „repräsentative" Stichproben realisiert werden können (siehe für Pro-Argumente z. B. Noelle-Neumann und Petersen 2005, S. 263–281; siehe für Contra-Argumente z. B. Westle 2009b, S. 172–174; Diekmann 2011, S. 390–398; Schnell et al. 2013, S. 292–296). Worum handelt es sich nun aber bei der Quotenauswahl?

Durch amtliche Statistiken (z. B. Volksbefragung, Zensus) stehen mehr oder weniger genaue Angaben über die Bevölkerung zur Verfügung.[37] Nach den Anga-

37 Beim Zensus 2011 handelt es sich um eine Stichprobe. Dabei wurden Auszüge aus verschiedenen Registern zusammengetragen und eine Haushaltsbefragung durchgeführt, bei der etwa zehn Prozent der Bevölkerung befragt wurden. Umfangreiche Informationen zum Zensus bietet die Homepage zum Zensus 2011 unter www.zensus2011.de.

ben des Zensus 2011 leben in Deutschland 80 219 695 Bürger. Davon sind 48,8 Prozent männlich und 51,2 Prozent weiblich. 39,9 Prozent der Bürger in Deutschland sind ledig, 45,8 Prozent sind verheiratet (inklusive eingetragene Lebensgemeinschaft) und sieben Prozent sind geschieden. Von den 80 219 695 Bürgern haben 92,3 Prozent die deutsche Staatsbürgerschaft. Diese Informationen werden bei der Quotenauswahl genutzt, um eine Stichprobe zu konstruieren, die möglichst exakt mit den (bekannten) Merkmalen (Quote) der (angestrebten) Grundgesamtheit übereinstimmt. Das Ziel besteht darin, eine Stichprobe zu erhalten, die sich ebenfalls aus 48,8 Prozent männlichen und 51,2 Prozent weiblichen Personen zusammensetzt. Selbstverständlich können bei einer solchen Stichprobe nicht alle Merkmale der Grundgesamtheit berücksichtigt werden. Am häufigsten werden Geschlecht, Alter, Bildung und Berufstätigkeit verwendet. Je nach Fragestellung können auch andere Merkmale berücksichtigt werden.

Der Interviewer, der die eigentliche Befragung durchführt, ist an der Festlegung der Quote bzw. der Stichprobe nicht beteiligt. Das Institut erstellt ein bestimmtes Anforderungsprofil desjenigen, der befragt werden soll. Der Interviewer soll beispielsweise zwei männliche Bürger zwischen 35 und 40 Jahren mit abgeschlossenem Studium befragen. Die konkrete Auswahl der Befragten bleibt dem Interviewer überlassen. Während bei einer Zufallsstichprobe sämtliche subjektiven Einflüsse bei der Auswahl der Befragten ausgeschaltet werden, entscheidet beim Quotenverfahren der Interviewer auf Basis vorgegebener Merkmale, wen er befragt. Die konkrete Auswahl der Befragungspersonen obliegt also den Interviewern, die aber in ihrer Entscheidung durch die Quoten eingeschränkt werden. Elisabeth Noelle-Neumann und Thomas Petersen (2005) erläutern das Ziel der Quotenauswahl:

> *„Die wirkliche Funktion der Quoten ist: sie sollen den Interviewer zu einer Zufallsauswahl veranlassen, bei der jedes Mitglied der Grundgesamtheit praktisch die gleiche Chance hat, in die Stichprobe zu gelangen." (Noelle-Neumann und Petersen 2005, S. 258)*

! Dieses Ziel, so entgegnen die Kritiker des Quotenverfahrens, wird durch eine Quotenauswahl aber vermutlich nicht erreicht. Die Interviewer werden verständlicherweise den einfachsten Weg wählen, um ihre Quote zu erfüllen. Es werden also in erster Linie Freunde und Bekannte und möglicherweise wieder deren Bekannte interviewt. In erster Linie werden damit Personen befragt, die leicht erreichbar sind (Diekmann 2011, S. 395; Schnell et al. 2013, S. 293). Das zentrale Kriterium der Zufallsstichprobe – jedes Element hat die gleiche Chance in die Stichprobe zu gelangen – wird damit verletzt. Die Befürworter des Quotenverfahrens verweisen im Gegenzug darauf, dass es auch bei einer Zufallsstichprobe Teilnahmeverweigerungen gibt.

„In der Praxis erfüllt auch die Random-Stichprobe nicht die Anforderungen der Wahrscheinlichkeitstheorie." (Noelle-Neumann und Petersen 2005, S. 268)

Die zufällige Auswahl der Elemente einer Grundgesamtheit stellt sicherlich das Ideal der Zufallsstichprobe dar. Alle Bemühungen zielen letztendlich darauf ab, dieses Ideal – also die Zufallsauswahl – bestmöglich zu realisieren. Beim Quotenverfahren wird dieses Ideal allerdings aufgegeben, da die Interviewerinnen einen mehr oder weniger großen Ermessensspielraum hat, wer befragt wird (Befürworter der Quotenauswahl stimmen dieser Aussage sicherlich nicht zu). Die Anwendung der Inferenzstatistik, also der Rückschluss von der Stichprobe auf die Grundgesamtheit, ist beim Quotenverfahren zumindest „erschwert, falls nicht unmöglich" (Schnell et al. 2013, S. 295).

Meier und Hansen (2014, S. 204) nennen neun – nicht weiter begründete – Bedingungen, wie Verzerrungen bei Quotenstichproben zu vermeiden sind. Dazu zählt, dass die Befragungen auf möglichst viele Interviewer verteilt werden, sodass jeder Interviewer nur wenige Befragungen durchführt. Die Quotenanweisung darf dem Interviewer keinen Spielraum für Interpretationen geben; die Erhebung ist zentral zu leiten. Dazu gehören regelmäßige Kontrollen der Einhaltung der Anweisungen und der korrekten Durchführung der Interviews. Zudem sollten die Interviews zu einem hohen Prozentsatz in den Wohnungen der Befragten durchgeführt werden, da die Kontaktierung auf der Straße zu einer Überrepräsentation des aktiven Teils der Bevölkerung führen.

MSCD und MDCD Bei einer Untersuchung von Staaten oder auch kleineren Gebietseinheiten (z. B. Bundesländern, Gemeinden) lassen sich in der vergleichenden Sozialforschung zwei Grundtypen der Fallauswahl unterscheiden: Most Similar Cases Design (MSCD) und Most Different Cases Design (MDCD). Beide Typen gehen auf die Methoden der Differenz und Konkordanz von John Stuart Mill zurück, die in den Arbeiten von Przeworski und Teune (1970) sowie Lijphart (1971, 1975) weiterentwickelt wurden (für eine ausführlichere Darstellung siehe z. B. Jahn 2011, S. 76–81; Lauth et al. 2015, S. 60–69; Pickel 2016).

Bei einem Most Similar Cases Design (MSCD) werden Untersuchungseinheiten (z. B. Staaten) ausgewählt, die sich in zahlreichen Eigenschaften ähnlich sind, aber sich hinsichtlich der eigentlich interessierenden unabhängigen und abhängigen Variablen unterscheiden. Ein Beispiel: Es soll der Zusammenhang zwischen Wahlpflicht (uV) und Wahlbeteiligung (aV) auf der Makroebene untersucht werden. Selbstverständlich müssen Länder mit und Länder ohne Wahlpflicht ausgewählt werden, um einen Zusammenhang zwischen Wahlpflicht und Wahlbeteiligung überhaupt untersuchen zu können. Neben der Wahlpflicht gibt es aber natürlich noch andere Merkmale auf der Makroebene, die einen Einfluss auf die

Forschungsprojekt: Europa im Kontext

Das Forschungsprojekt „Europa im Kontext“ untersucht den Einfluss des lokalen Umfelds auf politische Einstellungen und Verhaltensweisen (van Deth und Tausendpfund 2013; Tausendpfund 2013b; Tausendpfund und Vetter 2017). Für die empirischen Analysen wurden über 12 000 Personen in 28 Gemeinden befragt. Die Auswahl der Personen und der Gemeinden erfolgte zufällig, aber der Forschungsentwurf entspricht einem Most Similar Cases Design. In jeder Gemeinde sind die Bürgerinnen und Bürger einer unendlichen Reihe von Einflussfaktoren ausgesetzt, so unterscheiden sich beispielsweise die Gemeindeordnungen zwischen den Bundesländern. Durch die Begrenzung auf ein Bundesland unterscheiden sich die Gemeinden im Hinblick auf das lokale Umfeld, die überregionalen Einflussgrößen – insbesondere die institutionellen Besonderheiten – sind jedoch für alle Kommunen konstant. Die Festlegung auf hessische Gemeinden bietet zudem den Vorteil, dass es innerhalb des Bundeslands erhebliche Disparitäten in der Bevölkerungsdichte und der wirtschaftlichen Struktur gibt. Die im Rahmen des Forschungsprojekts zufällig ausgewählten hessischen Gemeinden unterscheiden sich zwar im Hinblick auf den lokalen Kontext (z. B. Arbeitslosenquote, parlamentarische Fragmentierung, Ausländeranteil), aber die überregionalen Einflussfaktoren sind für alle Gemeinden gleich.

Wahlbeteiligung haben können (z. B. Wahlsystem, Regierungssystem, politische Kultur). Bei einem MSCD-Design sollten sich die ausgewählten Länder möglichst nur im Hinblick auf die interessierende unabhängige Variable (hier: Wahlpflicht) unterscheiden, alle anderen möglichen Einflussfaktoren (z. B. Wahlsystem, Regierungssystem, politische Kultur) sollten möglichst ähnlich sein. Beschränkt man beispielsweise die Fallauswahl auf westeuropäische Länder (z. B. Luxemburg und Belgien (jeweils Wahlpflicht) sowie Deutschland und Frankreich (jeweils keine Wahlpflicht)) dann ähneln sich diese Länder stärker als wenn beispielsweise auch südamerikanische Länder berücksichtigt werden (z. B. Bolivien, Brasilien). Ein MSCD-Design strebt also eine möglichst homogene Fallauswahl an. Die Untersuchungseinheiten sollten sich nur im Hinblick auf die eigentlich zu untersuchende uV und die aV unterscheiden.

Bei einem Most Different Cases Design (MDCD) werden möglichst unterschiedliche Untersuchungseinheiten (z. B. Staaten) für den Vergleich ausgewählt. Ausgangspunkt ist meist eine Individualhypothese, zum Beispiel „Postmaterialisten wählen eher grüne Parteien“. Dieser Zusammenhang wird in einem Kontext

(z. B. Deutschland) untersucht. Wird der Zusammenhang bestätigt, folgt ein weiterer Test. Ideal wäre, wenn sich der Zusammenhang in möglichst vielen Systemen bestätigen würde. Aus forschungspraktischer Perspektive kann der Zusammenhang allerdings nicht in allen Kontexten und zu allen Zeitpunkten untersucht werden. Anstatt den bestätigten Zusammenhang also in einem ähnlichen Kontext einem weiteren Test zu unterziehen, sollte sich die Beziehung in möglichst unterschiedlichen Staaten bewähren. Bei einem MDCD-Design würde der Zusammenhang nicht in Frankreich oder Österreich untersucht werden, sondern möglicherweise in Bolivien oder Burkina Faso.

! Das MSCD- und das MDCD-Design skizzieren Idealvorstellungen von Fallauswahlen, die in der Forschungspraxis nur selten vollständig realisiert werden können. Häufig basiert die Fallauswahl auch auf Variationen und/oder Kompromissen (z. B. verfügbares Datenmaterial). Entsprechend sollte die Fallauswahl sorgfältig begründet werden und nicht durch Zufälligkeiten oder gar persönliche Vorlieben gekennzeichnet sein (Lauth et al. 2015, S. 67).

10.4 Auswahlverfahren in der Forschungspraxis

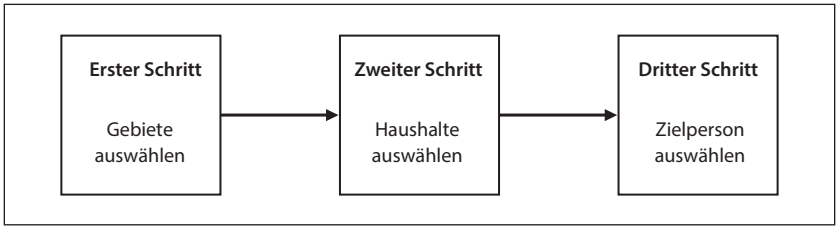
In der Forschungspraxis basieren die Auswahlverfahren für Personen in den seltensten Fällen auf der einfachen Zufallsstichprobe. Dafür gibt es einen einfachen Grund: Meist existiert keine vollständige „Liste" der Grundgesamtheit. Weder das Statistische Bundesamt noch eine andere Organisation verfügt beispielsweise über eine komplette (und vor allem aktuelle) Liste der Einwohner der Bundesrepublik Deutschland oder der wahlberechtigten Bevölkerung in Deutschland. Deshalb basieren Stichprobenziehungen in der Forschungspraxis meist auf mehrstufigen Auswahlverfahren. In diesem Abschnitt werden drei klassische Auswahlverfahren für Befragungen vorgestellt: Das ADM-Design für persönliche Befragungen, das Gabler-Häder-Design für telefonische Befragungen sowie Registerstichproben.

10.4.1 ADM-Design für persönliche Befragungen

ADM ist die Abkürzung für „Arbeitskreis Deutscher Markt- und Sozialforschungsinstitute e. V." (www.adm-ev.de). Als Wirtschaftsverband vertritt der ADM die Interessen der privatwirtschaftlichen Markt- und Sozialforschungsinstitute in Deutschland. Nach eigenen Angaben erzielen die Mitgliedsinstitute des ADM über 80 Prozent des Branchenumsatzes. Der Arbeitskreis stellt seinen Mitgliedern ein Stichprobensystem zur Verfügung, das diesen erlaubt, zufällige Stich-

proben der Privataushalte und der darin wohnenden Personen in Deutschland zu erstellen (eine ausführliche Darstellung bieten Heckel und Hofmann 2014; siehe auch Häder und Häder 2014, S. 289–292). Dieses Design gilt in Deutschland als „good practice" (Häder und Häder 2014, S. 290). Wie Abbildung 27 illustriert, lassen sich beim ADM-Design für persönliche Befragungen drei Auswahlschritte unterscheiden.

Abbildung 27 ADM-Design für persönliche Befragungen



Quelle: Eigene Darstellung

Bei allen drei Schritten wird eine zufällige Auswahl angestrebt, um auf Basis inferenzstatistischer Methoden Rückschlüsse von der Stichprobe auf die Grundgesamtheit ziehen zu können.

Erster Schritt: Gebiete auswählen Der erste Schritt des ADM-Designs für persönliche Befragungen ist die Festlegung des Gebiets. Das bewohnte Gebiet der Bundesrepublik Deutschland wird durch möglichst gleich große und in der inneren Struktur möglichst homogene Teilflächen abgebildet. Diese Teilflächen werden Sample Points bezeichnet.

> *„Für die Bildung der Sample Points verwendet man die zur Verfügung stehenden kleinsten administrativen Flächeneinheiten bis auf Baublockebene. Dazu können inzwischen amtliche innerstädtische Gliederungen und Sachdaten systematisch herangezogen werden." (Häder und Häder 2014, S. 290)*

Beim ADM-Auswahlrahmen werden knapp 53 000 solcher Teilflächen bzw. Sample Points unterschieden. Das heißt, das Gebiet der Bundesrepublik Deutschland wird in knapp 53 000 Teilflächen gegliedert. Auf jeden Sample Point entfallen durchschnittlich 750 Haushalte (Heckel und Hofmann 2014, S. 96). Diese Teilflächen bilden die Basis für die Konstruktion von Stichprobennetzen. Ein gesamt-

deutsches Stichprobennetz setzt sich aus 258 Sample Points zusammen. 202 dieser Sample Points befinden sich in den alten Bundesländern, 43 Sample Points in den neuen Bundesländern und 13 Sample Points in Berlin (für eine ausführliche Darstellung siehe Heckel und Hofmann 2014). Es werden zahlreiche Stichprobennetze realisiert, die den Mitgliedern des ADM für die eigene Arbeit überlassen werden. Die einzelnen Sample Points sind räumlich klar definiert, sodass im zweiten Schritt des Auswahlverfahrens die Zielhaushalte ermittelt werden können.

Zweiter Schritt: Haushalte auswählen Die Auswahl der Zielhaushalte innerhalb eines Sample Points erfolgt über Random Walk (oft auch Random Route genannt). Bei Random Walk handelt es sich um eine allgemeine Begehungsvorschrift:

> *„Das Prinzip des Random Walk basiert darauf, dass ein Interviewer zu einem vorgegebenen Startpunkt gebeten wird, von dem aus er mit Hilfe festgelegter Gehregeln (z. B. Hausnummernabwärts; bei Kreuzungen abbiegen und dabei die Straße überqueren; bei Sackgassen auf der gegenüberliegenden Seite umdrehen; etc.) in jedem x-ten (z. B. jeden dritten) Haushalt befragt." (Heckel und Hofmann 2014, S. 107)*

Mit der allgemeinen Begehungsvorschrift soll die Auswahl der Zielhaushalte durch den Interviewer möglichst eingeschränkt werden, sodass eine Zufallsauswahl der Haushalte entsteht. Für jede mögliche räumliche Situation muss eine klare Regelung existieren, damit der Interviewer zu keinem Zeitpunkt von der zufälligen Route abweicht. Die korrekte Einhaltung des Begehungsweges ist vom Befragungsinstitut regelmäßig zu überprüfen, um Abweichungen feststellen zu können (Heckel und Hofmann 2014, S. 108).

Bei Random Walk bzw. Random Route lassen sich zwei Varianten unterscheiden: Random Route mit Adressvorlauf und Random Route ohne Adressvorlauf. Bei einem Random Route mit Adressvorlauf übermittelt der Interviewer seine aufgelisteten Haushalte an das Ergebungsinstitut zurück. Die Liste umfasst eine zuvor festgelegte Anzahl von Haushalten, aus der dann eine zufällige Stichprobe gezogen wird. Diese Stichprobe wird anschließend von dem gleichen oder einem anderen Interviewer bearbeitet. Bei einem Random Route ohne Adressvorlauf versucht der Interviewer bereits bei der Begehung einen erfolgreichen Kontakt mit dem Zielhaushalt herzustellen und das Interview mit der Zielperson zu führen. Random Route ohne Adressvorlauf ist die kosten- und zeitgünstigere Variante, allerdings kann der Interviewer „ungünstige" Haushalte vorher aussortieren. Dann ist eine Zufallsauswahl der Haushalte nicht mehr gewährleistet. Die Kontrollmöglichkeiten des Instituts bei einem Random Route ohne Adressvorlauf sind deutlich eingeschränkt (Heckel und Hofmann 2014, S. 110).

Dritter Schritt: Zielperson auswählen Der dritte Schritt des ADM-Designs sieht die zufällige Auswahl der Zielperson im ermittelten Haushalt vor. Auch die eigentliche Zielperson muss zufällig ausgewählt werden, da sonst mobile Personen (z. B. Berufstätige, jüngere Personen) eine geringe Chance haben, in die Stichprobe zu gelangen. In der Forschungspraxis lassen sich zwei Varianten der Auswahl der Zielperson in einem Haushalt unterscheiden: die Geburtstagsmethode (Last-Birthday oder Next-Birthday) und der Schwedenschlüssel.

Bei der Geburtstagsmethode wird diejenige Person im Haushalt ausgewählt, die entweder als Letzte vor dem Kontaktdatum (Last-Birthday) oder als Nächste nach dem Kontaktdatum Geburtstag hat (Next-Birthday). Beim Schwedenschlüssel findet eine systematische Auflistung der einzelnen Haushaltsmitglieder durch den Interviewer statt (z. B. alle Haushaltsmitglieder werden nach dem Alter geordnet notiert). Die Auswahl der Zielperson erfolgt dann durch eine vom Befragungsinstitut erzeugte Zufallszahl; für jeden Zielhaushalt existiert eine eigene Zufallszahl (siehe Tabelle 14). Zunächst ermittelt der Interviewer die Anzahl der potenziellen Gesprächspartner (also die Personen im Haushalt, die zur Grundgesamtheit zählen). Diese Personen werden notiert. In unserem Muster-Haushalt handelt es sich um drei Personen. Unter der 3 im Schwedenschlüssel ist eine 2 aufgeführt. Mit der zweiten Person wird das Interview geführt. Selbstverständlich findet sich auf jedem Fragebogen ein neuer Schwedenschlüssel und die Zufallszahlen können nicht größer als die Haushaltsgröße sein. Der Schwedenschlüssel wird so genannt, da die Methode erstmalig in Schweden eingesetzt wurde. Das Verfahren hat Leslie Kish (1949) beschrieben, sodass der Schwedenschlüssel häufig auch als Kish Grid bezeichnet wird (Heckel und Hofmann 2014, S. 112).

Der Spielraum für mögliche Abweichungen von der Zufallsauswahl ist beim Schwedenschlüssel vermutlich etwas geringer als bei der Geburtstagsmethode. Deshalb wird diese Variante bei persönlich-mündlichen Befragungen bevorzugt (Häder und Häder 2014, S. 291). Falls die ermittelte Befragungsperson zum Zeitpunkt des Kontakts nicht anwesend ist, dann wird ein Termin vereinbart. Bei persönlichen Befragungen sind bis zu vier Kontaktversuche üblich (auch in den Abendstunden und an Samstagen). Keinesfalls darf die ermittelte Zielperson

Tabelle 14 Beispiel für einen Schwedenschlüssel

Haushaltsgröße	1	2	3	4	5	6
Auszuwählende Person	1	2	2	1	3	5

durch eine bereitwillige Befragungsperson ersetzt werden, da dies die Auswahlwahrscheinlichkeit mobiler Personen verringern würde.

Bei beiden Verfahren ist die Anzahl der zur Grundgesamtheit gehörenden Personen im Haushalt abzufragen. Denn die Auswahlwahrscheinlichkeit einer Person ist umso kleiner, je mehr potenzielle Zielpersonen zum Haushalt gehören. Bei einem Ein-Personen-Haushalt beträgt die Auswahlwahrscheinlichkeit 1, bei einem vier Personen-Haushalt nur noch 0,25. Die unterschiedlichen Auswahlwahrscheinlichkeiten sind allerdings mathematisch berechenbar und können bei der Datenanalyse berücksichtigt werden.

Probleme des ADM-Designs In einem Überblicksartikel haben Häder und Häder (2014, S. 292) auf die Grenzen und Fehlerquellen des ADM-Stichprobendesigns aufmerksam gemacht. Potenzielle Probleme lassen sich insbesondere beim zweiten (Haushalte auswählen) und dritten Schritt (Zielperson auswählen) feststellen. So kann der Interviewer erstens von Random Route abweichen und zweitens den Schwedenschlüssel bzw. die Geburtstagsmethode umgehen, um leichter die Interviews zu realisieren. Dies führt allerdings zu verzerrten Stichproben, da leicht erreichbare (und vor allem befragungswillige) Personen eine höhere Chance haben, in die Stichprobe zu gelangen. Die Einhaltung der Begehungsvorschrift bei Random Route und die korrekte Anwendung des Schwedenschlüssels bzw. der Geburtstagsmethode sind allerdings schwierig überprüfbare Fehlerquellen. In der Praxis wird die Arbeit der meist auf Honorarbasis tätigen Interviewer durch Stichproben kontrolliert. Dabei werden die Teilnehmer an einer Befragung gefragt, ob tatsächlich ein Interview geführt wurde, wie die Zielperson ausgewählt und welche Themen behandelt wurden (zu Fälschungen von Interviews siehe z. B. Blasius 2014). Die methodischen Defizite können durch Einwohnermeldeamtsstichproben zumindest verringert werden (siehe Abschnitt 10.4.3). Allerdings sind solche sogenannten Registerstichproben einerseits nicht in allen Fällen möglich und andererseits auch erheblich teurer als eine Realisierung der Stichprobe über Random Route und Schwedenschlüssel.

10.4.2 Stichprobendesigns für telefonische Befragungen

Für die Markt- und Sozialforschung sind telefonische Erhebungen von großer Bedeutung. Die ADM-Mitgliedsinstitute haben im Jahr 2013 über 21 Millionen Interviews durchgeführt. Der Anteil persönlich-mündlicher Interviews lag bei 22 Prozent; 36 Prozent waren telefonische Befragungen (ADM 2013, S. 22).[38] Vor allem durch ihre zentrale Funktion für repräsentative Erhebungen ist die herausragende Position von Computer Assisted Telephone Interviews (CATI) unstrittig.

Bei Telefonbefragungen sind Kontaktaufnahme mit dem Zielhaushalt sowie der Zielperson günstiger und schneller zu realisieren als bei persönlich-mündlichen Befragungen, da der Weg des Interviewers zum Zielhaushalt entfällt. Zudem sind weitere Kontaktversuche bei zunächst nicht erreichten Haushalten ebenfalls schneller und kostengünstiger zu realisieren. Darüber hinaus erlaubt die zentrale Erfassung der Antworten der Befragten auch Zwischenauswertungen.

Telefonstichproben in der Praxis

In der Regel werden die Zielpersonen bei einer telefonischen Befragung aus einem Telefonlabor angerufen. Eine Fachaufsicht (sogenannter Supervisor) kann bei Problemen sofort helfen und Korrekturen bei erkannten Schwächen der Stichprobenziehung vornehmen. Die Wahl der Rufnummer des Zielhaushalts erfolgt automatisiert, das Verhalten der Interviewer unterliegt ständiger Kontrolle. Zur Qualitätssicherung können Gespräche mitgeschnitten werden (Schnell 2012, S. 280–285; Meier et al. 2014, S. 119).

Ein zentraler Nachteil der telefonischen Erhebung ist offensichtlich: Wer keinen Telefonanschluss besitzt, der wird nicht in die Stichprobe aufgenommen. Eine Telefonbefragung ist daher erst bei einem gut entwickelten Fernmeldesystem möglich. In Deutschland war die Festnetzdichte erst zu Beginn der 1980er Jah-

38 Die verbliebenen 42 Prozent verteilen sich auf schriftliche Befragungen (sechs Prozent) und Online-Erhebungen (36 Prozent). Im Hinblick auf sogenannte „repräsentative“ Befragungen sind Online-Befragungen problematisch, da trotz steigender Zahlen nicht alle Bürger „online“ sind. Insbesondere ältere Menschen haben seltener Zugang zum Internet. Die jährlichen ARD-ZDF-Onlinestudien zeichnen ein sehr detailliertes Bild zur Entwicklung der Online-Nutzung in Deutschland (siehe auch Blasius und Brandt 2009; Schnell 2012, S. 287–290).

re ausreichend groß, um („repräsentative") Telefonbefragungen durchführen zu können (Meier et al. 2014, S. 122).[39]

Stichprobenziehung In Deutschland existiert kein allgemein zugängliches Verzeichnis aller privat genutzten Telefonanschlüsse. Eingetragene Telefonnummern in Telefonbüchern sind für eine Stichprobenziehung nicht geeignet, da diese Verzeichnisse einerseits unvollständig und andererseits auch sozial verzerrt sind. Deshalb ist es erforderlich, Telefonnummern zufällig zu generieren. Lange Zeit wurden die Rufnummern nach der Randomized-Last-Digit-Methode (RLD-Methode) generiert (Schnell et al. 2013, S. 280).[40]

> *„Von den aus Verzeichnissen gezogenen Nummern wurden die letzten Ziffern gestrichen (üblicherweise die letzten zwei Ziffern) und durch zufällig generierte Nummern ersetzt, was natürlich auch den Nummern eine Auswahlchance gibt, die nicht in Verzeichnisse eingetragen sind." (Meier et al. 2014, S. 124)*

Bei dieser Methode werden allerdings auch Telefonnummern generiert, die überhaupt nicht existieren. Bei CATI mit automatischer Anruffunktion (Autodialing) wählt ein Computer die Telefonnummern an und vermittelt einen erfolgreichen Kontakt, sodass die Interviewer von vergeblichen Kontaktversuchen kaum etwas mitbekommen. Bei manueller Rufnummernwahl ist die Belastung durch vergebliche Kontaktversuche bei nicht geschalteten Telefonnummern deutlich höher.

> *„Darüber hinaus gibt es auch immer wieder Anschlüsse, auf die automatische Ansagen geschaltet sind, bei denen nicht immer zweifelsfrei entschieden werden kann, ob es sich um eine aktive genutzte Rufnummer handelt, oder ob der Anschluss nicht, noch nicht oder nicht mehr genutzt wird." (Meier et al. 2014, S. 120)*

---

39 Nach der Wiedervereinigung waren telefonische Erhebungen in den neuen Bundesländern praktisch nicht realisierbar, da nur etwa 20 Prozent der Haushalte ein Telefon hatten. Erst zur Jahrtausendwende war die Telefondichte ausreichend, um die komplette Bevölkerung abzubilden. Von 1993 bis 2003 stieg der Anteil der Privathaushalte mit einem Festnetzanschluss in den alten Bundesländern von 94 auf 97 Prozent. In den neuen Bundesländern stieg der Anteil der Haushalte mit einem Festnetzanschluss im gleichen Zeitraum von 30 auf 94 Prozent (Meier et al. 2014, S. 123).

40 Schnell (2012, S. 268–269) unterscheidet zwei Lösungen, um das Problem nicht-eingetragener Rufnummern zu lösen: Es werden einerseits eingetragene Telefonnummern verwendet. Die letzten Ziffern werden durch zufällige Ziffern ersetzt. Dieses Verfahren wird als „Randomized Last Digit" (RLD) bezeichnet. Andererseits wird innerhalb der technisch möglichen Telefonnummern eine Teilmenge generiert. Dieses Verfahren wird als „Random Digit Dialing" (RDD) bezeichnet. Die Erzeugung von Telefonnummern ausschließlich auf Basis von Zufallszahlen ist technisch zwar unproblematisch, aber in hohem Maße ineffizient. Schließlich wird nur eine geringe Zahl der Telefonnummern überhaupt vergeben.

Auch Geschäftsanschlüsse müssen teilweise „manuell" aussortiert werden. Schließlich sind auch Anrufbeantworter problematisch, da kein direkter Kontakt hergestellt werden kann. Nach Schnell (2012, S. 275) ist in Deutschland das erneute Anrufen eines Anschlusses mit Anrufbeantworter, ohne eine Nachricht zu hinterlassen, die Standardpraxis.

Gabler-Häder-Design Siegfried Gabler und Sabine Häder (1997) konnten zeigen, dass die klassische RLD-Methode zu verzerrten Stichproben führt. Dieses Problem löst das sogenannte Gabler-Häder-Design. Dieses Verfahren gilt derzeit als die beste Praxis bei der Ziehung von Telefonstichproben (Gabler et al. 1998; Häder und Gabler 2002).

> *„Grundgedanke des Gabler-Häder-Modells ist es, die zu geringe Menge der eingetragenen Rufnummern durch geeignete generierte Rufnummern zu ergänzen, um damit möglichst das Universum aller aktiven Rufnummern verzerrungsfrei abzubilden." (Heckel et al. 2014, S. 141)*

Beim Gabler-Häder-Design werden zunächst alle im Telefonbuch registrierten Rufnummern gesammelt. Diese werden pro Ortsnetzbereich (Vorwahlen) der Größe nach geordnet. Es werden anschließend Blöcke vom Umfang 100 gebildet (z. B. 0 bis 99, 100 bis 199). Können die vorhandenen Rufnummern in einen Block eingeordnet werden (z. B. die Rufnummer 02331/40123 in den Block 40100–40199), dann werden Rufnummern von 02331/40100 bis 02331/40199 generiert. Die Grundannahme ist, dass sich die nicht gelisteten Telefonnummern ebenfalls innerhalb der mit mindestens einer Rufnummer eingetragenen Blöcke befinden. Die Menge aller dieser Ziffernfolgen in allen Ortsnetzbereichen bildet schließlich den Auswahlrahmen für Telefonbefragungen. Blöcke, in denen keine Rufnummer im Verzeichnis gelistet ist, werden aussortiert (Häder und Häder 2014, S. 294). Mittlerweile wurde das Gabler-Häder-Design durch den ADM weiter modifiziert (Heckel et al. 2014).

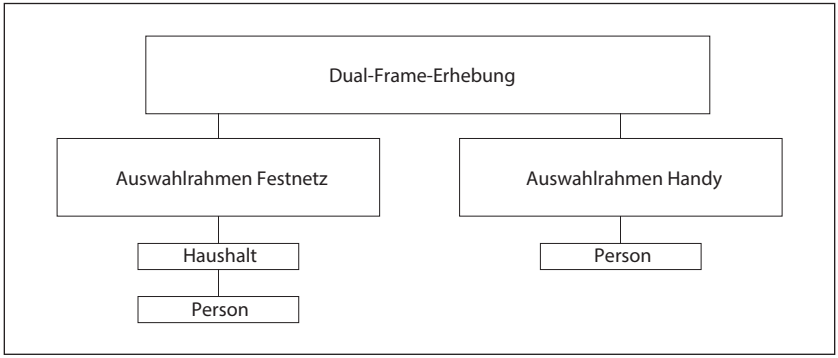
Die Telefonnummer ermöglicht die Auswahl des Haushaltes, nicht aber der Zielperson. Bei telefonischen Erhebungen wird die Zielperson in der Regel mit der Last-Birthday- oder Next-Birthday-Methode ausgewählt. Da es Haushalte mit mehreren Telefonnummern gibt, haben diese eine größere Chance in die Stichprobe zu gelangen. Deshalb werden bei einem telefonischen Interview die Personen gefragt, unter wie vielen Rufnummern sie erreichbar sind. Die Zahl der Rufnummern bzw. Anschlüsse wird bei der statistischen Analyse berücksichtigt (Korrektur der Inklusionswahrscheinlichkeit).

Dual-Frame-Ansatz Nach Zahlen des Statistischen Bundesamts (2009) waren Anfang 2008 rund neun Prozent aller Haushalte ausschließlich über Mobilfunktelefone erreichbar, ohne zusätzlich über einen Festnetzanschluss zu verfügen. Bisher konnten Handys vor allem bei jüngeren Menschen den festen Telefonanschluss verdrängen; der Anteil der ausschließlichen „Handy-Nutzer-Haushalte" nimmt mit steigendem Alter der Haupteinkommensbezieher ab. Die steigende Zahl an Haushalten, die ausschließlich über eine Mobilfunkrufnummer erreichbar sind, hat die Praxis der Stichprobenziehung bei telefonischen Befragungen verändert. Die Lösung bietet ein Dual-Frame-Ansatz.

Abbildung 28 illustriert den Dual-Frame-Ansatz. Bei einer Dual-Frame-Erhebung werden die Auswahlrahmen „Festnetz" und „Handy" kombiniert. Solche aus zwei Auswahlgrundlagen zusammengestellten Stichproben werden als Dual-Frame-Stichproben bezeichnet. Beim Auswahlrahmen „Festnetz" handelt es sich zunächst um eine Haushaltsstichprobe, da jedes Haushaltsmitglied den Hörer abnehmen kann. Innerhalb des Haushalts wird dann über die Last-Birthday- oder Next-Birthday-Methode die Zielperson ausgewählt. Beim Auswahlrahmen „Handy" handelt es sich eher um eine Personenstichprobe, da die meisten Personen ihr Handy überwiegend selbst nutzen (Gabler und Häder 2009, S. 245).

Probleme beim Auswahlrahmen „Handy" Insbesondere die Konstruktion eines geeigneten Rufnummern-Universums für Handynutzer stellt eine Herausforderung dar. Die mögliche Menge an Handy-Rufnummern ist erheblich größer als die tatsächlich vergebenen Rufnummern. Aus Effizienzgründen erfolgt daher ein Abgleich auf Basis von Referenzlisten der Bundesnetzagentur sowie eine technische

Abbildung 28 Dual-Frame-Ansatz



Quelle: Eigene Darstellung in Anlehnung an Gabler und Häder (2009, S. 244)

Vorprüfung (Häder und Häder 2009; Glemser et al. 2014). Forschungsergebnisse weisen darauf hin, dass Unterschiede bei politischen Einstellungen zwischen Personen, die ausschließlich über Mobiltelefone erreichbar sind, und Bürgern, die über einen Festanschluss verfügen, moderat ausfallen (Hunsicker und Schroth 2007, 2014). Allerdings können Mobilfunknummern nicht regional verortet werden, sodass bei lokalen oder regionalen Erhebungen Mobilfunknummern nicht sinnvoll berücksichtigt werden können.

10.4.3 Registerstichproben

Das bekannteste Beispiel für eine Registerstichprobe ist eine Einwohnermeldeamtsstichprobe (EWA-Stichprobe). Dabei handelt es sich um eine Stichprobe auf der Basis von Adressen aus Einwohnermeldeämtern (von der Heyde 2014b, S. 191). Gegenüber dem ADM-Design mit Random Route und Schwedenschlüssel bzw. Geburtstagsmethode hat eine solche Stichprobe zwei Vorteile. Erstens ist der Einfluss der Interviewerin auf die zufällige Auswahl des Zielhaushalts bzw. der Zielperson praktisch ausgeschlossen. Es existiert eine zufällig generierte Liste mit Personen, die befragt werden sollen. Zweitens stehen mit der Liste demographische Angaben über die befragten Personen zur Verfügung (z. B. Geschlecht, Alter), sodass die Arbeit der Interviewer besser kontrolliert werden kann.

> *„EWA-Stichproben verbinden also höchste Regelgebundenheit mit höchster Kontrollsicherheit und Dokumentierbarkeit. Sie sind deshalb die Bevölkerungsstichproben, die derzeit den besten methodischen Ruf haben." (von der Heyde 2014b, S. 192)*

Die beiden wichtigsten deutschen sozialwissenschaftlichen Datenerhebungen – die Allgemeine Bevölkerungsumfrage für Sozialwissenschaften (ALLBUS) und der European Social Survey (deutsche Teilstudie) – nutzen unter anderem aus diesen Gründen eine Registerstichprobe, um eine „repräsentative" Stichprobe zu realisieren.

Nachteile von EWA-Stichproben Allerdings sind mit einer EWA-Stichprobe auch vier Nachteile verbunden. Erstens sind solche Register nicht völlig aktuell. Umzüge oder Sterbefälle werden erst nach einer gewissen Zeit im Melderegister realisiert. Zweitens sind solche Register auch nicht völlig fehlerfrei. Nicht alle Einwohner einer Stadt haben sich ordnungsgemäß angemeldet oder wieder abgemeldet. Drittens ist der Aufwand für eine Einwohnermeldeamtsstichprobe deutlich höher als Random Route und Schwedenschlüssel. Für die Adressenlieferung ist das Institut auf die Kooperation mit der Gemeinde angewiesen. Dabei ist zum einen

nicht jede Gemeinde bereit, solche Adressen zur Verfügung zu stellen, zum anderen wird die Adresslieferung mit teilweise sehr hohen Gebührenforderungen verknüpft. Manchmal werden die Adressen auch in kaum nutzbaren Formaten geliefert, sodass erheblicher Aufbereitungsaufwand anfällt. Schließlich können zwischen der Anforderung der Stichprobe und der Lieferung der Daten durchaus mehrere Monate vergehen. Dies wirkt sich natürlich ungünstig auf die Datenqualität aus (Albers 1997). Viertens ist eine Registerstichprobe nur möglich, „wenn das Projekt, für das sie benutzt werden soll ‚im öffentlichen Interesse' liegt" (von der Heyde 2014b, S. 191). In der Regel wird Aufträgen privater Institutionen das öffentliche Interesse abgesprochen. Aber auch bei wissenschaftlichen Forschungsprojekten muss das öffentliche Interesse deutlich dargelegt werden. Die Entscheidung, ob dieses „öffentliche Interesse" vorliegt, treffen in vielen Fällen die Gemeinden bzw. das Stadtoberhaupt (von der Heyde 2014b, S. 191).

Datenerhebung

11

Nach der Festlegung des Forschungsdesigns und der Entscheidung für ein Auswahlverfahren muss die Datenerhebung konkretisiert werden. Traditionell werden drei Varianten der Datenerhebung unterschieden: Befragung, Beobachtung und Inhaltsanalyse (z. B. Atteslander 2010; Diekmann 2011, S. 195; Schnell et al. 2013, S. 311). Diese drei Formen der Datenerhebung werden in diesem Kapitel vorgestellt. Im Mittelpunkt steht dabei die standardisierte Befragung, die als „klassisches Instrument der Datenerhebung" (Reinecke 2014, S. 601) gilt und gerne auch als „Königsweg der praktischen Sozialforschung" (König 1974, S. 27) bezeichnet wird. Standardisierte Befragungen liefern heute die Datengrundlage für die Mehrheit der empirischen Studien (Diekmann 2011, S. 434; Schnell 2012, S. 28).

11.1 Befragung

Befragung ist zunächst ein loser Oberbegriff für eine Vielzahl an Befragungstechniken. Während qualitative Sozialforscher häufig Befragungsarten mit geringer Standardisierung nutzen (z. B. Leitfadeninterviews), wird innerhalb der quantitativen Sozialforschung eine möglichst hohe Standardisierung der Befragung angestrebt. Die Standardisierung bezieht sich auf die einheitliche Formulierung der Fragen, die vorgegebenen Antwortoptionen und die Reihenfolge der Fragen. Die Regeln standardisierter Befragungen verlangen, dass jeder Befragte die gleichen Fragen in gleicher Formulierung und Reihenfolge erhält.

- **Definition: Standardisierte Befragung**
  Bei standardisierten Befragungen werden alle Personen mit den gleichen Fragen und Antwortmöglichkeiten konfrontiert. Auch die Rahmenbedingungen einer Befragung (z. B. Interviewsituation) und die Durchführung der Befragung

© Springer Fachmedien Wiesbaden GmbH, ein Teil von Springer Nature 2018

M. Tausendpfund, *Quantitative Methoden in der Politikwissenschaft*,
Grundwissen Politik, https://doi.org/10.1007/978-3-658-20698-7_11

(z. B. Reihenfolge der Fragen, Hinweise zu einzelnen Fragen und Überleitungen) sind für alle Personen identisch. Die Standardisierung soll gewährleisten, dass unterschiedliche Antworten auf eine Frage auch tatsächlich auf unterschiedliche Merkmale, Einstellungen und Verhaltensweisen der Personen zurückzuführen sind und nicht auf variierende Bedingungen während der Befragungssituation (z. B. Schnell 2012, S. 26–28; Reinecke 2014, S. 602; Weischer 2015a).

Den Stellenwert der Befragung als Instrument der Datenerhebung verdeutlichen die Zahlen des Jahresberichts des Arbeitskreises Deutscher Markt- und Sozialforschungsinstitute (ADM 2013, S. 21–22). Im Jahr 2013 führten die Mitgliedsinstitute des ADM über 21 Millionen standardisierte Interviews durch. Schnell (2012, S. 30) dokumentiert einen Anstieg der wissenschaftlichen Veröffentlichungen auf Basis von Befragungsdaten. Nach Saris und Gallhofer (2014, S. 2) ist der Anteil an Publikationen mit Befragungsdaten in den vergangenen Jahrzehnten in der Soziologie, Politikwissenschaft, Sozialpsychologie sowie in der Ökonomie deutlich angewachsen. Für viele Fragestellungen (z. B. Einstellungen, Sozialstruktur) sind standardisierte Befragungen unverzichtbar (Diekmann 2011, S. 434–436).

11.1.1 Befragungsarten

Vier Formen standardisierter Befragungen werden unterschieden: Persönliche, telefonische, schriftliche und internetgestützte Interviews (z. B. Schnell 2012; Schnell et al. 2013, S. 314).[41] Diese unterschiedlichen Befragungsarten werden auch als Befragungsmodi bezeichnet.

Persönliche Befragung Die persönliche Befragung, häufig als Face-to-Face-Interview bezeichnet, setzt die Anwesenheit der Gesprächspartner (Interviewer und Befragter) voraus. Der Interviewer sucht die Zielperson auf und führt die Befragung in der Regel in dessen Privatwohnung durch. Die Befragung wird heute computergestützt realisiert. Der Interviewer liest die Fragen direkt vom Bildschirm des Laptops ab und die Antworten werden elektronisch erfasst. Diese Form des Interviews wird als CAPI bezeichnet (Computer Assisted Personal Interviewing). Die Computerunterstützung ermöglicht zudem eine komplexe Filterführung (siehe z. B. Scholl 2003, S. 52–57 für eine Darstellung der Vor- und Nachteile der computergestützten Befragung). Die klassische Vorgehensweise bei einer

41 Scholl (2003, S. 31) unterscheidet nur drei Gruppen der Befragung: Mündliche, telefonische und schriftliche Befragung. Die Online-Befragung stellt in seiner Darstellung nur eine Variante der schriftlichen Befragung dar.

persönlichen Befragung wird PAPI genannt (Paper-and-Pencil-Interviewing). Dabei liest der Interviewer die Fragen von einem Fragebogen vor und trägt die Antwort mit einem Stift ein.

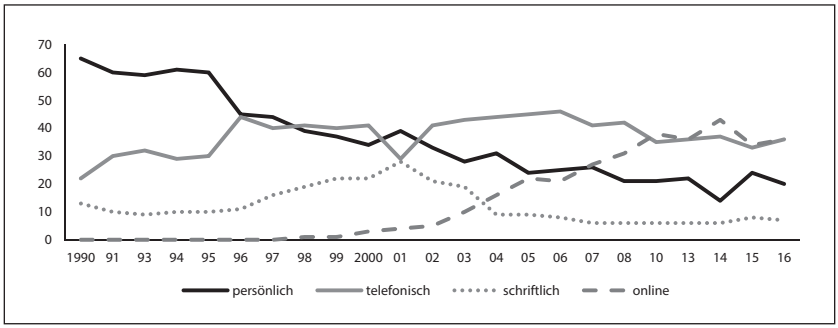
Telefonische Befragung Die telefonische Befragung setzt voraus, dass die Zielperson über einen Telefonanschluss verfügt bzw. telefonisch erreichbar ist. Telefonische Befragungen werden praktisch immer von zentralen Telefonstudios als computergestützte Telefoninterviews durchgeführt. Diese Variante der Befragung wird CATI bezeichnet (Computer Assisted Telephone Interview). Die Wahl der Telefonnummer, die Interviewfragen und Antwortkategorien, die Filterführung sowie die Speicherung der Daten werden durch einen Computer realisiert. Die Interviewerin sitzt mit einem Headset vor dem Monitor und wird bei der praktischen Durchführung durch den Computer bzw. die entsprechende Software unterstützt.

Schriftliche Befragung Bei der schriftlichen Befragung ist kein Interviewer anwesend und die Zielperson füllt den Fragebogen selbst aus. Der Fragebogen wird als Brief verschickt. Dieser Brief enthält neben dem Fragebogen in der Regel ein Anschreiben sowie einen adressierten und frankierten Rückumschlag. Bei der schriftlichen Befragung kann die Zielperson selbst bestimmen, wann und wo sie den Fragebogen ausfüllt. Dadurch sind allerdings auch die situativen Begleitumstände für den Forscher unklar. Auch kann nicht nachvollzogen werden, wer letztlich den Fragebogen ausgefüllt hat.

Online-Befragung Die Online-Befragung ist mit der schriftlichen Befragung vergleichbar, allerdings ersetzt das Internet die postalische Zustellung des Fragebogens. Die Einladung zu einer Online-Befragung kann persönlich, telefonisch, schriftlich oder auch online (via E-Mail) erfolgen. Die Bearbeitung des Fragebogens erfolgt dann online, sodass dieselben technischen Möglichkeiten wie bei computergestützten persönlichen und telefonischen Befragungen zur Verfügung stehen. Dabei kann der Befragte direkt auf Eingabefehler hingewiesen werden. Durch die elektronische Erfassung der Daten können Zwischenberichte schnell erstellt werden. Diese Variante wird als CAWI bezeichnet (Computer Assisted Web Interviewing).

Pluralisierung der Befragungsformen Abbildung 29 dokumentiert die Pluralisierung der Befragungsformen in Deutschland. Durch die zunehmende Telefondichte in der Bundesrepublik Deutschland lösten Telefoninterviews in den 1990er Jahren die persönliche Befragung als wichtigste Befragungsart ab. Während der Anteil persönlicher Befragungen von 1992 bis 1999 um 21 Prozentpunkte abnahm, stieg der Anteil telefonischer Erhebungen im gleichen Zeitraum von 32 auf 40 Pro-

Abbildung 29 Interviewanteile nach Befragungsart in Prozent



Quelle: Eigene Darstellung auf Basis der Jahresberichte des ADM

zent. Der Anteil schriftlicher Befragungen lag 1999 bei 22 Prozent. Mit einem Anteil von weniger als einem Prozent spielten Online-Befragungen vor der Jahrtausendwende keine bedeutsame Rolle.

Mit der Verbreitung des Internets ist der Anteil der Online-Befragungen von drei Prozent im Jahr 2000 auf 38 Prozent im Jahr 2010 gestiegen. Im gleichen Zeitraum haben sich die Anteile der persönlichen Befragungen um 13 Prozentpunkte und der telefonischen Befragungen um sechs Prozentpunkte verringert. Der Anteil schriftlicher Befragungen ist von 22 auf sechs Prozent gesunken. In den vergangenen Jahren haben sich die jeweiligen Anteile der Befragungsarten stabilisiert. 2016 betrug der Anteil persönlicher Interviews 20 Prozent; 36 Prozent waren telefonische Befragungen (ADM 2016, S. 15). Der Rückgang des Anteils schriftlicher Interviews seit der Jahrtausendwende auf acht Prozent im Jahr 2013 ist auf eine Substitution durch Online-Interviews zurückzuführen. Der Anteil der Online-Befragungen liegt bei den ADM-Mitgliedsinstituten im Jahr 2013 bei 34 Prozent (ADM 2015, S. 14). Da sich jede Befragungsart durch spezifische Vor- und Nachteile auszeichnet, ist mit einer vollständigen Ersetzung der „klassischen" Befragungsformen durch Online-Interviews nicht zu rechnen. Diese spezifischen Vor- und Nachteile werden im Folgenden genauer betrachtet.

Vor- und Nachteile In Tabelle 15 sind Vor- und Nachteile der Befragungsarten für sechs Vergleichsdimensionen dargestellt: Abbildung der Grundgesamtheit, maximale Dauer eines Interviews, visuelle Hilfsmittel, soziale Erwünschtheit, Dauer der Feldphase und Kosten je Interview (siehe auch Schnell 2012, S. 309; Häder 2015, S. 195). Auf Basis dieser Vergleichsdimensionen werden die vier Befragungsformen charakterisiert.

Tabelle 15 Vergleich der Vor- und Nachteile von Befragungsarten

	Persönliche Befragung	**Telefonische Befragung**	**Schriftliche Befragung**	**Online-Befragung**
Abbildung der Grundgesamtheit	hoch	mittel	mittel	gering
Maximale Dauer eines Interviews	lang	mittel	mittel	kurz
Visuelle Hilfsmittel	ja	nein	begrenzt	ja
Soziale Erwünschtheit	hoch	mittel	gering	gering
Dauer der Feldphase	lang	kurz	lang	mittel
Kosten je Interview	hoch	mittel	mittel	gering
Weiterführende Literatur	Stocké (2014)	Hüfken (2014)	Reuband (2014)	Wagner und Hering (2014)

Abbildung der Grundgesamtheit Aus zeitlichen, finanziellen und forschungspraktischen Gründen dominieren in den Sozialwissenschaften Teilerhebungen. Bei vielen Fragestellungen soll eine Zufallsstichprobe einem verkleinerten Abbild der Grundgesamtheit entsprechen (siehe Kapitel 10). Bei der Entscheidung für eine Befragungsform sind deshalb die Konsequenzen für die Abbildung der Grundgesamtheit zu berücksichtigen. Diese unterscheiden sich in Abhängigkeit von der festgelegten Grundgesamtheit und werden an einem Beispiel dargestellt. Für unser Beispiel wird die wahlberechtigte Bevölkerung in der Bundesrepublik Deutschland als Grundgesamtheit festgelegt.

Beim ADM-Design für persönliche Befragungen werden bestimmte Bevölkerungsgruppen praktisch aus der Befragung ausgeschlossen, zum Beispiel Bewohner von Alten- und Pflegeheimen, Patienten in Krankenhäusern oder auch Wohnungslose. Bei der Entscheidung für eine persönliche Befragung haben diese Personen keine Chance in die Stichprobe zu gelangen. Insgesamt ist der Ausschluss von bestimmten Bevölkerungsgruppen aber eher gering.

Auch bei einer telefonischen Befragung werden mehr oder weniger explizit bestimmte Bevölkerungsgruppen ausgeschossen. Wer keinen Telefonanschluss hat, kann nicht berücksichtigt werden. Hinzu kommen noch Personen, die nur über ein Mobiltelefon verfügen. Für deutschlandweite Erhebungen bietet der Dual-Frame-Ansatz eine Lösung, bei lokalen und regionalen Studien können Mobilfunknummern allerdings nicht sinnvoll berücksichtigt werden (siehe Abschnitt 10.4.2).

! Bei einer schriftlichen Befragung müssen die Postadressen bekannt sein, um den Personen den Fragebogen (mit Rücksendeumschlag) überhaupt zusenden zu können. Eine deutschlandweite Datenbank mit den Postadressen der Bürger existiert allerdings nicht. Eine schriftliche Befragung setzt daher mindestens zwei Auswahlschritte voraus: Erstens müssen die Gemeinden ausgewählt werden. Zweitens müssen innerhalb der Gemeinden die Postadressen über die Einwohnermeldeämter bezogen werden. Dies ist allerdings mit mehreren Herausforderungen und Problemen verbunden (siehe Kapitel 10). Zudem gelten nach Schnell (2012, S. 247) etwa 7,5 Millionen Personen im Alter von 18 bis 64 Jahre als funktionale Analphabeten. Diese werden bei einer postalischen Befragung praktisch aus der Stichprobe ausgeschlossen.

Eine allgemeine Bevölkerungsumfrage („wahlberechtigte Bevölkerung") auf Basis einer Online-Befragung lässt sich nur unter Einschränkungen realisieren. Nach Daten der ARD-ZDF-Onlinestudie 2017 liegt der Anteil der Onlinenutzer der deutschsprachigen Bevölkerung ab 14 Jahren bei 90 Prozent. Bei einer Online-Befragung werden somit rund zehn Prozent der Bevölkerung ausgeschlossen, da diese das Internet nicht nutzen bzw. keinen Zugang zum Internet haben. Hinzu kommen noch erhebliche Unterschiede nach Alter. In der Altersgruppe 14 bis 19 Jahre nutzen praktisch alle Personen das Internet, in der Altersgruppe ab 60 Jahren sind es rund 74 Prozent (Koch und Frees 2017). Da zudem keine Datenbank der E-Mail-Adressen der Internetnutzer in Deutschland existiert, stellt auch der Erstkontakt der Befragten eine Herausforderung dar. Eine Lösung bieten „offline" rekrutierte Online-Erhebungen, zum Beispiel das German Internet Panel (Blom et al. 2015) und das GESIS Panel (Bosnjak et al. 2017).

Der Vergleich zeigt: Für eine bundesweite – sogenannte repräsentative – Befragung bieten sich aktuell nur persönliche und telefonische Interviews an. Die Entscheidung für die konkrete Befragungsform muss allerdings immer die definierte Grundgesamtheit berücksichtigen. Eine lokale Bürgerbefragung lässt sich beispielsweise mit einer schriftlichen Befragung realisieren, da die Gemeinde über die notwendigen Adressinformationen verfügt. Eine Befragung der Studierenden der FernUniversität in Hagen als Online-Befragung ist durchaus möglich, da erstens alle Studierenden über einen Internetzugang verfügen und zweitens eine „Liste" der E-Mail-Adressen existiert. Allerdings sind Bestimmungen des Datenschutzes zu berücksichtigen.

Maximale Dauer eines Interviews Die Teilnahme an den meisten Befragungen ist freiwillig.[42] Die Forscherinnen bzw. stellvertretend die Interviewerinnen, die die Befragung durchführen, sind deshalb auf die Kooperation der Befragten angewiesen.

Bei den meisten Befragten treten nach 30 Minuten Ermüdungserscheinungen auf, die zum Abbruch des Interviews führen können (Brace 2013, S. 13). Deshalb sind längere Befragungszeiten nur bei Themen möglich, für die sich der Befragte persönlich interessiert. Dies dürfte bei den meisten sozialwissenschaftlichen Themen nicht der Fall sein. Deshalb betont Schnell (2012, S. 117–118) sehr deutlich:

> *„Die Einzigen, die über eine maximale Befragungsdauer entscheiden können, sind die Befragten. Und die Mehrheit der Befragten toleriert bei Themen, die sie nicht interessieren, keine Befragungen, die länger als 30 Minuten sind." (Schnell 2012, S. 117–118)*

Eine Befragungsdauer von 30 Minuten gilt damit als Obergrenze. Die maximale Befragungsdauer variiert allerdings auch in Abhängigkeit von der Befragungsform. Bei einer persönlichen Befragung sind Erhebungszeiten von bis zu 60 Minuten möglich (z. B. ALLBUS oder ESS). Einerseits verringert die Anwesenheit eines Interviewers die Abbruchwahrscheinlichkeit, andererseits erhalten Personen bei längeren Befragungen auch kleinere Geschenke (sogenannte Incentives). Häufig werden persönliche Befragungen auch schriftlich angekündigt.

Bei einer telefonischen Befragung sollte die Obergrenze von maximal 30 Minuten ernst genommen werden, da viele Personen längere Befragungen nicht akzeptieren. In der Regel wird die Befragungszeit zu Beginn des Interviews angekündigt. Starke Abweichungen von der realistischen Befragungszeit sind zu vermeiden, da Täuschungen und Irreführungen vom Gesprächspartner nicht toleriert werden.

Die Bereitschaft zur Teilnahme an einer schriftlichen Befragung ist unter anderem vom Seitenumfang abhängig. Bei einer Befragung mit einem kurzen Fragebogen (vier bis sechs Seiten) werden sich mehr Personen beteiligen als an einer Befragung mit einem 30-seitigen Fragebogen. Bei Porst (2001, S. 4) gelten 16 DIN-A4-Seiten als Obergrenze, und Reuband (2014, S. 12) nennt zwölf Seiten als Maximum. Bei einer Online-Befragung nennt Schnell (2012, S. 309) 15 Minuten als maximale Befragungsdauer.

---

42 Eine Ausnahme ist der Mikrozensus. Dabei handelt es sich um eine Haushaltsbefragung der amtlichen Statistik in Deutschland. Rund 830 000 Personen in etwa 370 000 privaten Haushalten und Gemeinschaftsunterkünften werden stellvertretend für die gesamte Bevölkerung zu ihren Lebensbedingungen befragt. Dies sind etwa ein Prozent der Bevölkerung, die zufällig ausgewählt werden. Beim Mikrozensus besteht Auskunftspflicht (siehe § 7 Mikrozensusgesetz).

Warum beteiligen sich Personen freiwillig an einer Befragung?

Schnell (1997, S. 157–216; siehe auch Esser 1986) erklärt die Teilnahme an einer Befragung auf Basis der Theorie des rationalen Handelns. Danach ist eine Teilnahme wahrscheinlich, wenn der erwartete Nutzen der Teilnahme die erwarteten Kosten überschreitet. Die Teilnahme ist wahrscheinlicher, wenn der Befragte durch die Teilnahme einen Nutzen hat. Dies kann beispielsweise das Interesse an der Befragung, die Unterstützung der Wissenschaft oder auch einfach Abwechslung sein. Andererseits sprechen Zeitmangel, Desinteresse oder grundsätzliche Abneigung gegenüber Umfragen gegen eine Teilnahme. Die Entscheidung über die Teilnahme an einer Befragung lässt sich dabei als eine sogenannte Low-Cost-Situation konzeptualisieren (allgemein zur Low-Cost-Hypothese siehe z. B. Best und Kroneberg 2012), der erwartete Nutzen und die Kosten sind relativ gering. Deshalb kann die Teilnahmewahrscheinlichkeit häufig durch geringe Veränderungen der Erhebungssituation gefördert werden (z. B. kleine Geschenke, das Geschlecht oder die Sprache des Interviewers). Eine theoretische Weiterentwicklung bietet die Leverage-Salience-Theorie (Groves et al. 2000).

Visuelle Hilfsmittel Bei sozialwissenschaftlichen Erhebungen werden teilweise sehr komplexe Inhalte abgefragt, die die Befragten innerhalb weniger Sekunden beantworten sollen. Einige Beispiele aus dem European Social Survey: Wie sicher fühlen Sie sich – oder würden Sie sich fühlen – wenn Sie nach Einbruch der Dunkelheit alleine zu Fuß in Ihrer Wohngegend unterwegs sind oder wären? Gibt es eine politische Partei, der Sie näher stehen als allen anderen Parteien? Was würden Sie sagen, ist es im Allgemeinen gut oder schlecht für die deutsche Wirtschaft, dass Zuwanderer hierher kommen? Durch visuelle Hilfsmittel können die Befragten bei der Beantwortung einer Frage unterstützt werden. Damit ist die visuelle Präsentation von Antwortvorgaben oder Antwortkategorien gemeint.

Je nach Befragungsform lassen sich visuelle Hilfsmittel einfacher oder schwieriger einsetzen. Bei einer persönlichen Befragung kann der Interviewer dem Befragten entsprechende Listen vorlegen (Stocké 2014, S. 622). Bei einer Online-Befragung können Multimediainhalte wie Bilder, Töne und Videos relativ einfach integriert werden (Wagner und Hering 2014, S. 662). Schwieriger können visuelle Hilfsmittel bei schriftlichen Befragungen eingesetzt werden. Bei telefonischen Befragungen sind visuelle Hilfsmittel praktisch nicht realisierbar.

Soziale Erwünschtheit Mit sozialer Erwünschtheit wird die Tendenz einer Person beschrieben, auf eine Frage nicht wahrheitsgemäß, sondern entsprechend der wahrgenommenen gesellschaftlichen Erwartung zu antworten. Befragte behaupten dann im Interview, dass sie sich bei der nächsten Bundestagswahl beteiligen, obwohl dies nicht zutrifft. Oder umgekehrt: Personen geben „sozial unerwünschtes" Verhalten (z. B. Drogenkonsum, Diebstahl) seltener an, da dies nicht der gängigen gesellschaftlichen Norm entspricht (siehe ausführlich Diekmann 2011, S. 447–455).

Die Wirksamkeit der sozialen Erwünschtheit ist von der Befragungsform abhängig. Durch die Anwesenheit einer Interviewerin sind bei einer persönlichen Befragung stärkere Effekte durch soziale Erwünschtheit zu erwarten als bei einer postalischen Erhebung oder einer Online-Befragung. Bei einer telefonischen Befragung sollten mögliche Effekte durch soziale Erwünschtheit geringer sein als bei einer persönlichen Befragung, da der Gesprächspartner nicht unmittelbar sichtbar ist.

Dauer der Feldphase Die Datenerhebung des European Social Survey (ESS) erfolgt mittels persönlicher Befragung. Bei der sechsten Erhebung wurden in Deutschland 2958 Personen befragt. Die Feldphase, der Zeitraum der eigentlichen Datenerhebung, begann am 6. September 2012 und endete am 22. Januar 2013. Für die Durchführung der knapp 3000 Interviews waren damit etwa fünf Monate erforderlich. Beim Politbarometer, einer telefonischen Befragung der Forschungsgruppe Wahlen im Auftrag des ZDF, werden innerhalb von drei Werktagen etwa 1200 zufällig ausgewählte Wahlberechtigte befragt.

Der Vergleich zwischen ESS und Politbarometer verdeutlicht: Die Dauer der Feldphase variiert in Abhängigkeit von der Befragungsform. Bei einer persönlichen Befragung muss der Interviewer den Haushalt bzw. die Zielperson aufsuchen, während eine telefonische Befragung zentral in einem Telefonlabor durchgeführt wird.

Auch Online-Befragungen können zeitlich und räumlich unabhängig realisiert werden. Die Dauer der Feldphase ist sicherlich etwas länger als bei einer telefonischen Erhebung, aber deutlich kürzer als bei einer persönlichen Befragung. Für die Durchführung einer postalischen Befragung werden mehrere Schritte empfohlen (ausführlich siehe Dillman et al. 2014, S. 351–397; siehe auch Schnell 2012, S. 249–251): Ankündigungsschreiben, Versand des Fragebogens und mindestens zwei Erinnerungsschreiben. Für die praktische Durchführung sind daher mehrere Wochen zu veranschlagen.

Kosten je Interview Schließlich variieren auch die Kosten je Interview in Abhängigkeit von der Befragungsart. Bei Koch (2002, S. 29–31) findet sich eine Über-

sicht der Entwicklung der Kosten der ALLBUS (persönliche Befragung). Danach sind die Kosten von 1980 bis 2000 von 40 auf 130 Euro je Interview gestiegen. Die Kosten beinhalten unter anderem das Honorar und die Spesen der Interviewer, Stichprobenziehung, Interviewerschulungen, Projektleitung und die Bereitstellung des Datensatzes. Selbstverständlich variieren die Kosten einer persönlichen Befragung auch in Abhängigkeit vom Design der Erhebung. Eine Registerstichprobe vom Einwohnermeldeamt ist teurer als Random Route. Ein Random Route mit Adressvorlauf ist teurer als ein Random Route ohne Adressvorlauf. Die Kosten für eine einstündige persönliche Befragung auf Basis einer Einwohnermeldeamtsstichprobe liegen heute zwischen 200 und 250 Euro – je Interview. Mit anderen Worten: Eine Befragung von etwa 3000 Personen kostet rund 675 000 Euro (ohne Mehrwertsteuer). Allerdings liegen nur wenige veröffentlichte Informationen zu den Kosten einer bundesweiten Bevölkerungsumfrage vor.

Im Vergleich zu persönlichen Befragungen sind die Kosten für eine telefonische Erhebung deutlich geringer (Hüfken 2014, S. 46). Durch die zentrale Durchführung im Telefonlabor entstehen beispielsweise keine Reisekosten. Auch die Stichprobenziehung ist günstiger, da im Vergleich zur Einwohnermeldeamtsstichprobe keine Gebühren anfallen. Die Kosten für eine geschlossene Frage mit mehreren Antwortkategorien beginnen beim PolitikBus von Infratest dimap bei 990 Euro.[43] Dafür werden dann 1000 Personen telefonisch befragt. Bei 30 Fragen werden etwa 30 000 Euro fällig (ohne Mehrwertsteuer). Die meisten sozialwissenschaftlichen Befragungen umfassen allerdings deutlich mehr als 30 Fragen.

Die Kosten einer schriftlichen Befragung sind im Vergleich zu einer persönlichen oder telefonischen Befragung geringer. Allerdings sind bei der Kalkulation neben Druck- und Portokosten für Fragebogen und Anschreiben auch Kosten für die Stichprobenziehung, Adress- und Rücklaufverwaltung, Dateneingabe und Datenbereinigung sowie die Entsorgung der Dokumente zu berücksichtigen. Schnell (2012, S. 244) kalkuliert die Kosten einer sorgfältig durchgeführten postalischen Befragung von 1000 Personen mit rund 30 000 Euro und betont:

> *„Diese Kalkulation stellt die Untergrenze der Kosten dar." (Schnell 2012, S. 244)*

Die Online-Befragung ist die Befragungsart mit den geringsten Kosten (Wagner und Hering 2014, S. 48). Je nach Gestaltung der Teilnehmerverwaltung (telefonisch, postalisch oder E-Mail) fallen nur geringe bis keine Kosten für Einladun-

43 Beim PolitikBus von Infratest dimap handelt es sich um eine klassische Mehrthemenbefragung (auch Omnibusbefragung genannt). Dabei werden die Fragen mehrerer Auftraggeber gebündelt und unterschiedliche Themen in einer Erhebung abgefragt. Eine Darstellung des PolitikBus findet sich unter www.infratest-dimap.de/leistungen/politikbus/.

gen, Fragebogenversand und Erinnerungen an. Auch entstehen keine Kosten für Interviewer oder Dateneingabe. Allerdings sind Aufwendungen für Hard- und Software sowie die technische Kompetenz zur Erstellung eines Online-Fragegogens erforderlich.

Fazit Je nach Forschungsfrage und Festlegung der Grundgesamtheit kann eine persönlich-mündliche Befragung obligatorisch oder eine Online-Befragung ausreichend sein. Die Entscheidung für eine konkrete Befragungsart ist damit in erster Linie von der Zielsetzung der Studie abhängig.

Mixed-Mode-Surveys Eine Kombination mehrerer Befragungsformen wird als Mixed-Mode-Survey bezeichnet. Dabei können die Zielpersonen den Fragebogen beispielsweise klassisch als Papierversion ausfüllen oder als Online-Variante. Durch die Kombination der Befragungsformen soll die Teilnahmebereitschaft der Personen gefördert werden. Zudem können mögliche Nachteile einzelner Befragungsformen (z. B. erforderliche Computerkompetenz beim Ausfüllen eines Online-Fragebogens) reduziert werden (Schnell 2012, S. 312–316; Hox et al. 2015; Dillman und Edwards 2016; zu methodischen Problemen siehe Eifler und Faulbaum 2017).

Tailored Design Method

Die Tailored Design Method (TDM) ist ein von Don Dillman et al. (2014) entwickelter Ansatz zur Optimierung von standardisierten Befragungen. Die Methode umfasst alle Schritte einer Erhebung – von der Konzeption einer Erhebung über die Formulierung von Fragen und Antworten bis hin zur praktischen Durchführung (z. B. die Papierfarbe). Die Tailored Design Method ist eine Weiterentwicklung der Total Design Methode und gibt zahlreiche Hinweise zur Durchführung von Befragungen.

11.1.2 Fragearten

Bei einer Befragung stellt die Interviewerin dem Gesprächspartner Fragen. Diese Fragen können nach ihrem Inhalt oder ihrer Form unterschieden werden (Porst 2014a, S. 53). Die Differenzierung der Fragen nach Inhalt ist beliebig – es gibt unter anderem Fragen nach Einstellungen, Überzeugungen, Verhalten und sozialstatis-

tischen Merkmalen (z. B. Diekmann 2011, S. 471). Diese Liste lässt sich sicherlich noch ergänzen (z. B. Fragen nach Wertorientierungen, Wissen). Deshalb spielt die Differenzierung von Fragen nach ihrem Inhalt in der Literatur keine besonders große Rolle, wichtiger ist hingegen die Unterscheidung nach der Frageform. Porst (2014a, S. 53) unterscheidet traditionell drei Fragearten: geschlossene, halboffene und offene Fragen.

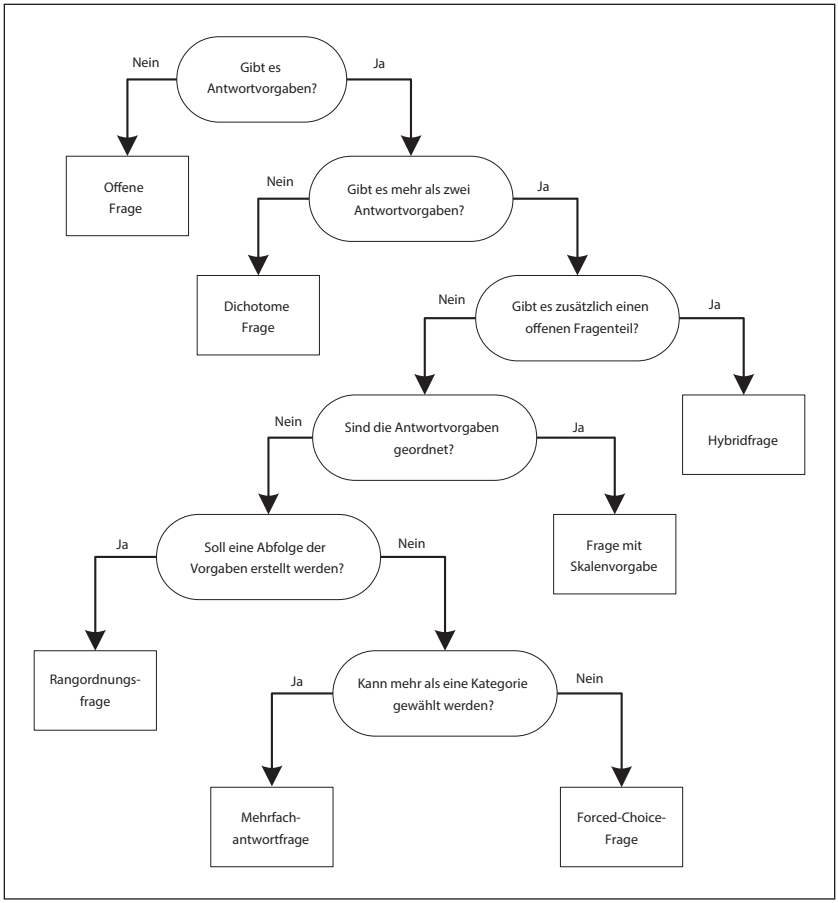
Drei klassische Fragearten Bei geschlossenen Fragen gibt es eine begrenzte und definierte Anzahl möglicher Antwortoptionen, aus der der Befragte „seine" Antwort auswählen muss. Geschlossene Fragen können danach unterschieden werden, ob es nur eine zulässige Antwort oder mehrere zulässige Antwortmöglichkeiten gibt. Bei halboffenen Fragen (Hybridfragen) wird – wie bei einer geschlossenen Frage – eine Anzahl möglicher Antwortoptionen vorgegeben, aus der der Befragte seine Antwort auswählen kann. Darüber hinaus enthält eine Hybridfrage eine „Restkategorie", die meist mit „Sonstige, und zwar" eingeleitet wird. Bei einer offenen Frage wird der Fragetext vorgelesen und der Befragte antwortet in eigenen Worten.

Klassifikation von Fragearten Eine detailliertere Klassifikation von Fragearten hat Schnell (2012, S. 82) vorgelegt. Auf der Basis von Antwortvorgaben präsentiert er einen Klassifikationsbaum (siehe Abbildung 30). Auf diese Weise unterscheidet Schnell (2012, S. 82) sieben unterschiedliche Fragearten: Offene Frage, dichotome Frage, Hybridfrage, Frage mit Skalenvorgaben, Rangordnungsfrage, Mehrfachantwortfrage und Forced-Choice-Frage. Diese Fragearten werden an Beispielen dargestellt.

Offene Frage Als offene Fragen werden alle Fragearten bezeichnet, bei denen es keine Antwortvorgaben gibt. Der Befragte kann in eigenen Worten antworten; er übernimmt selbst die Formulierung einer Antwort (Schnell 2012, S. 81–86). Der Vorteil offener Fragen besteht darin, dass der Befragte innerhalb seines eigenen Referenzsystems antworten kann. Er wird nicht durch vorgegebene Antwortoptionen in seiner Antwort beeinflusst. Dadurch unterstützen offene Fragen Antworten, die auch tatsächlich im Wissensbestand des Befragten verankert sind (Schnell et al. 2013, S. 324; siehe auch Züll und Menold 2014; Porst 2014a, S. 66–69).

Ein Beispiel: Aus früheren Studien ist zwar bekannt, was Befragte in den 1980er- und 1990er-Jahren mit den Begriffen „links" und „rechts" verbanden, aber möglicherweise hat sich die Bedeutung der Begriffe verändert. Um diese mögliche Veränderung zu erfassen, bietet sich eine offene Frage an (Züll und Menold 2014). Deshalb wurde in der Allgemeinen Bevölkerungsfrage der Sozialwissenschaften

Abbildung 30 Klassifikation von Fragearten



Quelle: Schnell (2012, S. 82).

(ALLBUS) 2008 den Befragten im Anschluss an die Links-Rechts-Selbsteinstufung unter anderem folgende offene Frage gestellt:

Würden Sie mir bitte sagen, was Sie mit dem Begriff „links" verbinden?

Der Interviewer hat die Antwort des Befragten notiert. Die Angaben der Befragten wurden mittels einer computerunterstützten Inhaltsanalyse sehr aufwändig kodiert (Züll et al. 2010). Auf diese Weise kann einerseits ein möglicher Bedeutungswandel der Begriffe identifiziert und andererseits das ideologische Verständnis der Bürger mit dem der politischen Eliten verglichen werden.

Den Vorteilen offener Fragen stehen allerdings auch mehrere Nachteile gegenüber. Erstens unterscheiden sich die Befragten in ihrer Artikulationsfähigkeit. Antwortunterschiede sind dann nicht etwa auf Einstellungsunterschiede zurückzuführen, sondern basieren auf der Fähigkeit des Befragten, seine Haltung in Worte zu fassen. Zweitens sind Interviewereffekte bei offenen Antworten wahrscheinlicher, da der Interviewer die Antworten genau notieren muss. Dabei können einzelne Antwortteile verändert werden. Drittens verlängern offene Fragen die Interviewdauer. Zeit stellt bei einem Interview allerdings eine wertvolle Ressource dar. Viertens erfordern offene Fragen einen erheblichen zusätzlichen Auswertungsaufwand. Es muss ein Kategorienschema zur Codierung der Angaben entwickelt und jede Antwort muss einzeln kodiert werden (Schnell et al. 2013,

Links-Rechts-Schema

Als ideologische Orientierungsdimension ist das Links-Rechts-Schema in den Sozialwissenschaften von besonderer Bedeutung (Klingemann und Welzel 2002; Trüdinger und Bollow 2011). Die Bezeichnungen „links" und „rechts" übersetzen politische Sachfragen in ein relativ einfaches räumliches Muster. Dadurch kommt es zu einer erheblichen Komplexitätsreduktion, die es der Bürgerin erleichtert, sich politisch zu orientieren und politische Streitfragen einzuordnen (z. B. Arzheimer 2009; Rattinger 2009). Nach Warwick (2002) lässt sich die Mehrzahl der relevanten politischen Streitfragen durch die Links-Rechts-Skala abbilden, Niedermayer (2008) plädiert dagegen für eine zweidimensionale Lösung. Trotz dieser unterschiedlichen Auffassungen stellt die individuelle Verortung auf der Links-Rechts-Skala eine Heuristik dar, die es erlaubt, auch bei unvollständiger Information, Entscheidungen zu treffen bzw. Bewertungen abzugeben (z. B. Rudi 2010; Neundorf 2012).

S. 325). Angesichts dieser Nachteile ist es kaum überraschend, dass in standardisierten Befragungen nur selten offene Fragen eingesetzt werden.

Dichotome Frage Bei dichotomen Fragen existieren genau zwei Antwortalternativen. Häufig handelt es sich um die Antwortoptionen „Ja – Nein" oder „Vorhanden – Nicht vorhanden". In der ALLBUS 2014 findet sich folgendes Beispiel für eine dichotome Frage:

> *Haben Sie im Laufe Ihres Lebens privaten Musik- oder Gesangsunterricht erhalten, den Schulunterricht nicht mitgerechnet?*

Die Verwendung einer dichotomen Frage setzt voraus, dass die vorgegebenen Alternativen sich gegenseitig ausschließen und keine weitere Antwortmöglichkeit existiert. Die obige Frage können die Personen nur mit den Antwortoptionen „Ja" oder „Nein" beantworten.[44]

Hybridfrage Gelegentlich werden auch Fragen mit einer Kombination von geschlossenen Antwortvorgaben und einer offenen Antwortvorgabe eingesetzt, die die Möglichkeit bietet, zusätzlich zu den formulierten Antwortvorgaben „bei Bedarf" eine andere Antwort aufzuführen. Diese Frageart wird Hybridfrage genannt. Ein Beispiel ist die Frage nach dem höchsten allgemeinbildenden Schulabschluss:

> *Als nächstes kommen jetzt Fragen zu Ihrer Ausbildung und Ihrem Beruf. Beginnen wir mit Ihrer Ausbildung: Welchen allgemeinbildenden Schulabschluss haben Sie? [Nur eine Nennung möglich! Nur höchsten Schulabschluss angeben lassen!]*
> - *Noch Schüler*
> - *Schule beendet ohne Abschluss*
> - *Volks-/Hauptschulabschluss bzw. Polytechnische Oberschule mit Abschluss 8. oder 9. Klasse*
> - *Mittlere Reife, Realschulabschluss bzw. Polytechnische Oberschule mit Abschluss 10. Klasse*
> - *Fachhochschulreife (Abschluss einer Fachoberschule etc.)*
> - *Abitur bzw. Erweiterte Oberschule mit Abschluss 12. Klasse (Hochschulreife)*
> - *Anderen Schulabschluss, und zwar: . . . . . .*

44 Bei jeder Frage haben die Befragten natürlich auch die Möglichkeit, die Frage mit „weiß nicht" zu beantworten oder die Antwort zu verweigern (Item-Nonresponse). Dabei handelt es sich in der Regel aber nicht um inhaltliche Antwortoptionen.

Eine Hybridfrage bietet dem Befragten die Möglichkeit, eine individuelle Antwort zu geben, wenn die formulierten Antwortoptionen seiner persönlichen Situation nicht gerecht werden. Individuelle Antwortmöglichkeiten verlängern allerdings die Befragung und können bei der Analyse häufig nicht berücksichtigt werden.

Frage mit Skalenvorgabe Die (geschlossene) Frage mit einer Skalenvorgabe ist der „Klassiker" in den meisten standardisierten Befragungen. Die folgende Frage erfasst das politische Interesse. Die Befragte muss sich „nur" für eine der Antwortvorgaben entscheiden.

Wie stark interessieren Sie sich für Politik?
- *Sehr stark,*
- *stark,*
- *mittel,*
- *wenig oder*
- *überhaupt nicht?*

Solche geschlossenen Fragen mit klaren Antwortvorgaben haben zwei Vorteile: Einerseits können sie in einer Befragung schnell abgearbeitet und andererseits bei der Datenanalyse ohne größeren Aufwand analysiert werden.[45] Allerdings sind die Befragungspersonen manchmal mit den Antwortvorgaben nicht einverstanden und geben eine Antwort, die der Fragebogen nicht enthält.

Die nachfolgende Frage zielt darauf ab, die Zufriedenheit des Befragten mit der Demokratie zu erfassen. Der Befragte muss sich dabei für eine von sechs Antwortoptionen entscheiden.

Kommen wir nun zu der Demokratie in Deutschland: Wie zufrieden oder unzufrieden sind Sie – alles in allem – mit der Demokratie, so wie sie in Deutschland besteht?
- *Sehr zufrieden,*
- *ziemlich zufrieden,*
- *etwas zufrieden,*
- *etwas unzufrieden,*
- *ziemlich unzufrieden oder*
- *sehr unzufrieden?*

45 Die verbalen Antwortkategorien können unkompliziert in Zahlen „umgewandelt" werden: 1 steht für „sehr stark", 2 für „stark", 3 für „mittel", 4 für „wenig" und 5 für „überhaupt nicht". Bei Bedarf können die Antworten auch so kodiert werden, dass höhere Zahlen ein stärkeres Politikinteresse widerspiegeln (siehe auch Kapitel 7).

Die Frageformulierung ist durchaus umstritten (Canache et al. 2001), sie findet sich allerdings in nahezu allen nationalen und internationalen Surveys. In der Praxis haben die meisten Befragten auch kein Problem, die Frage zu beantworten (z. B. Tausendpfund 2013a).

Rangordnungsfrage Die Erhebung materialistischer und postmaterialistischer Werte basiert auf einer Rangordnungsfrage. Dem Befragten werden vier Aussagen vorgelesen, zwei gelten eher als materialistische und zwei eher als postmaterialistische Ziele. Die Frageformulierung lautet:

> *Auch in der Politik kann man nicht alles auf einmal haben. Ich lese Ihnen nun verschiedene Ziele vor, die man in der Politik verfolgen kann:*
> A. *Aufrechterhaltung von Ruhe und Ordnung in diesem Land*
> B. *Mehr Einfluss der Bürger auf die Entscheidungen der Regierung*
> C. *Kampf gegen die steigenden Preise*
> D. *Schutz des Rechtes auf freie Meinungsäußerung*
> *Wenn Sie zwischen diesen verschiedenen Zielen wählen müssten, welches Ziel wäre Ihnen persönlich am Wichtigsten? Und welches Ziel erschiene Ihnen am Zweitwichtigsten?*

Entsprechend der gewählten Prioritäten – unabhängig von der Reihenfolge der Nennung – erfolgt die Klassifikation als Materialist, Postmaterialist oder Mischtyp. Die Wahl der Items A und C hat die Einordnung als Materialist zur Folge. Die Nennung der Items B und D die Klassifikation als Postmaterialist. Befragte, die ein materialistisches (Item A oder C) und ein postmaterialistisches Ziel (Item B oder D) genannt haben, werden als Mischtyp identifiziert.[46]

Mehrfachantwortfrage Bei Fragen mit mehreren Antwortkategorien ist es möglich, dass mehr als eine Antwort zutrifft (Mehrfachnennungen). Bei der ALLBUS 2014 wurde den Befragten beispielsweise eine Liste mit häufigen Krankheiten vorgelegt und folgende Frage gestellt:

46 Dieses Erhebungsinstrument basiert auf den Arbeiten von Ronald Inglehart (1971, 1977, 1995) zum Wertewandel. Es existiert auch eine erweiterte Variante des Instruments, doch in der Forschung wird fast ausschließlich auf die Kurzversion zurückgegriffen (Schmitt-Beck 1992, S. 528). Für eine ausführliche Diskussion des Instruments siehe die Beiträge von Inglehart (1998, S. 158–188), Inglehart und Abramson (1999) sowie Inglehart und Klingmann (1996). Bei Marcus (2009) findet sich eine Übersicht, in welchen Erhebungen diese Frage eingesetzt wurde.

Bitte sagen Sie mir, an welchen Krankheiten oder Beschwerden Sie seit mindestens 12 Monaten oder chronisch leiden (Mehrfachnennungen möglich).

☐ *Allergie*
☐ *Migräne*
☐ *Bluthochdruck, Hypertonie*
☐ *Durchblutungsstörung am Herzen, Angina Pectoris*
☐ *Rheuma, chronische Gelenkentzündung, Arthritis, Arthrose, Gicht*
☐ *Wirbelsäulenschäden*
☐ *Chronische Bronchitis*
☐ *Asthma*
☐ *Leberentzündung, Hepatitis, Leberschrumpfung, Leberzirrhose*
☐ *Zuckerkrankheit, Diabetes*
☐ *Krebs*
☐ *Osteoporose*
☐ *Sonstige, und zwar:* . . . . . .
☐ *Keine chronischen Krankheiten oder Beschwerden*

Der Befragte kann jetzt mehrere Antwortoptionen auswählen. Die Möglichkeit der mehrfachen Antwort sollte bei der Befragung bzw. im Fragebogen deutlich betont werden, um Fehler bzw. Nachlässigkeiten bei der Antwort zu vermeiden. Günstiger als dieses Frageformat ist allerdings eine explizite Vorgabe für jede Antwortmöglichkeit („trifft zu“ – „trifft nicht zu“). Aus einer Mehrfachantwortfrage wird dann eine dichotome Frage. Dadurch können fehlerhafte Angaben verringert werden.

Forced-Choice-Frage Bei Forced-Choice-Fragen kann die Befragte eine Antwort aus einer Liste von möglichen Antwortoptionen wählen. Dieses Frageformat wird insbesondere bei Wissensfragen genutzt. Die Antwortvorgaben können beliebig geordnet werden.

Durch wen wird der Bundeskanzler der Bundesrepublik Deutschland gewählt?

- *Durch das Volk.*
- *Durch den Bundesrat.*
- *Durch den Bundestag.*
- *Durch die Bundesversammlung.*

Die Befragte wählt die nach ihrer Ansicht richtige Antwort aus. Bei der ALLBUS 2008 haben knapp 50 Prozent der Befragten die korrekte Antwort „Durch den Bundestag“ ausgewählt (Westle 2012, S. 57).

Sozialwissenschaftliche Items und Skalen (ZIS)

Für zahlreiche sozialwissenschaftliche Konzepte existieren bereits etablierte Fragen oder Skalen. Die Zusammenstellung sozialwissenschaftlicher Items und Skalen (ZIS) umfasst über 200 sozialwissenschaftliche Erhebungsinstrumente, zum Beispiel Items zum Messen von politischen Einstellungen oder Persönlichkeit. Dokumentiert sind die Instrumente, die Entwicklung, die Gütekriterien und Informationen zum theoretischen Hintergrund. Alle dokumentierten Instrumente sind frei zugänglich und können für Forschungszwecke eingesetzt werden. ZIS ist ein kostenfreies Angebot von GESIS. Eine Orientierung bieten auch etablierte wissenschaftliche Erhebungen (z. B. ALLBUS, ESS). Die Tatsache, dass eine Frage in eine Erhebung aufgenommen wurde, ist aber kein Qualitätskriterium. Es finden sich in der Literatur viele Beispiele für ungeeignete Frageformulierungen, die immer wieder verwendet werden (Schnell 2012, S. 106).

11.1.3 Frageformulierung

Während Payne (1958) noch von „The art of asking questions" sprach, bezeichnen Schaeffer und Presser (2003) die Frageformulierung als Wissenschaft. Die Formulierung von Fragen ist eine höchst sensible Phase bei der Konstruktion eines Fragebogens. Entsprechend existieren unzählige Veröffentlichungen zur Frageformulierung (z. B. Converse und Presser 1986; Bradburn et al. 2004; Faulbaum et al. 2009; Dillman et al. 2014; Porst 2014a; Kallus 2016). Die wichtigsten Grundregeln der Frageformung hat Porst (2014a, S. 99–118) in zehn „Geboten" zusammengefasst. Diese Gebote stellen grobe Faustregeln dar, um typische Fehler bei der Fragebogenkonstruktion zu vermeiden.

1. Gebot: Einfache Begriffe Du sollst einfache, unzweideutige Begriffe verwenden, die von allen Befragten in gleicher Weise verstanden werden!
Bei standardisierten Erhebungen wird allen Befragten der gleiche Fragebogen vorgelegt (siehe Abschnitt 11.1). Für die Erfassung theoretischer Konzepte wie Vertrauen oder Demokratiezufriedenheit ist es daher von zentraler Bedeutung, dass eine Frage von allen Befragten in gleicher Weise verstanden wird. Bei einer allgemeinen Bevölkerungsumfrage sollten die einzelnen Fragen deshalb keine Fremdwörter oder schwierige Begriffe enthalten. Begriffe wie Subsidiarität, Eurobonds oder Finanztransaktionssteuer sind einerseits für viele Befragte zu schwierig und andererseits auch nicht immer eindeutig. Bei der Wahl der Begriffe muss der Blick

immer auf diejenigen gerichtet werden, „die letztendlich die Fragen beantworten oder den Fragebogen ausfüllen sollen“ (Porst 2014a, S. 101). Die meisten Fragebögen werden von Akademikern konstruiert. Deshalb muss ein Fragebogen vor der Datenerhebung auch Personen mit niedrigerem Bildungsniveau vorgelegt werden, um schwierige Begriffe identifizieren zu können (siehe Abschnitt 11.1.6).

2. Gebot: Kurze Fragen Du sollst lange und komplexe Fragen vermeiden!
Payne (1958, S. 136) bezeichnet die Verwendung von 20 Wörtern in einer Frage als Höchstgrenze. Grundsätzlich sollten Fragen so kurz wie möglich formuliert sein. Unnötige Füllwörter und Wiederholungen sind zu streichen. Fragen sollten auch nicht „bürokratisch gestelzt klingen“ (Diekmann 2011, S. 479).

3. Gebot: Keine hypothetischen Fragen Du sollst hypothetische Fragen vermeiden!
Fragen sollten nicht hypothetisch formuliert werden. Fragen nach dem Prinzip „Was wäre, wenn“ bezeichnen Schnell et al. (2013, S. 327) schlicht als „unzulässig“. Porst (2014a, S. 104) bewertet hypothetische Fragen als vertretbar, sofern sich der Befragte schon einmal in die hypothetische Situation hineinversetzt hat. Danach sind Fragen wie „Einmal angenommen, Sie würden am Samstag in Lotto gewinnen“ akzeptabel, aber Fragen wie „Stellen Sie sich vor, Sie wären Generalsekretär der Vereinten Nationen. Was wäre Ihre erste Amtshandlung?“ sollten nicht in den Fragenkatalog aufgenommen werden.

4. Gebot: Keine Verneinungen Du sollst doppelte Stimuli und Verneinungen vermeiden!
„Wie häufig gehen Sie in Ihrer Freizeit ins Theater oder ins Kino?“ Sowohl „Kino“ als auch „Theater“ sind Freizeitaktivitäten, die aber in einer Frage nicht einfach zu beantworten sind, wenn man beispielsweise nur ins Theater, aber nie ins Kino geht. Mit „Kino“ und „Theater“ liegt ein doppelter Stimulus vor, der die Befragten überfordert. Für die Erfassung des Freizeitverhaltens müssen daher zwei Fragen gestellt werden – einmal nach der Häufigkeit des Theaterbesuchs und einmal nach der Häufigkeit des Kinobesuchs.

Bei einer Befragung werden ziemlich viele Fragen in relativ kurzer Zeit gestellt. Längeres Nachdenken über eine Frage ist in der Regel nicht möglich. Deshalb sollten Fragen nicht unnötig verkompliziert werden. Verneinungen – insbesondere doppelte Verneinungen – können zu Irritationen bei den Befragten führen. Die Frage „Stimmen Sie der Aussage zu, dass es keine Bürgerpflicht ist, sich an Wahlen zu beteiligen“ ist sicherlich komplizierter als die Formulierung „Die Wahlteilnahme ist Bürgerpflicht“.

5. Gebot: Keine Unterstellungen Du sollst Unterstellungen und suggestive Fragen vermeiden!
Fragen sollen keine bestimmte Antwort provozieren. Deshalb ist die Frage „Haben Sie den Film ‚Avatar – Aufbruch nach Pandora" gesehen?" besser als die Formulierung: „Avatar – Aufbruch nach Pandora war der erfolgreichste Film aller Zeiten. Haben Sie diesen Film gesehen?". Grundsätzlich sind Fragen neutral zu formulieren. Dabei sollte auch auf stark wertbesetzte Begriffe wie Gerechtigkeit oder Verbrechen verzichtet werden. Unabhängig vom Frageinhalt kann allein die Verwendung eines solchen Begriffs die Antwortreaktion in eine bestimmte Richtung lenken (Diekmann 2011, S. 412).

6. Gebot: Keine Überfragung Du sollst Fragen vermeiden, die auf Informationen abzielen, über die viele Befragte mutmaßlich nicht verfügen.
Wie hoch ist der jährliche Etat der NASA? Wie viele Unfälle gab es im vergangenen Jahr in Deutschland? Welches Verkehrsmittel ist in Nordrhein-Westfalen am beliebtesten? Solche Fragen stellen für die meisten Befragten eine Überforderung dar, da sie nicht über die erforderlichen Informationen verfügen, um die Frage korrekt beantworten zu können. Im besten Fall geben die Befragten ihre Unwissenheit zu Protokoll („weiß nicht"), häufig geben die Personen aber irgendeine Antwort. Die Nicht-Beantwortung von Fragen erhöht zudem die (psychologischen) Kosten der Befragungsteilnahme. Nach 15 „weiß nicht"-Antworten wird selbst die geduldigste Person überlegen, ob sie weiterhin an der Befragung teilnehmen möchte.

7. Gebot: Klarer zeitlicher Bezug Du sollst Fragen mit eindeutigem zeitlichen Bezug verwenden!
Bei Fragen, die auf Verhalten oder Einstellungen innerhalb eines bestimmten Zeitraums abzielen, muss der Zeitraum klar definiert werden. Die Frage „Haben Sie in den vergangenen Wochen an einer Unterschriftensammlung teilgenommen?" ist schlecht formuliert, da der Bezugspunkt „in den vergangenen Wochen" vage ist. Deshalb wird in der ALLBUS gefragt: „Haben Sie in den letzten 12 Monaten an einer Unterschriftensammlung teilgenommen?" Nach Porst (2014a, S. 110) sind Formulierungen wie „in der letzten Zeit", „in naher Zukunft", „früher" oder „in Ihrer Kinderzeit" mehrdeutig und sollten deshalb nicht in einem Fragebogen verwendet werden.

8. Gebot: Klare Antwortvorgaben Du sollst Antwortkategorien verwenden, die erschöpfend und disjunkt (überschneidungsfrei) sind!
Die Antwortkategorien von geschlossenen Fragen sollten nicht überlappend (disjunkt), erschöpfend und präzise sein (siehe Abschnitt 11.1.4). Antwortkategorien

sind disjunkt, wenn sich eine Person bei einer geschlossenen Frage (ohne Mehrfachantworten) genau einer Antwort zuordnen kann. Die Frage „Wie viele Stunden pro Woche haben Sie durchschnittlich für die Lektüre der Studienmaterialien aufgewendet?“ mit den Antwortoptionen „0 bis 5 Stunden“, „5 bis 10 Stunden“, „10 bis 15 Stunden“, „15 bis 20 Stunden“ und „20 Stunden und mehr“ ist problematisch, da sich die einzelnen Antwortoptionen überschneiden. Welche Antwort soll die Studierende ankreuzen, die sich 15 Stunden pro Woche mit den Lehrmaterialien auseinandersetzt?

Die Antwortoptionen sollten auch erlauben, ausreichend zwischen verschiedenen Sachverhalten unterscheiden zu können. Das individuelle (monatliche) Nettoeinkommen mit nur drei Kategorien abzufragen (z. B. „weniger als 1000 Euro“, „zwischen 1000 und 2000 Euro“ und „mehr als 2000 Euro“) dürfte für eine angemessene statistische Auswertung nicht ausreichend sein. Denn: Mit der Wahl der Antwortkategorie wird eine Vorentscheidung über das Messniveau der Variable und damit für die möglichen Analyseverfahren getroffen (Diekmann 2011, S. 411).

Dokumentation der Fragebogenentwicklung

Die Entwicklung eines Fragebogens ist ein längerer Prozess. In der Regel werden einzelne Fragen geändert, die Reihenfolge modifiziert und neue Fragen in den Fragebogen aufgenommen. Mit dem Questionnaire Development Documentation System (QDDS) kann der Fragebogenerstellungprozess mit seinen Überarbeitungen sowohl für die Datenanalyse und Datenarchivierung als auch im Hinblick auf die Weiterentwicklung in späteren Befragungen detailliert dokumentiert werden. QDDS bietet damit eine dauerhafte elektronische Dokumentation der Fragebogenentwicklung und des Endzustandes des Erhebungsinstruments. Für weitere Informationen siehe www.qdds.org. Dort steht die Software auch zum kostenlosen Download zur Verfügung.

9. Gebot: Kontext beachten Du sollst sicherstellen, dass der Kontext einer Frage sich nicht (unkontrolliert) auf deren Beantwortung auswirkt!

Die einzelnen Fragen in einem Fragebogen sind nicht isoliert, sondern müssen immer im Kontext anderer Fragen betrachtet werden. Vorausgegangene Fragen können die Antworten auf nachfolgende Fragen beeinflussen. Schwarz (1991, S. 6; siehe auch Schwarz und Bless 1992) unterscheidet dabei zwischen Assimilations- und Kontrasteffekt. Bei einem Assimilationseffekt ist die Antwort auf die zwei-

te Frage ähnlich der Antwort auf die erste Antwort, bei einem Kontrasteffekt tritt hingegen ein gegenteiliger Effekt auf: Die Antwort auf die zweite Frage wird der Antwort auf die erste Frage unähnlicher.

Ein Beispiel für einen klassischen Assimilationseffekt: Wenn die Befragten gebeten werden, ihre Zufriedenheit mit verschiedenen Lebensbereichen (z. B. Beruf, Familie) anzugeben, dann werden die Antworten auf diese Fragen die anschließende Frage nach der „allgemeinen Lebenszufriedenheit" beeinflussen. Mit anderen Worten: Die Abfrage konkreter Lebensbereiche wirkt sich auf das zusammenfassende Urteil „Allgemeine Lebenszufriedenheit" aus. In dem „zusammenfassenden" Urteil der allgemeinen Lebenszufriedenheit werden alle Informationen berücksichtigt, die den Befragten einfallen, wenn sie an ihr Leben denken. Durch die vorherige Abfrage werden spezifische „Lebensbereiche" kognitiv aktiviert, die das zusammenfassende Urteil beeinflussen. Wer also zunächst eine hohe Zufriedenheit mit Beruf und Familie berichtet, der wird anschließend auch eine höhere Lebenszufriedenheit angeben.

Ein Kontrasteffekt tritt auf, wenn ein Gegenstand (z. B. eine Person) und der Kontext, in dem er präsentiert wird, nicht als zusammengehörig empfunden werden. In einem klassischen Fragebogenexperiment (Schwarz 1991) wurden die Befragten gebeten, die CDU zu bewerten.

> *„Um dies zu tun, rufen sich die Befragten vermutlich ein Bild der CDU vor Augen und ihr Urteil sollte von den Informationen abhängen, die in diese kognitive Repräsentation eingehen." (Schwarz 1991, S. 6)*

Diese „kognitive Repräsentation" wurde manipuliert. Einer Gruppe wurde vorher die Frage „Wissen Sie zufällig, welcher Partei Richard von Weizsäcker seit mehr als 20 Jahren angehört?", einer anderen Gruppe die Frage „Wissen Sie zufällig, welches Amt Richard von Weizsäcker bekleidet, das ihn außerhalb des Parteiengeschehens stellt?" gestellt. Bei der ersten Variante antworteten fast alle Befragten mit CDU. Dadurch wurde der über die Parteigrenzen angesehene Bundespräsident Richard von Weizsäcker in die kognitive Repräsentation der CDU aufgenommen und die Partei wurde positiver bewertet. Bei der zweiten Gruppe konnte Weizsäcker nicht in das mentale Bild der CDU aufgenommen werden, da er ja ein Amt bekleidet, das ihn außerhalb des Parteiengeschehens stellt. Folglich wird die CDU von der zweiten Gruppe schlechter bewertet als von der ersten Gruppe.

In der Literatur finden sich viele Beispiele für solche Fragebogenreihenfolgeeffekte (z. B. Schwarz und Strack 1991; Weller 1996). Solche Effekte lassen sich bei der Konstruktion von Fragebögen häufig „erahnen", aber nur durch systematische Tests belegen (Porst 2014a, S. 116). Deshalb sind Pretests von Fragebögen von großer Bedeutung.

10. Gebot: Unklare Begriffe definieren Du sollst unklare Begriffe definieren! Begriffe sind dann unklar, wenn sie nicht von allen Befragungspersonen in der gleichen Weise verstanden werden (siehe Gebot 1). Wann ein Begriff unklar ist, hängt von der Personengruppe ab, die die Frage beantworten soll. Sicherlich können bei einer Befragung von Allgemeinmedizinern Begriffe wie Hypertonie, Angina pectoris oder Synkope eher vorausgesetzt werden als bei einer Befragung der allgemeinen Bevölkerung. Falls bei einer Befragung die konkreten Begriffe erforderlich sind, dann sollten Sie entsprechend definiert bzw. erläutert werden.

Fazit Porst (2014a, S. 117) bewertet seine zehn Gebote selbst nur als „allgemeine Wegweiser", die helfen können, typische Fehler bei der Konstruktion zu vermeiden. Es handelt sich auch nicht um unumstößliche Regeln, die man „blind" anwenden sollte. Die meisten der zehn Gebote lassen Interpretationsspielraum zu und stehen gelegentlich auch in Konkurrenz zueinander. Die zehn Gebote sollen allerdings für die Herausforderungen bei der Konstruktion von Fragen sensibilisieren. Die Entwicklung geeigneter Fragen für eine Erhebung ist kein Kinderspiel, das man mal „so nebenbei" erledigen kann. Denn:

> *„Schlechte Fragen führen zu schlechten Daten, und kein Gewichtungs- und kein Analyseverfahren der Welt kann aus schlechten Daten gute Ergebnisse machen." (Porst 2014a, S. 118)*

Deshalb sei an dieser Stelle an eine Aussage von Schnell (2012, S. 23) erinnert:

> *„Der notwendige Aufwand für Befragungen wird fast immer unterschätzt. Dies beginnt mit der erforderlichen Zeit. Selbst wenn der seltene Fall einer präzise definierten inhaltlichen Fragestellung gegeben sein sollte, werden allein für die Konstruktion und den Test eines Fragebogens in der Regel mehrere Wochen erforderlich sein."*

Mit Blick auf eine eigene (quantitative) Datenerhebung im Rahmen einer Abschlussarbeit kommen deshalb Kopp und Lois (2014, S. 31) zu einer eindeutigen Empfehlung:

> *„… die Zahl an Fallstricken und Problemen, denen man sich bei der Planung, Konzeptualisierung und Durchführung einer empirischen – und damit ist hier immer gemeint: einer quantitativ orientierten empirischen – Studie gegenüber sieht, so groß, dass man in aller Regel erst gar nicht versuchen sollte im Rahmen von Lehrforschungsprojekten oder gar Qualifikationsarbeiten wie Bachelor-, Master-, aber auch Promotionsarbeiten eigenständig Daten zu erheben."*

Diese Einschätzung wird an dieser Stelle ausdrücklich geteilt (siehe auch Schnell 2012, S. 204). Für empirische (quantitative) Haus- oder Abschlussarbeiten stehen mittlerweile umfangreiche Datensätze zur Verfügung, die für eine sekundäranalytische Auswertung genutzt werden können (siehe Kapitel 8). Wer sich mit dem Wahlverhalten bei Bundestagswahlen beschäftigen möchte, der kann auf Daten der GLES zurückgreifen. Wer sich mit sozialer Ungleichheit in Deutschland auseinandersetzen möchte, dem stehen Daten der ALLBUS zur Verfügung. Für ländervergleichende Analysen kann der ESS oder das Eurobarometer genutzt werden. Insgesamt existieren zahlreiche hochwertige Datensätze, die sich hervorragend für sekundäranalytische Auswertungen im Kontext von Haus- und Abschlussarbeiten eignen. Eine eigene Individualdatenerhebung ist in der Regel weder sinnvoll noch erforderlich (einzige Ausnahme: Experimente).

Geschlecht, Alter und Bildung: Demographische Standards

Die Demographischen Standards (Statistisches Bundesamt 2016) zielen darauf ab, soziostrukturelle Erhebungsmerkmale in Bevölkerungsumfragen (Haushalts- und Personenbefragungen) zu vereinheitlichen, um eine größere Vergleichbarkeit zwischen einzelnen Erhebungen zu ermöglichen. Konkret handelt es sich um Frageformulierungen und -kategorisierungen für die Erhebung klassischer soziodemographischer Merkmale wie Geschlecht, Alter, Familienstand und Bildung. Es werden Muster-Fragebögen für persönlich-mündliche, telefonische und schriftliche Befragungen bereitgestellt. Bei den Demographischen Standards handelt es sich um eine gemeinsame Empfehlung des Arbeitskreises Deutscher Markt- und Sozialforschungsinstitute (ADM), der Arbeitsgemeinschaft Sozialwissenschaftlicher Institute (ASI) sowie des Statistischen Bundesamtes. Das Dokument steht beim Statistischen Bundesamt (www.destatis.de) zum kostenlosen Download zur Verfügung.

11.1.4 Antwortvorgaben

Eine (geschlossene) Frage ohne Antwortvorgaben ist unvollständig. Das achte Gebot der Frageformulierung (Porst 2014a, S. 112) nimmt bereits die Antwortkategorien in den Blick, die erschöpfend und disjunkt sein sollen. In einem Überblicksartikel formuliert Franzen (2014, S. 702) sechs Fragen, die bei der Konstruktion von Antwortskalen zu beantworten sind (siehe auch Porst 2014a, S. 77–97):

1) Wie viele Antwortkategorien soll die Skala enthalten?
2) Sollte eine gerade oder ungerade Anzahl verwendet werden?
3) Ist die Beschriftung aller Kategorien besser als nur die Bezeichnung der Endpunkte?
4) Sollte die Skalenbeschriftung bipolar oder unipolar erfolgen?
5) Sollte die Skala positiv (zustimmend) oder negativ (ablehnend) beginnen?
6) Sollte die Verbalisierung fragenspezifisch oder standardisiert vorgenommen werden?

Ausgehend von diesen sechs Leitfragen werden Hinweise zur Konstruktion von Antwortvorgaben vorgestellt.

Anzahl der Antwortkategorien? Insbesondere bei Fragen mit Skalenvorgabe, die in der Regel das Ausmaß an Zustimmung oder Ablehnung zu einer Aussage erfassen, stellt sich die Frage nach der Anzahl der Antwortkategorien. Bei der Frage nach der Demokratiezufriedenheit werden in der ALLBUS 2014 beispielsweise sechs Antwortkategorien vorgegeben – von „sehr zufrieden" bis „sehr unzufrieden" (siehe oben). Im ESS findet sich für die identische Frage eine 11-Punkt-Skala von „äußerst unzufrieden" bis „äußerst zufrieden". Solche Skalenunterschiede erschweren zwar den direkten Vergleich der Ergebnisse, aber grundsätzlich gewährleisten beide Varianten reliable Ergebnisse. Miller (1956) nennt als optimale Anzahl an Antwortkategorien die Zahl 7 plus/minus 2. Franzen (2014, S. 706) und Faulbaum et al. (2009, S. 66) sprechen sich für fünf- oder siebenstufige Skalen aus. Bei der Anzahl der Antwortkategorien sollte auch der Befragungsmodus berücksichtigt werden. Persönliche, schriftliche und Online-Erhebungen erlauben mehr Antwortkategorien als eine telefonische Befragung, da bei telefonischen Befragungen keine Hilfsmittel zur Visualisierung eingesetzt werden können. Skalen mit mehr als zehn Skalenpunkten betrachtet Porst (2014a, S. 87) zurückhaltend, da diese eine extreme Abstraktionsfähigkeit der Befragten verlangen. Selbstverständlich kann die Zufriedenheit mit der Demokratie auch mit einer Skala von 0 bis 100 abgefragt werden. Allerdings dürften die wenigsten Befragten zwischen den Werten 34 und 37 differenzieren können.

Gerade oder ungerade Antwortskalen? Die obige ALLBUS-Antwortskala ist „gerade", die ESS-Antwortskala ist hingegen „ungerade". Eine gerade Anzahl von Antwortkategorien zwingt die Befragten zu einer eher zustimmenden oder ablehnenden Entscheidung. Ungerade Antwortmöglichkeiten enthalten dagegen eine „neutrale" Mitte. Krosnick (1991) bewertet eine Mittelkategorie als Einladung für Befragte, die keine Meinung haben, diese Kategorie zu wählen. Faulbaum et al. (2009) befürworten solche Skalenmittelpunkte, „wenn Individuen wirklich neu-

trale Positionen einnehmen können". Auch Schnell (2012, S. 95–96) spricht sich für die Aufnahme einer Mittelkategorie aus.

Klassiker der Sozialwissenschaften: Antwortskalen beeinflussen Antworten

Antwortskalen geben den Befragten Hinweise zur Beantwortung von Fragen und können das Antwortverhalten beeinflussen. Eine klassische Studie zum Einfluss von Antwortskalen auf das Antwortverhalten haben Schwarz et al. (1985) vorgelegt. Den Befragten wurde eine Frage zu ihrem täglichen Fernsehkonsum gestellt, die sich allerdings in den Antwortvorgaben unterschied. In einer Version begann der tägliche Fernsehkonsum mit einem kleinen Intervall (0 bis 0,5 Stunden, 0,5 bis 1,0 Stunden, 1 bis 1,5 Stunden, usw.), in einer anderen Version mit einem großen Intervall (bis zu 2,5 Stunden, 2,5 bis 3 Stunden, 3 bis 3,5 Stunden usw.). Die Anzahl der Antwortkategorien war in beiden Versionen identisch. In der Version, die mit einem großen Intervall begann, ergab sich allerdings ein bedeutend höherer berichteter Fernsehkonsum als bei der ersten Variante. Die Teilnehmer haben sich bei der Antwort an der vorgegebenen Antwortskala orientiert. In einem neueren Experiment konnte Franzen (2014) die Ergebnisse replizieren.

Beschriftung der Antwortkategorien? Eine verbale Beschriftung der einzelnen Antwortkategorien empfiehlt Franzen (2014, S. 706) nur bei Skalen bis zu 9 Kategorien, da es bei längeren Skalen schwierig sei, eine angemessene verbale Abstufung zu konstruieren. Bei 11-Punkt-Skalen werden meist nur die Skalenendpunkte bezeichnet (z. B. 0 = „äußerst unzufrieden" und 10 = „äußerst zufrieden").

Nach Krosnick und Fabrigar (1997) hat der Verzicht auf verbale Beschriftungen zwei Vorteile: Erstens sind numerische Werte präziser als verbale Angaben und zweitens können sie bei einer telefonischen Befragung leichter gemerkt werden. Für Faulbaum et al. (2009, S. 68) ist die verbale Bezeichnung aller Antwortkategorien allerdings näher an der Realität der Befragten.

> *„Da im Alltag Bedeutungen seltener durch Zahlen ausgedrückt werden, erscheinen verbale Benennungen natürlicher. Die Vergabe von Namen für alle Skalenpunkte kann daher dazu beitragen, die Bedeutung der Skalenpunkte zu klären und damit die Leichtigkeit erhöhen, mit der Personen antworten können." (Faulbaum et al. 2009, S. 68)*

Bei einer hohen Anzahl an Skalenpunkten ist die Beschriftung der einzelnen Antwortkategorien allerdings schwierig, zumal die Interpretation der Skalenpunkte auch vom Bildungsniveau der Befragten abhängig sein kann. Bei Rohrmann (1978) findet sich eine Übersicht für die verbale Beschriftung einzelner Antwortkategorien.

Bipolare oder unipolare Beschriftung? Bei einer bipolaren Beschriftung werden Gegensatzpaare verwendet, um die Skalenendpunkte zu benennen. Beispiele für Gegensatzpaare sind „Misstrauen – Vertrauen", „unzufrieden – zufrieden" oder auch „lehne sehr stark ab – stimme sehr stark zu". Bipolare Skalen werden auch zweidimensionale Skalen genannt, da die Skala von einer negativen Angabe (z. B. unzufrieden) über einen Mittelwert zu einer positiven Angabe verläuft (z. B. zufrieden). Bei einer unipolaren Beschriftung wird die Abstufung einzelner Kategorien eindimensional erreicht. Die Endpunkte einer Skala werden etwa als „vertraue überhaupt nicht – vertraue voll und ganz", „überhaupt nicht wichtig – äußerst wichtig" oder „überhaupt nicht zufrieden – sehr zufrieden" bezeichnet. Nach Franzen (2014, S. 707) kommen Personen bei einer Befragung mit einfachen Antwortskalen besser zurecht. Deshalb votiert er für die Verwendung unipolarer Skalen.

Die Beschriftung von Antwortskalen kann mit Zahlen kombiniert werden. Allerdings wird die gleichzeitige Präsentation von verbalen Beschriftungen und numerischen Werten kritisch gesehen (z. B. Schnell 2012, S. 93; Porst 2014a, S. 91), da Befragte den verwendeten Ziffern eine zusätzliche Bedeutung zuschreiben. Insbesondere auf negative Ziffern bzw. ein negatives Vorzeichen sollte verzichtet werden, da dieser Bereich der Skala negativer beurteilt wird als bei der Verwendung positiver Werte (z. B. Schwarz et al. 1991; Schwarz und Hippler 1995; Amoo und Friedman 2001).

Positiv zu negativ oder umgekehrt? Eine Antwortskala kann mit den positiven bzw. zustimmenden Kategorien oder aber mit den negativen bzw. ablehnenden Kategorien beginnen. Bei einer optischen Präsentation der Antwortskalen befürwortet Probst (2014a, S. 88–92) einen Verlauf von negativ zu positiv, also beispielsweise von „überhaupt nicht wichtig" zu „sehr wichtig". Bei einer akustischen Präsentation (telefonische Befragung) empfiehlt er dagegen eine umgekehrte Reihenfolge (von „sehr wichtig" zu „überhaupt nicht wichtig"). Bei einer Skala von „positiv zu negativ" belegen Krebs u. Hoffmeyer-Zlotnik (2010) einen höheren Anteil positiver Antworten als bei einer Skala von „negativ zu positiv". Diese Unterschiede bewertet Franzen (2014, S. 708) allerdings als gering. Deutlichere Effekte lassen sich bei ungeordneten mehrkategorialen Antwortskalen nachweisen. Dabei wählen Befragte verstärkt die Antwortkategorie aus, die ihnen als erstes „akzeptabel" erscheint (Krosnick 1999).

Standardisierte oder spezifische Skalen? Für weitgehend standardisierte Antwortskalen spricht die Einfachheit. Dabei wird der Fragebogen so konzipiert, dass bei den Fragen weitgehend die gleichen Antwortskalen verwendet werden (z. B. siebenstufige Zustimmungsskalen). Gegen standardisierte Antwortskalen sprechen erstens Ermüdungseffekte der Befragten, die immer wieder mit der gleichen Antwortskala konfrontiert werden. Zweitens können spezifische Antwortskalen das Frageverständnis verbessern (Franzen 2014).

Fazit *„Die optimale Antwortskala gibt es nicht."*

Dieses Fazit von Porst (2014a, S. 94) ist möglicherweise enttäuschend. Jede Antwortskala hat Vor- und Nachteile, die zudem vom Befragungsmodus abhängen. Drei Faustregeln sind mit Blick auf Fragen mit Skalenvorgabe in der Literatur allerdings (weitgehend) unstrittig: Erstens sollte die Anzahl an Antwortkategorien bei 7 plus/minus 2 liegen. Allerdings haben sich bei vielen Fragen auch 11-Punkt-Skalen etabliert (z. B. Lebenszufriedenheit). Zweitens sollten die Antwortkategorien ungerade sein. Drittens sollte auf eine gleichzeitige Präsentation von verbalen Beschriftungen und numerischen Werten verzichtet werden. Weitere Regeln zur Gestaltung von Antwortkategorien finden sich bei Franzen (2014).

11.1.5 Gestaltung des Fragebogens

Bei der Fragebogenkonstruktion sind zwei Aspekte zu unterscheiden: Zum einen die Anordnung der Frageblöcke und zum anderen das Layout des Fragebogens.

Anordnung der Frageblöcke Fragen, die ähnliche Inhalte abfragen, werden zu Frageblöcken zusammengefasst (sogenannte Module). Solche Module bündeln – in Abhängigkeit von der konkreten Forschungsfrage – beispielsweise Fragen zur Politik, zum gesellschaftlichen Engagement, zur Mediennutzung oder zum subjektiven Wohlbefinden. Fragen, die zum selben Thema gehören, werden damit hintereinander abgefragt, da ein hin- und herspringen zu Irritationen führt. Der Zusammenhang verschiedener Frageblöcke wird durch einfache Überleitungen gewährleistet (z. B. „Nun möchten wir Ihnen einige Fragen über Staat und Politik stellen"). Innerhalb eines Frageblocks werden erst allgemeine und dann spezielle Fragen gestellt. Werden innerhalb eines Moduls Einstellungs- und Verhaltensfragen gestellt, dann sollten nach Schnell (2012, S. 121) zunächst Verhaltensfragen gestellt werden. Diese sind leichter zu beantworten und die umgekehrte Reihenfolge könnte das Antwortverhalten beeinflussen. Da die Aufmerksamkeit bei einer Befragung zunächst steigt und mit zunehmender Fragedauer abnimmt (Diekmann

2011, S. 414), sollten die zentralen Themenkomplexe in der Mitte des Fragebogens stehen (Klöckner und Friedrichs 2014, S. 676).

Das erste Fragemodul bzw. die ersten Fragen in einer Befragung sind von besonderer Bedeutung. Sie sind entscheidend dafür, ob eine Zielperson zur Befragungsperson wird. Deshalb sollten die Einstiegsfragen spannend, thematisch interessant, persönlich relevant und von allen Befragten zu beantworten sein (Porst 2014a, S. 142–143).

Praktisch in jeder Befragung findet sich ein Modul mit soziodemographischen Fragen (z. B. Alter, Bildung, Beruf). Diese werden zumeist am Ende des Fragebogens gestellt, da das Modul zum Teil sensible Fragen enthält (z. B. Einkommen), die zum Abbruch der Befragung oder zur Verweigerung führen können (Klöckner und Friedrichs 2014, S. 676). Schnell (2012, S. 121) empfiehlt eine Erklärung, warum die Erhebung demographischer Fragen erfolgt. In der telefonischen Bevölkerungsumfrage des Projekts „Europa im Kontext" (van Deth und Tausendpfund 2013) wurde das soziodemographische Fragemodul mit folgender Erklärung eingeleitet:

> *„Vielen Dank für Ihre bisherigen Antworten. Sie wissen, dass wir das gesamte Interview nicht in Verbindung mit Ihrem Namen auswerten. Wichtig ist allerdings, wie einzelne Bevölkerungsgruppen denken und handeln, also Männer und Frauen, jüngere und ältere Menschen, Erwerbstätige und Rentner. Deshalb habe ich abschließend noch ein paar Fragen zu Ihrer Person und zu Ihrem Haushalt."*

Die Befragung sollte allerdings nicht unmittelbar mit der Abfrage demographischer Informationen aufhören, sondern mit einem Dankeschön an die Befragte enden. Zuvor empfiehlt Porst (2014a, S. 161) der Person die Möglichkeit zu einer abschließenden Reaktion einzuräumen (z. B. „Damit sind wir am Ende unserer Befragung. Haben Sie noch Anmerkungen zum Thema unserer Befragung oder zur Befragung selbst?").

Layout des Fragebogens Das Layout des Fragebogens ist insbesondere bei schriftlichen und Online-Befragungen von Bedeutung (z. B. Dillman et al. 2014; Porst 2014a, S. 169–176). Einerseits soll ein optisch ansprechend gestalteter Fragebogen zur Teilnahme motivieren, andererseits soll er die Navigation durch das Erhebungsinstrument erleichtern. Grundsätzlich sollte die Schrift gut lesbar sein und es sollten nicht zu viele Fragen auf einer Seite platziert werden.

Durch Filterfragen bzw. Filterführungen werden in Abhängigkeit von der Antwort auf eine Frage nachfolgende Fragen ausgelassen. Filterfragen helfen, überflüssige Fragen zu vermeiden und die Befragungszeit zu reduzieren. So ist es bei-

spielsweise nicht zielführend, Merkmale eines Ehepartners abzufragen, wenn die Befragungsperson gar keinen Ehepartner hat. Bei einer computerunterstützten Befragung ist die Filterführung (korrekte Programmierung vorausgesetzt) unproblematisch, da der Computer den Befragten bzw. den Interviewer zur nächsten Frage führt. Bei einer schriftlichen Befragung muss die Filterführung allerdings eindeutig gekennzeichnet sein, damit keine Fragen übersprungen werden.

Umfangreiche Hinweise für die Gestaltung von Fragebögen finden sich bei Schnell (2012), Dillman et al. (2014) und Porst (2014a, S. 169–176).

11.1.6 Pretest

Die Entwicklung eines gelungenen Fragebogens ist – dies sollten die vorherigen Abschnitte verdeutlicht haben – eine Herausforderung. Geeignete Fragen müssen entwickelt, angemessene Antwortkategorien ausgewählt werden und bei der Anordnung der Fragen sind Reihenfolgeeffekte zu berücksichtigen. Deshalb sind Fragen bzw. ist der Fragebogen vor der eigentlichen Datenerhebung zu testen. Diese Prüfung wird allgemein als Pretest bezeichnet (Weichbold 2014).

- **Definition: Pretest**
  Mit Pretest ist die Prüfung eines Erhebungsinstruments vor der eigentlichen Datenerhebung gemeint. Ein Pretest zielt darauf ab, die Mängel eines Erhebungsinstruments zu identifizieren, um die Qualität der Datenerhebung nicht zu gefährden. Weichbold (2014, S. 299) versteht Pretests „nicht als punktuelle Verfahren oder einmalig durchzuführende Methode (typischerweise kurz bevor ein Fragebogen ins Feld geht), sondern als eine Menge von Verfahren zur Qualitätssicherung des Erhebungsinstruments, ja des gesamten Erhebungsdesigns, die im Zuge der Fragebogenerstellung und Erhebungsplanung Anwendung finden".

Drei Formen von Pretests werden unterschieden: Kognitive Interviews, Fragebewertungssysteme und Feldpretests (auch Standard-Pretest genannt). Die drei Varianten werden zu unterschiedlichen Phasen der Fragebogenentwicklung eingesetzt. Kognitive Interviews eignen sich beispielsweise dazu, die Verständlichkeit einzelner Fragen zu überprüfen. Deshalb werden kognitive Interviews in einem frühen Entwicklungsstadium des Erhebungsinstruments eingesetzt. Das Fragebewertungssystem (FBS) ist eine umfangreiche Checkliste, um die Qualität der Fragen systematisch zu überprüfen (Faulbaum et al. 2009). Eine solche Checkliste wird im mittleren Entwicklungsstadium des Fragebogens genutzt. Kurz vor der

eigentlichen Datenerhebung wird ein Feldpretest durchgeführt. Dabei handelt es sich um eine Testerhebung unter realistischen Bedingungen der geplanten Haupterhebung (Reinecke 2014, S. 614–615).

Kognitive Interviews Kognitive Interviews werden in der Entwicklungsphase eines Fragebogens durchgeführt. Sie eignen sich besonders dazu, die Verständlichkeit einzelner Fragen zu überprüfen, Probleme der Befragten bei der Beantwortung zu identifizieren und die Ursachen dieser Schwierigkeiten aufzuzeigen (z. B. Prüfer und Rexroth 2005; Willis 2005, 2016; Lenzner et al. 2015).

> *„Das eigentliche Ziel kognitiver Interviews besteht darin, Hinweise darauf zu erhalten, ob eine Frage diejenigen Informationen generiert, die von ihrem Konstrukteur intendiert sind." (Faulbaum et al. 2009, S. 67)*

Prüfer und Rexroth (2005; siehe auch Faulbaum et al. 2009, S. 98) unterscheiden fünf Techniken eines kognitiven Interviews: Nachfragetechniken (Probing), Bewertung der Verlässlichkeit der Antwort (Confidence Rating), Paraphrasieren (Paraphrasing), Sortiertechniken (Card Sorting) und die Technik des lauten Denkens (Think Aloud).

Mit Hilfe von Nachfragetechniken wird mittels einer oder mehrerer Zusatzfragen das Verständnis der Frage oder von Frageelementen gezielt hinterfragt. Die Technik zur Bewertung der Verlässlichkeit der Antwort wird zur Überprüfung von Faktenfragen eingesetzt. Der Person wird beispielsweise folgende Frage gestellt: „Seit wie vielen Jahren leben Sie in dieser Stadt?" Mittels Nachfragen werden die Befragten aufgefordert, nach der Antwort den Grad der Verlässlichkeit der Antwort zu bewerten. Beim Paraphrasieren sollen die Personen nach der Beantwortung der zu überprüfenden Frage den Fragetext in eigenen Worten wiederholen. Daraus können dann Rückschlüsse auf das Frageverständnis gezogen werden. Die Sortiertechnik wird genutzt, um zu untersuchen, wie Befragte Begriffe verstehen. Ein Beispiel: Um zu erfahren, was Befragte unter einem Verkehrsunfall verstehen, erhalten diese mehrere Kärtchen, die verschiedene Situationen beschreiben. Bei jeder Situation muss der Befragte entscheiden, ob es sich dabei – nach seiner Meinung – um einen Verkehrsunfall handelt (z. B. „Zwei Autos stoßen auf einer Kreuzung zusammen" oder „Ein Busfahrgast stürzt beim Aussteigen und bricht sich ein Bein"). Bei der Technik des lauten Denkens werden die Befragten gebeten, ihre Gedanken im Zuge der Beantwortung der Fragen auszusprechen. Auf diese Weise sollen Frageverständnis und Antwortprozess offengelegt werden, um Hinweise auf mögliche Probleme mit dem Fragenverständnis zu erhalten.

GESIS Pretest-Datenbank

Die Pretest-Datenbank dokumentiert die Ergebnisse von kognitiven Fragebogenpretests, die bei GESIS durchgeführt wurden. Die Datenbank bietet damit die Möglichkeit, sich über die Stärken und Schwächen bereits getesteter Survey-Fragen zu informieren und nachzuvollziehen, welche Konzepte mit den Fragen erfasst werden. So lassen sich zu den getesteten Survey-Fragen beispielsweise folgende Informationen finden: Wie verstehen Befragte die Frage oder bestimmte Begriffe? Wird die Frage im intendierten Sinne beantwortet? Ist die Frage einfach zu beantworten? Die Ergebnisse können entweder projektweise durchsucht oder mit Hilfe der Suchfunktion recherchiert werden. Weitere Informationen finden sich auf der Homepage von GESIS: http://pretest.gesis.org.

Kognitive Interviews finden selbstverständlich nicht unter Feldbedingungen, sondern im Labor, also unter künstlichen Bedingungen, statt. GESIS führt kognitive Interviews üblicherweise mit kleinen Fallzahlen von 15 bis 30 Testpersonen durch. Die Einzelinterviews sind auf eine Gesamtdauer von 60 bis 90 Minuten angelegt, in deren Rahmen etwa 30 Fragen bzw. Items mittels kognitiver Pretesttechniken getestet werden.

Fragebewertungssystem Mit Fragebewertungssystemen (FBS) werden Defizite von Fragen und Antwortvorgaben in einem Fragebogen systematisch identifiziert. Faulbaum et al. (2009, S. 111) betrachten das Fragebewertungssystem (FBS) als ein Evaluationsinstrument, „mit dem systematisch Fragen eines Erhebungsinstruments auf Optimierungsmöglichkeiten hin überprüft werden können". Beim FBS handelt es sich um eine Checkliste. Diese Checkliste ist in zwölf Bereiche gegliedert (z. B. Probleme mit Worten bzw. Texten, unzutreffende Annahmen über Befragte und Erinnerungsvermögen). Für jeden Bereich werden verschiedene Aspekte näher beschrieben. Diese Checkliste wird für jede einzelne Frage verwendet – ein äußerst zeitaufwändiges Verfahren. Eine ausführliche Fassung eines FBS findet sich bei Faulbaum et al. (2009, S. 111–245).

Feldpretest Beim Feldpretest (Standard-Pretest) wird das Erhebungsinstrument unter möglichst realistischen Bedingungen vor der geplanten Datenerhebung eingesetzt. Es handelt sich um einen abschließenden Test des Fragebogens mit einer Stichprobe von etwa 10 bis 200 Personen. Dieser Pretest wird im selben Befragungsmodus durchgeführt, der auch in der Haupterhebung vorgesehen ist.

Ziel des Feldpretests ist es, den gesamten Ablauf einer Erhebung zu überprüfen. Wie lange dauert die durchschnittliche Befragung? Gibt es Probleme bei der Filterführung? Haben die Interviewer Schwierigkeiten bei der Handhabung des Erhebungsinstruments? Ein solcher Feldpretest oder auch Standard-Pretest ist für jede Datenerhebung obligatorisch. Es handelt sich um ein passives Verfahren, bei dem das Verhalten der Befragten lediglich beobachtet wird, ohne das Verständnis von Fragen aktiv zu hinterfragen. Verständnisprobleme auf Seiten der Befragten werden nur aufgedeckt, wenn der Interviewte dies selbst anspricht. Trotz der geringen Stichprobengröße lassen sich erste Rückschlüsse über Häufigkeitsverteilungen ziehen. Ein hoher Anteil an Antwortverweigerung oder „weiß nicht"-Angaben kann auf Probleme beim Frageverständnis hindeuten.

Fazit Generell sollte kein Fragebogen ohne Pretest bei einer Datenerhebung eingesetzt werden. Deshalb sind bereits bei der Studienplanung Pretests zu berücksichtigen. Pretests sind wichtige Instrumente der Qualitätskontrolle. Entsprechend sollten die Ergebnisse von Pretests ernst genommen werden. Ungeeignete Fragen oder Antwortvorgaben sind zu verändern. Auf Grundlage eines Pretest neuformulierte Fragen sind wieder in einem Pretest zu überprüfen, sodass unter Umständen mehrere Pretest-Runden erforderlich sind (Weichbold 2014, S. 303).

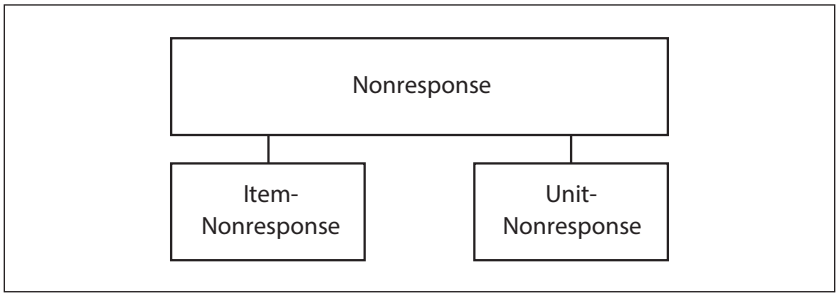
11.1.7 Nonresponse

Im Mittelpunkt einer Befragung steht der Befragte. Die Befragte muss die Fragen verstehen, mit den Antwortvorgaben zurechtkommen und elegant durch den Fragebogen geführt werden. Vor allem muss die (hoffentlich) zufällig ausgewählte Zielperson aber auch die Fragen beantworten. Die Nichtantwort bzw. Nichtbeteiligung an einer Befragung wird allgemein als Nonresponse bezeichnet.

- **Definition: Nonresponse**
  Schnell (1997, S. 17) bezeichnet „das Fehlen von Daten eines Teils oder aller Variablen für die ausgewählten Einheiten einer Stichprobe als Nonresponse". Je nach Ausmaß der fehlenden Daten wird zwischen Item-Nonresponse und Unit-Nonresponse unterschieden.

In Abbildung 31 werden die beiden Varianten von Nonresponse dargestellt. Bei Item-Nonresponse hat der Befragte an der Erhebung teilgenommen und einzelne Fragen nicht beantwortet. Bei Unit-Nonresponse hat der (ausgewählte) Befragte die Teilnahme an der Befragung verweigert.

Abbildung 31 Nonresponse



Quelle: Eigene Darstellung

Item-Nonresponse Bei Item-Nonresponse werden einzelne Fragen bei einer Befragung nicht beantwortet. Der Befragte nimmt an einer Befragung teil und beantwortet die meisten Fragen. Allerdings möchte er einzelne Fragen (Items) nicht beantworten – er verweigert die Angabe einer „zulässigen" Antwort. Ein klassisches Item mit einer hohen Verweigerungsquote ist die Frage nach dem Einkommen. Auch intime Fragen (z. B. Sexualverhalten, Drogenkonsum) werden häufiger verweigert. Diese werden von vielen Befragten als Eingriff in die Privatsphäre betrachtet (Proner 2011, S. 33). Anstatt einer expliziten Verweigerung geben viele Befragte häufig auch einfach an, man wisse keine Antwort („weiß nicht"). Ein hoher Anteil an „weiß nicht"-Angaben in einer Erhebung könnte daher auch auf Verweigerung zurückgeführt werden. Für das Problem Item-Nonresponse und die damit fehlenden Daten gibt es in der Forschungspraxis verschiedene Lösungsstrategien. Diese können danach unterschieden werden, zu welchem Zeitpunkt sie eingesetzt werden – bei der Datenerhebung oder der Datenanalyse.

Bei der Datenerhebung bietet die Randomized-Response-Technik die Möglichkeit, den Anteil an Item-Nonresponse bei sensiblen Fragen zu verringern. Die Randomized-Response-Technik gewährleistet bei einer Befragung vollständige Anonymität. Vor einer sensiblen Frage entscheidet ein Zufallsmechanismus (z. B. Münzwurf) mit bekannter Wahrscheinlichkeit der Ergebnisse, ob der Befragte eine Frage zum Ladendiebstahl oder zur Kalenderwoche (gerade versus ungerade) beantworten soll. Der Befragte wirft vom Interviewer verdeckt eine Münze. Bei „Kopf" soll er wahrheitsgemäß die Frage zum Ladendiebstahl, bei „Zahl" die Frage zur Kalenderwoche beantworten. Die vollständige Anonymität ist gewährleistet, da der Interviewer nicht wissen kann, auf welche Frage sich die Antwort bezieht. Mit Hilfe der Wahrscheinlichkeitsrechnung können die Anteile in der Stichprobe geschätzt werden (Diekmann 2011, S. 488–495).

Bei der Datenanalyse existieren mittlerweile zahlreiche (statistische) Verfahren zum Umgang mit fehlenden Daten (für eine Übersicht siehe z. B. Göthlich 2009; Spieß 2010). Am häufigsten werden sogenannte Imputationsverfahren genutzt, bei denen die bekannten Informationen eines Befragten (z. B. Geschlecht, Alter, Bildung) zur Schätzung der fehlenden Information genutzt werden (z. B. Einkommen). Bei der einfachen Imputation wird nur ein Wert geschätzt, bei der multiplen Imputation werden mehrere Werte geschätzt. Die multiple Imputation stellt dabei das Standardvorgehen dar (zur Methode siehe z. B. Rubin 1987; King et al. 2001; Allison 2002; Graham 2009; Waal et al. 2011).

Unit-Nonresponse Während bei Item-Nonresponse in der Regel nur wenige einzelne Angaben eines Befragten fehlen (z. B. zum Einkommen), stehen bei Unit-Nonresponse meistens überhaupt keine Informationen über die Person zur Verfügung.[47] Es handelt sich um den vollständigen Ausfall einer Erhebungseinheit (Stoop 2016). Bei Unit-Nonresponse werden traditionell drei Formen unterschieden: Nicht-Erreichte, Nicht-Befragbare und Verweigerer (Proner 2011, S. 37; ähnlich auch Schnell 2012, S. 157).

Nicht-Erreichte sind diejenigen Personen, die innerhalb einer Feldzeit vom Interviewer nicht kontaktiert bzw. nicht erreicht werden können. Dies sind insbesondere mobile Personen (jüngere Personen, Berufspendler). Der Anteil an Nicht-Erreichten variiert natürlich in Abhängigkeit von der Feldzeit und der Kontakthäufigkeit. Bei einer längeren Feldphase und häufigen Kontaktversuchen sollte der Anteil an Nicht-Erreichten geringer sein als bei einer kurzen Feldphase und wenigen Kontaktversuchen. Bei einer telefonischen Erhebung sind bis zu zehn Kontaktversuche (an unterschiedlichen Tagen und Uhrzeiten) üblich.

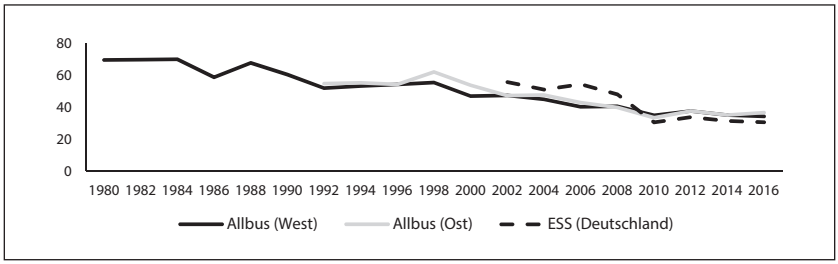
Zu den Nicht-Befragbaren werden diejenigen Personen gezählt, mit denen beispielsweise aufgrund von Sprachbarrieren keine Verständigung möglich ist oder aufgrund einer Erkrankung keine Befragung durchgeführt werden konnte.

Verweigerer sind die Personen, die angetroffen bzw. erreicht werden und mit denen ein Gespräch möglich ist, die aber eine Teilnahme verweigern. Die Gründe für eine Verweigerung sind vielfältig. Schnell (2012, S. 158–159) nennt unter anderem die zeitliche Belastung durch die Befragung, die wahrgenommene Verletzung der Privatsphäre, Desinteresse oder auch Angst vor möglichen Konsequenzen.

Entwicklung der Response-Quote Wie Abbildung 32 zeigt, ist die Beteiligung an allgemeinen Bevölkerungsumfragen in Deutschland rückläufig. In den 1980er Jahren lag die Beteiligungsrate bei der ALLBUS noch bei über 70 Prozent, aktuell

47 Bei einer Einwohnermeldeamtsstichprobe stehen für alle Personen einer Stichprobe zumindest Angaben zum Geschlecht und zum Alter zur Verfügung.

Abbildung 32 Entwicklung der Beteiligungsrate in Deutschland (Anteile in Prozent)



Quelle: Eigene Darstellung.

liegt die Beteiligungsquote zwischen 30 und 40 Prozent. Vereinfacht formuliert: Von etwa 100 kontaktierten Personen beteiligen sich etwa 30 bis 40 Personen an der Befragung. Auch beim European Social Survey (ESS) ist die sogenannte Response-Quote rückläufig. Von 55,7 Prozent (2002) ist die Quote auf 30,6 Prozent (2016) gesunken.

Ausschöpfung In der Praxis gilt die Ausschöpfung (Response-Rate) als eines der wichtigsten Merkmale zur Beurteilung der Qualität von allgemeinen Bevölkerungsumfragen. Unter Ausschöpfung wird dabei – vereinfacht formuliert – das Verhältnis der realisierten Interviews zur Anzahl der tatsächlich gezogenen Elemente aus einer Auswahlgesamtheit verstanden. Die Definition der Ausschöpfungsquote ist allerdings uneinheitlich. Deshalb bewertet Schnell (2012, S. 163) die ausschließliche Angabe der Ausschöpfungsrate als Indikator für die Bewertung der Qualität einer Erhebung als „unzureichend". Statt einer einfachen Zahlenangabe sollte ein umfassender Methodenbericht der Erhebung vorliegen, der eine Darstellung der Stichprobenziehung sowie eine detaillierte Aufgliederung aller „Ausfallursachen" enthält.

Die American Association for Public Opinion Research (AAPOR) hat Definitionen von Ausschöpfungsraten vorgeschlagen. An dieser Stelle wird die Response-Rate 1 (RR 1) dargestellt, die auch als minimum response rate bezeichnet wird. Diese ist definiert als:

$$RR\,1 = \frac{I}{(I + P) + (R + NC + O) + (UH + UO)}$$

I sind die vollständig realisierten Interviews. Die Anzahl der vollständig realisierten Interviews wird dividiert durch die Summe aus vollständig realisierten Inter-

views (I), unvollständig realisierten Interviews (P), Verweigerern (R), Nicht-Erreichten (NC), andere Ausfallgründe (O) sowie unbekannte Haushalte (UH) und andere Ursachen (UO). Der Unterschied zwischen O und UH bzw. UO besteht darin, dass bei O bekannt ist, dass die Zielperson zur Grundgesamtheit gehört und nicht verweigerte, es aber trotzdem zu keiner Befragung kam (z. B. Sprachproblem). Bei UH bzw. UO ist nicht bekannt, ob überhaupt eine Zielperson existiert bzw. diese zur Grundgesamtheit gehört (Schnell 2012, S. 163). Dieses Problem tritt auf, wenn „gezogene" Adressen nicht bearbeitet wurden. Für die Einordnung eines „Ausfalls" als R, NC oder O existieren umfangreiche Dispositionscodes, die sich in Abhängigkeit von der Befragungsart unterscheiden (AAPOR 2011, S. 57–60).

Konsequenzen von Nonresponse Die rückläufige Entwicklung der Beteiligungsrate ist in erster Linie auf einen Anstieg der Verweigerungsraten zurückzuführen (Schnell 1997, S. 84; Aust und Schröder 2009; Lynn 2009; Proner 2011, S. 39). Deshalb steht die Gruppe der Verweigerer auch im Mittelpunkt der Nonresponse-Forschung. Mit Hilfe von aufwändigen Maßnahmen wird versucht, die Beteiligung an Erhebungen zu fördern (Neller 2005; Meier et al. 2005; Blohm und Koch 2013; Pforr et al. 2015; Stähli und Joye 2016; siehe auch Schnell 2012, S. 180–184 für Vorschläge zur Verringerung von Nonresponse).

Eine geringe Ausschöpfung muss nicht zwingend negative Auswirkungen auf die Datenqualität haben. Die Deutsche Forschungsgemeinschaft weist in einer Denkschrift darauf hin, dass die Ausschöpfung nur misst, „wie groß der Spielraum für Selektivität durch nonresponse ist. Sie besagt nichts über die tatsächliche Selektivität" (Deutsche Forschungsgemeinschaft 1999, S. 104). Selektivität entsteht erst, wenn sich Teilnehmer und Nicht-Teilnehmer an einer Erhebung systematisch unterscheiden. Forschungsergebnisse deuten darauf hin, dass zufallsbasierte Stichproben mit einer geringen Ausschöpfung keine systematischen Verzerrungen aufweisen müssen (z. B. Krosnick 1999, S. 540; Schneekloth und Leven 2003; Proner 2011). Dennoch stellt Nonresponse eine zentrale Herausforderung für die Surveyforschung dar (siehe auch Weichbold et al. 2009; Schupp und Wolf 2015).

11.2 Beobachtung

Die Beobachtung stellt eine weitere Datenerhebungsmethode in den Sozialwissenschaften dar. Bei einem weiten Verständnis lassen sich alle Datenerhebungsverfahren als Beobachtung begreifen. So kann etwa eine Antwort bei einer Befragung als Beobachtung der verbalen Reaktion auf eine gestellte Frage betrachtet werden. Im engeren Verständnis ist Beobachtung der organisierte Prozess der Erfassung von sozialem Handeln (Weischer 2015b, S. 38).

- **Definition: Beobachtung**
  Beobachtung ist „das direkte, unmittelbare Registrieren der für einen Forschungszusammenhang relevanten Sachverhalte. Bei den beobachteten Sachverhalten kann es sich um Sprache, um bestimmte Verhaltensweisen, auch um nonverbales Verhalten (Mimik, Gestik), um soziale Merkmale (Fahrzeugtyp, Kleidung, Abzeichen und ähnliche Symbole) und um geronnenes Verhalten (Schaufensterauslagen, Klingelschilder und ähnliche Dinge) handeln" (Häder 2015, S. 306).

Beobachten ist eine alltägliche Angelegenheit. Auf dem Balkon wird das Spielen der Kinder beobachtet, an der Kasse des Supermarkts das Verhalten des Personals verfolgt und im Zug das laute Telefonieren des Mitfahrers registriert. Jede Person hat durch diese alltäglichen Wahrnehmungen eine gewisse Erfahrung im Beobachten. Diese alltäglichen Formen der Beobachtung sind allerdings noch keine Beobachtung im wissenschaftlichen Verständnis. Zur besseren Differenzierung zwischen der alltäglichen und der wissenschaftlichen Beobachtung bezeichnen Behnke et al. (2010, S. 257) die alltägliche Beobachtung als „Wahrnehmung". Bei Häder (2015, S. 307) finden sich vier Kriterien, um eine wissenschaftliche Beobachtung von einer Alltagswahrnehmung abzugrenzen (siehe auch Behnke et al. 2010, S. 257–258).

Vier Kriterien einer wissenschaftlichen Beobachtung Erstens bilden Hypothesen die Grundlage für eine wissenschaftliche Beobachtung. In den Hypothesen werden die Zusammenhänge beschrieben, die mit Hilfe einer Beobachtung untersucht werden sollen. Die wissenschaftliche Beobachtung erfüllt damit ein konkretes Ziel. Die Formulierung von Hypothesen setzt eine Auseinandersetzung mit Theorien und empirischen Befunden vor der eigentlichen Beobachtung voraus.

Zweitens müssen wissenschaftliche Beobachtungen einer Form der Kontrolle unterzogen werden. Die Ergebnisse der Beobachtungen mehrerer Personen müssen verglichen oder die Beobachtung muss systematisch dokumentiert werden.

Drittens ist die Auswahl der zu beobachtenden Einheiten nach wissenschaftlichen Kriterien zu begründen.

Viertens sind wissenschaftliche Beobachtungen so zu konzipieren, dass „sie einer systematischen, intersubjektiv nachvollziehbaren Auswertung und Replikation unterzogen werden können" (Häder 2015, S. 307).

Vorteile von Beobachtungen Im Vergleich zur Befragung haben wissenschaftliche Beobachtungen mehrere Vorteile (Häder 2015, S. 309). Erstens wird bei Beobachtungen tatsächliches Verhalten erfasst, bei Befragungen wird hingegen Verhalten bzw. werden Verhaltensabsichten berichtet. Berichtetes Verhalten und

tatsächliches Verhalten müssen zudem nicht immer übereinstimmen (z. B. LaPiere 1934; Diekmann 2011, S. 575). Zweitens sind Beobachtungen gut geeignet, wenn unbewusste Verhaltensweisen dokumentiert werden sollen. Drittens können mit Beobachtungen auch nonverbale Verhaltensweisen (z. B. Mimik, Gestik) dokumentiert werden.

Nachteile von Beobachtungen Diesen Vorteilen stehen allerdings auch mehrere Nachteile gegenüber (Häder 2015, S. 309–310; siehe auch Westle und Krumm 2009, S. 270–271). Erstens sind Beobachtungen im Vergleich zu Befragungen zeit- und kostenintensiver. Deshalb ist die Zahl der Untersuchungsobjekte (z. B. Personen) bei Beobachtungen in der Regel deutlich geringer als bei einer allgemeinen Bevölkerungsumfrage. Zweitens ist bei einer Beobachtung die Untersuchung sogenannter repräsentativer Stichproben kaum möglich. Dadurch ergeben sich Grenzen bei der Verallgemeinerung der Befunde. Drittens können mit einer Befragung potenziell mehr Themenbereiche untersucht werden als mit einer Beobachtung, da sich nicht jeder interessante sozialwissenschaftliche Sachverhalt beobachten lässt (z. B. keinen Zugang). Viertens können Beobachterfehler die Validität und Reliabilität der erhobenen Daten beeinflussen. In dieser Hinsicht ist der Forschungsstand zur Befragung deutlich umfassender als zur Beobachtung. Eine allgemeine Theorie der Beobachtung liegt indes noch nicht vor (Schnell et al. 2013, S. 381). Fünftens ist insbesondere bei verdeckten Beobachtungen auf ethische Grenzen hinzuweisen.

11.2.1 Beobachtung in den Sozialwissenschaften

In der quantitativ orientierten Sozialforschung wird die Beobachtung als Methode der Datenerhebung eher zurückhaltend genutzt. Das Verfahren der Beobachtung wird auf der Grundlage von drei Studien illustriert.

„Die Arbeitslosen von Marienthal“ In der klassischen Studie „Die Arbeitslosen von Marienthal“ (Jahoda et al. 1975 [1933]) wird die Beobachtung – neben anderen Methoden der Datenerhebung – genutzt, um die Konsequenzen der Arbeitslosigkeit zu untersuchen (siehe Abschnitt 2.5). Im Rahmen einer Kleideraktion besuchten die Forscherinnen 100 Familien, um nach benötigten Kleidungsstücken zu fragen. In Beobachtungsprotokollen wurde unter anderem die Größe, Sauberkeit und Ordnung der Wohnung erfasst. Die Unterkunft von Familie 363 wird wie folgt beschrieben (Jahoda et al. 1975 [1933], S. 68):

> *„Die Wohnung – Zimmer, Küche, Vorraum in einer Baracke – ist in furchtbarem Zustand. Sehr schmutzig und unaufgeräumt. Die Kinder und die Erwachsenen haben fast nichts zum Anziehen.“*

Im Hausbesuchsprotokoll findet sich folgende Passage zur Unterkunft von Familie 467 (Jahoda et al. 1975 [1933], S. 69):

> *„Die Wohnung – Zimmer, Kabinett, Küche – ist sehr schön gehalten, die Kleider aller Familienmitglieder machen einen guten, sauberen Eindruck."*

Eine weitere abhängige Variable in der klassischen Untersuchung war der Umgang mit der Zeit. Welche Konsequenzen hat die Arbeitslosigkeit für die Zeitverwendung? Die Forscher ermittelten aus einem verborgenen Fensterplatz die Gehgeschwindigkeit und die Häufigkeit des Stehenbleibens. Die Ergebnisse zeigen deutliche geschlechtsspezifische Unterschiede, die Jahoda et al. (1975 [1933], S. 84) wie folgt interpretieren:

> *„Doppelt verläuft die Zeit in Marienthal, anders den Frauen und anders den Männern. Für die letzteren hat die Stundeneinteilung längst ihren Sinn verloren. Aufstehen – Mittagessen – Schlafengehen sind die Orientierungspunkte im Tag, die übriggeblieben sind."*

Arbeit strukturiert den Tagesablauf. Zur damaligen Zeit war dies bei Frauen die Hausarbeit und bei Männern die Fabrikarbeit. Der Verlust des Arbeitsplatzes hatte daher Auswirkungen auf die Zeitstruktur der Männer. Die Ergebnisse der Studie wurden auch mit alternativen Methoden abgesichert.

Politiker im Wahlkreis Im Rahmen des Forschungsprojekts „Citizens and Representatives in France and Germany" (CITREP) wurden 64 Abgeordnete des Deutschen Bundestages für jeweils drei Tage bei ihrer Wahlkreisarbeit begleitet (Siefken 2013). Dabei wurde jede begleitete Veranstaltung mit Hilfe eines einheitlichen Erhebungsbogens erfasst. In offenen Fragen und in den frei zu formulierenden Protokollen konnten die Beobachter zudem ihre persönlichen Eindrücke schildern, aussagekräftige Zitate aufnehmen oder Situationen detailliert darstellen. Die Beobachter sammelten unter anderem Daten zur Anwesenheit von Medienvertretern, zur Anzahl der Teilnehmenden sowie zur Relevanz einzelner Politikfelder. Insgesamt liegen Informationen zu 618 Veranstaltungen mit einer Nettodauer von 969 Stunden vor.

Die Analyse der beobachteten Veranstaltungen zeigt, dass die Inhalte der Wahlkreisarbeit auch von Themen jenseits der Bundespolitik geprägt sind. Insbesondere die politische Situation und Sachthemen vor Ort nehmen einen größeren Raum ein: Bei 49 Prozent der Veranstaltungen haben diese einen großen oder sehr großen Anteil, gefolgt von der Bundespolitik (35 Prozent), der Kommunal- (21 Prozent) und Landespolitik (20 Prozent). Eine untergeordnete Rolle spielen

die Dienstleistungen für Einzelpersonen sowie Themen der internationalen Politik (Siefken 2013, S. 492).

Zusätzlich zur Beobachtung wurde mit den Abgeordneten auch ein leitfadengestütztes Interview geführt. Dies ermöglicht einen Vergleich der Beobachtungen mit den Angaben in der Befragung.

Beobachtungen bei Parteitagen Die Inszenierung von Europawahlparteitagen nimmt die teilnehmende Beobachtung von Nieland (2011) in den Blick. Er vergleicht die Parteitage der Grünen und der Linken, um Informationen über die Befindlichkeit der Partei, über die Stimmung, aber auch innerparteiliche Spannungen zu erhalten. Seine stark explorativ angelegte Beobachtung enthält auch eine Beschreibung der Gestaltung der Veranstaltungen.

> *„Das Setting der Bundesdelegiertenkonferenz war schlicht. Der Bundesvorstand nahm erhöht rechts und links sowie leicht nach hinten versetzt zu dem zentral aufgestellten Rednerpult Platz. Die Parteifarben prägten den Bühneneindruck, das Hintergrundbild war mit Wahrzeichen europäischer Hauptstädte aufgelockert." (Nieland 2011, S. 278)*

> *„Auch die Partei Die Linke verfügte über ein professionelles Bühnenbild. Zentral war der mit großen weißen Buchstaben auf rotem Grund gehaltene Schriftzug der Partei als Hintergrund zu sehen. Die Parteitagsregie hatte sich für bunte Projektionen, die Flaggen der EU-Mitgliedstaaten zeigten, in der Mitte des Bühnenbildes und zwei große Videoleinwände entschieden. Auf der großen Bühne in der Essener Grugahalle nahm auf sechs Plätzen die Parteiführung Platz. Das Rednerpult befand sich in der Mitte der Bühne." (Nieland 2011, S. 279)*

Die Inszenierung der zwei Europawahlparteitage war durchaus vergleichbar, aber bei der inhaltlichen Auseinandersetzung konnte Nieland (2011) erhebliche Unterschiede feststellen. Während bei den Grünen die europapolitischen Themen im Vordergrund standen, fand bei den Linken eine konstruktive Debatte über Europa nicht statt.

11.2.2 Varianten der Beobachtung

Allgemein werden fünf Varianten der Beobachtung unterschieden: Teilnehmende versus nichtteilnehmende Beobachtung, offene versus verdeckte Beobachtung, Feld- versus Laborbeobachtung, strukturierte versus unstrukturierte Beobachtung sowie Fremd- versus Selbstbeobachtung (z. B. Pötschke 2010, S. 56; Diekmann

2011, S. 564; Häder 2015, S. 310–312; Kromrey et al. 2016, S. 328). Diese fünf Formen der Beobachtung werden näher vorgestellt.

Teilnehmend versus Nichtteilnehmend Bei einer teilnehmenden Beobachtung ist der Beobachter selbst am Geschehen der Gruppe beteiligt. Er wird selbst zum Bestandteil der beobachteten Gruppe. Eine teilnehmende Beobachtung kann detaillierte Einblicke in die Gruppendynamik ermöglichen, möglicherweise werden durch die eigene Beteiligung aber auch die Personen beeinflusst. Zwar kann sich der Beobachter bei einer teilnehmenden Beobachtung stärker mit den Personen identifizieren, allerdings besteht dadurch auch die Gefahr, die wissenschaftliche Distanz zum Untersuchungsobjekt zu verlieren. Der Prozess der starken Identifikation des Beobachters mit den beforschten Personen wird als „going native" bezeichnet (Behnke et al. 2010, S. 261). Bei einer nichtteilnehmenden Beobachtung nimmt der Beobachter eine passive Rolle ein. Er beteiligt sich nicht an den Interaktionen und hält zum Geschehen stärkeren räumlichen Abstand.

Offen versus Verdeckt Eine Beobachtung kann offen oder verdeckt sein. Bei einer verdeckten Beobachtung ist der Beobachter als solcher nicht zu erkennen. Den beobachteten Personen ist nicht bewusst, dass ihr Verhalten beobachtet wird. Dadurch ist das Verfahren nichtreaktiv. Allerdings sind verdeckten Beobachtungen ethische Grenzen gesetzt. Bei einer offenen Beobachtung sind die Personen über die Beobachtung informiert. Offene und verdeckte Beobachtungen sind nur Endpunkte dieser Dimension. Denkbar sind auch Situationen, in denen einige Personen über die Beobachtung informiert sind, andere aber nicht.

Feld versus Labor Von einer Feldbeobachtung wird gesprochen, wenn die Beobachtung unter natürlichen Bedingungen stattfindet. Eine Beobachterin kann beispielsweise das Verhalten der Politiker bei einem Parteitag beobachten. Laborbeobachtungen finden dagegen in einer künstlich geschaffenen Umgebung statt. Solche künstlichen, das heißt gezielt hergestellten, Beobachtungssituationen finden sich häufig bei experimentellen Studien. Den Versuchspersonen wird ein Treatment verabreicht (z. B. gewalttätiges Video versus Naturfilm) und anschließend wird das Verhalten der Personen beobachtet.

Strukturiert versus Unstrukturiert Bei einer strukturierten Beobachtung wird ein Beobachtungsprotokoll oder Beobachtungsschema eingesetzt. Die zu erfassenden Sachverhalte sind vorab festgelegt. Dadurch wird die Aufmerksamkeit des Beobachters auf bestimmte Ausschnitte der Wirklichkeit gelenkt. Solche Sachverhalte können etwa die Anzahl, Dauer, Reihenfolge und Intensität der Wortbeiträ-

ge einzelner Teilnehmer bei einer Entscheidungsfindung sein. Bei einem solch strukturierten Vorgehen können die Beobachtungen verschiedener Personen miteinander verglichen und die Zuverlässigkeit einer Beobachtung ermittelt werden. Bei einer unstrukturierten Beobachtung erfolgt das Vorgehen eher unsystematisch und dem spontanen Interesse des Beobachters. Ein solches Vorgehen bietet sich beispielsweise zur Exploration eines zunächst unbekannten Forschungsgegenstands an. Zunächst werden viele Eindrücke gesammelt, um nicht relevante Kategorien zu übersehen. Diese Erfahrungen werden genutzt, um ein Beobachtungsschema zu entwickeln. Qualitative Sozialforscherinnen nutzen entsprechend eher eine unstrukturierte Beobachtung als Forschungsmethode, quantitative Sozialforscherinnen arbeiten stärker strukturiert.

Selbst- versus Fremdbeobachtung Schließlich wird noch zwischen Selbst- und Fremdbeobachtung differenziert. Bei einer Selbstbeobachtung wird auf Beobachter verzichtet und die Person beobachtet sich selbst. Dies wird auch Introspektion genannt (Häder 2015, S. 311). Der Regelfall ist die Fremdbeobachtung, bei der geschulte Beobachter soziales Verhalten anderer Personen beobachten.

11.2.3 Phasen einer Beobachtung

Bei Diekmann (2011, S. 574) und Häder (2015, S. 318–327) finden sich exemplarische Darstellungen der Entwicklung und Durchführung einer standardisierten Beobachtungsstudie. Die grundsätzlichen Schritte sind dabei mit den Phasen des quantitativen Forschungsprozesses vergleichbar. Folgende Schritte werden unterschieden: Fragestellung und Entwicklung von Hypothesen, Operationalisierung, Entwicklung eines Kodierschemas, Auswahl der Untersuchungsobjekte, Pretest, Haupterhebung sowie Datenaufbereitung und Auswertung.

Fragestellung und Hypothesen Vor der Datenerhebung sind Forschungsfrage und überprüfbare Hypothesen zu formulieren. Dies setzt die Auseinandersetzung mit dem existierenden Forschungsstand voraus. Nur wer einschlägige Theorien und empirische Befunde kennt, kann gehaltvolle Hypothesen entwickeln. Diese Hypothesen werden dann mittels einer Beobachtung empirisch überprüft.

Operationalisierung Zuvor müssen die theoretischen Konzepte in empirisch beobachtbare Sachverhalte „übersetzt“ werden. Welche Indikatoren sind für die Operationalisierung geeignet? Können die festgelegten Indikatoren auch empirisch beobachtet werden? Bereits bei der Operationalisierung sind die spezifischen Herausforderungen der Beobachtungsmethode zu berücksichtigen.

Entwicklung eines Kodierschemas Bei der Entwicklung eines geeigneten Beobachtungsprotokolls unterscheiden Schnell et al. (2013, S. 384; ähnlich auch Kromrey et al. 2016, S. 326–327) zwischen dem rationalen und dem empirischen Ansatz. Bei der rationalen Vorgehensweise werden relevante Kategorien auf Basis theoretischer Überlegungen konzipiert. Beim empirischen Ansatz werden Kategorien auf Grundlage von bereits zuvor durchgeführten Beobachtungen konstruiert.

In Anlehnung an Westle und Krumm (2009, S. 265; ähnlich auch Schnell et al. 2013, S. 383) lassen sich drei Grundformen von Beobachtungsschemata unterscheiden: Erstens Zählschemata, bei denen lediglich das Auftreten bestimmter Ereignisse verzeichnet wird, zweitens Kategorienschemata, bei denen die auftretenden Ereignisse nach weiteren Kriterien klassifiziert werden, und drittens Schätzschemata, bei denen zusätzlich die Eigenschaften der auftretenden Ereignisse einzuschätzen sind (z. B. Lautstärke bei Wortbeiträgen).

Westle und Krumm (2009, S. 265) nennen zudem fünf Minimalanforderungen an ein Beobachtungsschema: Erstens müssen alle relevanten Sachverhalte in den Kategorien verzeichnet werden können (Vollständigkeit). Zweitens darf jeder Sachverhalt nur einer Kategorie zuzuordnen sein (Ausschließlichkeit der Kategorien). Drittens dürfen die Kategorien nicht gleichzeitig mehrere Aspekte ansprechen (Eindimensionalität). Viertens müssen die beobachtbaren Sachverhalte eindeutig einer Kategorie zugeordnet werden können (Konkretisierung). Fünftens sollte das Schema ausreichend differenzieren, aber den Beobachter auch nicht überfordern (Sensibilität).

Auswahl der Untersuchungsobjekte Auch bei einer Beobachtungsstudie ist die Auswahl der Untersuchungsobjekte festzulegen und zu begründen. Falls keine Zufallsstichprobe realisiert werden kann, sind theoretische Argumente für die Stichprobenziehung anzuführen. Die Auswahl der 64 beobachteten Bundestagsabgeordneten der CITREP-Studie wird wie folgt begründet:

> *„Insgesamt sind über ein Zehntel (elf Prozent) der Abgeordneten des 17. Deutschen Bundestages Teil der Stichprobe. Sie kam nicht durch ein Zufallsverfahren zustande, sondern mittels eines theoretischen Sampling, das 19 Kriterien der Person und des Wahlkreises berücksichtigte, die aus theoretischer Sicht das Repräsentationsverhalten beeinflussen können und vorab festgelegt wurden. Dazu gehören für die Abgeordneten das Alter, die Seniorität, die Parteizugehörigkeit und der Mandatstyp (direkt oder über Liste gewählt) und für die Wahlkreise der Urbanisierungsgrad, die Arbeitslosenquote und die Stärke der Partei vor Ort („Stimmenhochburg“ oder „Diaspora“). Die Verteilung der Auswahlkriterien in der Stichprobe entspricht jener im Deutschen Bundestag der 17. Wahlperiode. So sind auf Basis dieser systematischen Auswahl zwar keine inferenzstatistisch abgesicherten Ge-*

neralisierungen von der Stichprobe auf die Grundgesamtheit aller Abgeordneten möglich, doch liefern die erhobenen Daten Befunde, die wohl für die allermeisten Abgeordneten im Bundestag zutreffen.“ (Siefken 2013, S. 488)

Pretest Vor der Datenerhebung ist das Erhebungsinstrument einem Pretest zu unterziehen. Dabei sind unter anderem die Eindeutigkeit der Zuordnungen der Beobachtungen zu den vorgesehenen Kategorien, die Vollständigkeit der Kategorien und die Handhabbarkeit des Schemas für die Beobachterin zu prüfen. Auch eine Beobachterschulung wird empfohlen (Diekmann 2011, S. 571).

Haupterhebung Mit dem entwickelten und in einem Pretest überprüften Kodierschema wird die Haupterhebung durchgeführt. Wünschenswert, aber in der Forschungspraxis häufig nicht realisierbar, ist der Einsatz mehrerer Beobachter je Untersuchungsobjekt. Dies fördert die Reliabilität der Daten.

Datenaufbereitung und Auswertung Die einzelnen Beobachtungsprotokolle werden aufbereitet und die Daten in einer Datenmatrix organisiert. Dabei werden typischerweise die Untersuchungsobjekte in Zeilen und die Kategorien in Spalten aufgelistet. Daran schließt sich die Datenauswertung an.

11.3 Inhaltsanalyse

In der Methodenliteratur wird die Inhaltsanalyse traditionell als ein Datenerhebungsverfahren eingeordnet (z. B. Diekmann 2011, S. 576; Schnell et al. 2013, S. 397; Häder 2015, S. 327). Diese Charakterisierung ist nach Kuckartz (2009, S. 335) auf die Anfänge der Inhaltsanalyse zurückzuführen. Das Verfahren entstand im Kontext der amerikanischen Massenkommunikationsforschung der 1940er Jahre, bei der vornehmlich Medienprodukte untersucht wurden (Kuckartz 2009, S. 335). Erste inhaltsanalytische Auswertungen wurden allerdings schon Mitte des 18. Jahrhunderts vorgelegt (siehe Früh 2017 für eine knappe historische Darstellung dieser Methode). Kuckartz (2009, S. 335) bezeichnet die Inhaltsanalyse als eine sozialwissenschaftliche Methode, „die keineswegs nur als ein Daten*erhebungsverfahren* gilt, sondern ebenso als ein Verfahren der Daten*analyse*“ (Hervorhebungen im Original).

Inhalt und formale Aspekte Im Vergleich zu einer Befragung werden bei einer Inhaltsanalyse keine neuen Daten generiert, sondern existierende Daten dienen als Grundlage für eine inhaltsanalytische Auswertung. Entsprechend befasst sich die Inhaltsanalyse mit der systematischen Erhebung und Auswertung von Texten,

Bildern und Filmen (z. B. Diekmann 2011, S. 576). Die Inhaltsanalyse muss sich dabei nicht auf den Inhalt von Texten beschränken, sondern kann auch formale Aspekte in den Blick nehmen. In Abhängigkeit von der Forschungsfrage können die Textlänge, die Verwendung von Fremdwörtern oder auch bestimmte Stilmittel (z. B. Konjunktiv) interessante Einblicke ermöglichen. Brettschneider et al. (2009) untersuchen beispielsweise Wahlprogramme auf ihre Verständlichkeit. In der klassischen Studie „Die Arbeitslosen von Marienthal" (Jahoda et al. 1975 [1933], S. 75–76) werden die in Aufsätzen geäußerten Weihnachtswünsche von Kindern ausgewertet. Im Vergleich zu Kindern der Nachbarorte wünschten sich die Mädchen und Jungen in Marienthal nicht nur günstigere Weihnachtsgeschenke, fast ein Drittel dieser Aufsätze ist zudem im Konjunktiv verfasst.

- **Definition: Inhaltsanalyse**
  Früh (2017, S. 29) definiert Inhaltsanalyse als „eine empirische Methode zur systematischen, intersubjektiv nachvollziehbaren Beschreibung inhaltlicher und formaler Merkmale von Mitteilungen, meist mit dem Ziel einer darauf gestützten interpretativen Inferenz auf mitteilungsexterne Sachverhalte".

Vorteile der Inhaltsanalyse Bei einer Inhaltsanalyse werden Texte aller Art untersucht. Diekmann (2011, S. 586) nennt drei Vorteile von Inhaltsanalysen: Vergangenheitsbezug, sozialer Wandel und Nichtreaktivität.

Mit Vergangenheitsbezug ist gemeint, dass grundsätzlich auch in der Vergangenheit produziertes Material inhaltsanalytisch ausgewertet werden kann. Schriftliche Dokumente sind häufig für sehr lange Zeiträume verfügbar, sodass beispielsweise die Berichterstattung zur RAF auch heute inhaltsanalytisch ausgewertet werden kann.

Die zeitliche Verfügbarkeit des Textmaterials ermöglicht die Analyse von (sozialen) Veränderungen. Mit der Inhaltsanalyse können Veränderungen über längere Zeiträume untersucht werden (sozialer Wandel).

Bei einer Inhaltsanalyse wird (in der Regel) auf nichtreaktives Datenmaterial zurückgegriffen. Die Produzenten der Dokumente sind sich in der Regel nicht bewusst, dass die erstellten Dokumente Gegenstand einer wissenschaftlichen Untersuchung sind bzw. werden können. Eine Verfälschung durch soziale Erwünschtheit oder eventuelle Erinnerungsfehler ist deshalb auszuschließen.

11.3.1 Inhaltsanalyse in den Sozialwissenschaften

Für die Durchführung von Inhaltsanalysen gibt es vielfältige Anwendungsmöglichkeiten. An dieser Stelle sollen exemplarisch einige sozialwissenschaftliche Studien vorgestellt werden (siehe für weitere Beispiele z. B. Häder 2015, S. 328–330).

Analyse von Todesanzeigen In einer explorativ angelegten Studie hat Schmied (2002, S. 211–212) die Todesanzeigen in einer südhessischen Kleinstadt analysiert. Er unterscheidet zwischen öffentlichen und privaten Beisetzungen. Bei privaten Beisetzungen („in aller Stille") erscheint die Todesanzeige erst nach erfolgter Beisetzung. Der Anteil dieser privaten Beisetzungen ist im Zeitverlauf (von 1970 bis 2000) von 0,01 über 15 auf 24 Prozent gestiegen. Diesen Befund interpretiert Schmied als Hinweis auf eine zunehmende Familienbezogenheit bei Trauerfällen.

Medien und Wahlkampf Viele Menschen erfahren aktuelle Ereignisse und Problemlagen ausschließlich über die Berichterstattung in Zeitung, Fernsehen, Radio und Internet. Massenmedien prägen damit sehr stark das Bild, das „sich die Bürger vom politischen Geschehen machen" (Niedermayer 2005, S. 161). Mittlerweile liegen umfangreiche Medieninhaltsanalysen für nationale (z. B. Holtz-Bacha 2010; Reinisch 2017) und europäische (z. B. Brettschneider und Rettich 2005; Schuck et al. 2011) Wahlkämpfe vor.

Politik bei Bibi Blocksberg Medien gehören zu den zentralen Instanzen der politischen Sozialisation (Rippl 2008). Der Schwerpunkt der wissenschaftlichen Forschung liegt zwar seit Jahrzehnten auf der politischen Sozialisation Jugendlicher (z. B. Roller et al. 2006; Tenscher und Scherer 2012; siehe auch die Shell- und die DJI-Jugendstudien), aber die jüngere Forschung hat verstärkt Arbeiten zur politischen Sozialisation von Kindern vorgelegt (van Deth et al. 2007; Tausendpfund 2008; Abendschön 2013; Götzmann 2015). In diesem Kontext ist die Studie von Strohmeier (2005) einzuordnen, der die Darstellung von Politik in den Hörspielen von Benjamin Blümchen und Bibi Blocksberg untersucht. Dabei werden überwiegend Schwarz-Weiß-Schemata vermittelt und Politiker als „lächerliche und inkompetente Figuren" dargestellt (Strohmeier 2005, S. 15). Als politische Inhalte werden zudem nur Umweltfragen thematisiert, andere Probleme werden hingegen als nachrangig und unbedeutend dargestellt (Strohmeier 2005, S. 15). Untersuchungen zum Einfluss der Politikdarstellung bei Benjamin Blümchen und Bibi Blocksberg auf politische Orientierungen im Jugend- und Erwachsenenalter stehen allerdings noch aus.

Sichtbare Demokratie In einer Studie der Bertelsmann-Stiftung wird die mediale Präsenz von Bundestagsdebatten mittels einer quantitativen Inhaltsanalyse in deutschen Leitmedien zu zwei Zeitpunkten verglichen (2005/06 und 2013/14). Die Ergebnisse sind ernüchternd: Die Präsenz ist gering und hat über die Zeit abgenommen. Über die DAX30-Unternehmen wird häufiger berichtet als über die Debatten des Deutschen Bundestages (Hierlemann und Sieberer 2014, S. 38–43).

Wahlprogramme In Wahlprogrammen nehmen politische Parteien zu Themen Stellung und formulieren ihre Zielsetzungen. Zwar werden Wahlprogramme von den Bürgern vermutlich kaum gelesen (Rölle 2002), aber durch Politiker und Medien werden die Inhalte der Wahlprogramme zu den Menschen transportiert (Binder und Wüst 2004, S. 38). Im Rahmen von sogenannten Manifesto-Projekten werden Wahlprogramme inhaltsanalytisch ausgewertet. Das bekannteste Manifesto-Projekt ist am Wissenschaftszentrum Berlin (WZB) angesiedelt und wertet Wahlprogramme politischer Parteien in mehr als 50 Ländern für alle freien demokratischen Wahlen seit 1945 aus. Die Wahlprogramme der Parteien bei Europawahlen werden im Rahmen des Euromanifesto-Projekts gesammelt und ausgewertet. Eine jüngere Inhaltsanalyse der Politikangebote der im Deutschen Bundestag vertretenen Parteien im Zeitraum von 1976 bis 2009 haben Pappi et al. (2011) vorgelegt. Die Wahlprogramme auf Landesebene nehmen Seher et al. (2011) sowie Bräuninger und Debus (2012) in den Blick. Eine praktische Anleitung zur Messung der innerparteilichen Demokratie auf Basis von Parteistatuten haben von dem Berge et al. (2013) vorgelegt.

11.3.2 Varianten der Inhaltsanalyse

Vier Varianten der (quantitativen) Inhaltsanalyse werden unterschieden: Frequenzanalysen, Valenzanalysen, Intensivitätsanalysen und Kontingenzanalysen (z. B. Schnell et al. 2013, S. 398–399; Kromrey et al. 2016, S. 322–324).

Frequenzanalyse Frequenzanalysen stellen die einfachste Variante einer Inhaltsanalyse dar. Textelemente werden klassifiziert und die Häufigkeit ihres Vorkommens wird erfasst. Im Vorfeld einer Bürgermeisterwahl könnte beispielsweise die Häufigkeit der Nennung der Kandidaten in den lokalen Tageszeitungen gezählt werden. Dadurch könnte einerseits untersucht werden, ob Amtsinhaber vor einer Wahl häufiger genannt werden als Herausforderer, und andererseits, ob die Häufigkeit der Nennung zwischen verschiedenen Tageszeitungen variiert.

Valenzanalyse Bei Valenzanalysen werden die Bewertungen (negativ, neutral, positiv), die im Zusammenhang mit der Nennung von interessanten Begriffen stehen, erfasst. So könnte im genannten Beispiel nicht nur die Häufigkeit der Nennung gezählt, sondern es könnten auch mögliche Bewertungen erfasst werden.

Intensitätsanalyse Die Intensitätsanalyse stellt eine Erweiterung der Valenzanalyse dar. Dabei wird auch die Intensität von Bewertungen erfasst. Wie stark positiv oder negativ werden die Kandidaten einer Bürgermeisterwahl in der Berichterstattung bewertet?

Kontingenzanalyse Bei Kontingenzanalysen wird das Auftreten bestimmter sprachlicher Elemente im Zusammenhang mit anderen Begriffen untersucht. Im genannten Beispiel könnte beispielsweise untersucht werden, ob die einzelnen Bürgermeisterkandidaten mit anderen Kategorien (z. B. Parteien) erscheinen. Kontingenzanalysen untersuchen, ob relevante Begriffe in einem Text besonders häufig gemeinsam auftreten.

11.3.3 Phasen einer Inhaltsanalyse

Bei einer Inhaltsanalyse lassen sich verschiedene Phasen unterscheiden, die allerdings in der Literatur nicht einheitlich bezeichnet oder abgegrenzt werden (Behnke et al. 2010, S. 293–296; Diekmann 2011, S. 586–596; Schnell et al. 2013, S. 400–404; Kromrey et al. 2016, S. 317–322). Früh (2017, S. 141–200) unterscheidet folgende Phasen einer Inhaltsanalyse: Planungsphase, Entwicklungsphase, Testphase, Anwendungsphase und Auswertungsphase (ähnlich auch Rössler 2010; Krippendorff 2013). Diese einzelnen Phasen werden an einem fiktiven Beispiel illustriert: die Berichterstattung der Medien bei Wahlen.

Planungsphase Die Planungsphase umfasst zwei Schritte: Erstens muss die Fragestellung konkretisiert und zweitens die Grundgesamtheit und ggf. die Stichprobe festgelegt werden.

Zu Beginn jeder Inhaltsanalyse muss das Untersuchungsziel präzisiert werden. Das oben genannte Beispiel „Berichterstattung der Medien bei Wahlen“ lässt sich in dieser allgemeinen Form inhaltsanalytisch nicht bearbeiten. In einem ersten Schritt müssten zumindest die Form der Medien (z. B. Zeitung, Fernsehen oder Radio) und die Wahlebene (z. B. Bundestagswahl, Landtagswahl oder Kommunalwahl) präzisiert werden. Das Thema „Berichterstattung der Tageszeitungen bei Bundestagswahlen“ ist deutlich präziser als die allgemeine Formulierung. Auf

Basis dieser inhaltlichen Präzisierung kann der existierende Forschungsstand aufgearbeitet und erste Hypothesen können entwickelt werden.

Im nächsten Schritt muss die Grundgesamtheit der Elemente bestimmt werden, die als Merkmalsträger zu berücksichtigen sind (Behnke et al. 2010, S. 293). Bei einer Analyse der Tageszeitungen ist festzulegen, welche Zeitungen überhaupt Gegenstand der Erhebung sein sollten. Bei einer Betrachtung der Berichterstattung bei Bundestagswahlen können vermutlich nicht alle deutschen Tageszeitungen berücksichtigt werden. Häufig beschränkt sich die Auswahl auf überregionale Zeitungen; dabei kann grob zwischen Qualitätszeitungen und Boulevardpresse differenziert werden. Schließlich muss der Erhebungszeitraum festgelegt werden. Je nach Größe der Grundgesamtheit kann eine Vollerhebung oder eine Stichprobe angestrebt werden.

Entwicklungsphase In der Entwicklungsphase wird das Kategoriensystem erstellt. Die Kategorien bzw. das Kategoriensystem (auch Kategorienschema genannt) sind das entscheidende Instrument einer Inhaltsanalyse (z. B. Schnell et al. 2013, S. 401; Kromrey et al. 2016, S. 315–317). Behnke et al. (2010, S. 294) beschreibt ein Kategoriensystem „als eine Liste von Merkmalen oder Eigenschaften, die die Erhebungseinheiten aufweisen müssen". Bei der Entwicklung eines Kategoriensystems für die Auswertung der Berichte der Tageszeitungen bei Bundestagswahlen wird festgelegt, welche Informationen über die einzelnen Berichte festgehalten werden. Dies sind sicherlich zunächst einmal allgemeine Informationen (z. B. Zeitung, Erscheinungstermin, Art des Beitrags), aber auch inhaltliche Merkmale (z. B. Nennung von Personen, Anlass der Berichterstattung, Zuordnung zu einem Politikfeld, Wertungen). Für jede Kategorie wird ein Code festgelegt.

Kuckartz (2009, S. 340) unterscheidet zwischen deduktiver und induktiver Kategorienbildung. Bei einer deduktiven Kategorienbildung werden auf Grundlage theoretischer Überlegungen und/oder vorliegender empirischer Befunde Kategorien abgeleitet. Die Kategorien werden so gebildet, dass mit ihnen Hypothesen überprüft werden können, ohne zuvor das Material zu berücksichtigen. Bei einer induktiven Kategorienbildung werden die Kategorien auf Grundlage des vorliegenden Materials entwickelt.

Vier formale Anforderungen an ein Kategorienschema nennen Behnke et al. (2010, S. 295; siehe auch Schnell et al. 2013, S. 401): Erstens sollte ein Kategorienschema eindimensional sein. Jede Kategorie sollte nur eine einzige Bedeutungsdimension abbilden. Zweitens sollte ein Kategoriensystem vollständig sein. Jede Untersuchungseinheit (hier: Zeitungsbericht) muss hinsichtlich der untersuchten Dimension einer Kategorie des Kategorienschemas zugeordnet werden können. Drittens müssen die Kategorien überschneidungsfrei sein. Jedes Merkmal eines

Textes muss eindeutig einer Kategorie zugeordnet werden können. Viertens sollten die Kategorien unabhängig sein.

Testphase Die Testphase umfasst bei Früh (2017, S. 155–189) die Probecodierung, die Codierschulung, die Reliabilitäts- und die Validitätsprüfung.

Bei einer Probecodierung wird das entwickelte Kategorienschema getestet. Die beteiligten Forscher und Mitarbeiter (Codierer) wenden das vorläufige Kategoriensystem gemeinsam an.

> *„Jede Zuordnung wird besprochen, und die Codierentscheidung durch expliziten Bezug auf die bisher vorhandenen Kategoriendefinitionen begründet. Differierende Zuordnungen oder nicht mögliche Zuordnungen werden durch Präzisierung der Definition oder die Hinzunahme neuer Kategorien gelöst. Abgrenzungen und Vereinbarungen werden sofort notiert. Es können auch Unterkategorien ausgetauscht, zusammengefasst oder ganz gestrichen werden." (Früh 2017, S. 156)*

Nach der Probecodierung wird das Kategorienschema überarbeitet. Die anschließende Codierschulung zielt darauf ab, die Codierer mit dem Kategorienschema und den im Codebuch formulierten Regeln vertraut zu machen. Häufig arbeiten die Codierer zunächst mit einem gemeinsamen Text, der anschließend besprochen wird.

Im Anschluss an die Codierschulung folgt ein Reliabilitätstest. Zwei Formen der Reliabilität werden unterschieden: die Intercoder-Reliabilität und die Intracoder-Reliabilität (z. B. Früh 2017, S. 179–187; Schnell et al. 2013, S. 403). Bei der Intercoder-Reliabilität wird geprüft, ob der gleiche Text von verschiedenen Codierern einheitlich vercodet wird. Unterschiedliche Codierungen können verschiedene Ursachen haben (z. B. mehrdeutige Codieranweisungen, unzuverlässige Codierer). Bei langer Projektdauer ist auch die Intracoder-Reliabilität zu bestimmen. Die Intracoder-Reliabilität ist gewährleistet, wenn der gleiche Text von einer Person zu verschiedenen Zeitpunkten gleich vercodet wird. Für die Berechnung der Reliabilität existieren verschiedene Reliabilitätsmaße, beispielsweise Holsti CR, Scott Pi oder auch Krippendorfs Kappa (für eine Übersicht siehe z. B. Früh 2017, S. 179–187; siehe auch Krippendorff 2013, S. 267–328).

Zum Abschluss der Testphase ist schließlich auch die Validität der Codierungen in den Blick zu nehmen (Krippendorff 2013, S. 329–353). Allgemein ist mit Validität die Gültigkeit eines Messinstruments gemeint. Ein Messinstrument ist valide (gültig), wenn es tatsächlich das misst, was es messen soll (siehe Kapitel 7). Das generelle Problem der gültigen Übersetzung theoretischer Konzepte in empirische Indikatoren existiert selbstverständlich auch bei der Inhaltsanalyse (Kromrey et al. 2016, S. 322). Deshalb betrachtet Früh (2017, S. 187–189) die Übereinstimmung der

Codierung zwischen Forscherin und Codierer als einen begrenzten Hinweis der Validität. Eine hohe Übereinstimmung der Codierung zwischen Forscherin und Codierer deutet darauf hin, dass die von der Forscherin als valide gesetzten Operationalisierungen auch von den Codierern angemessen umgesetzt wurden.

Anwendungsphase Auf Grundlage des entwickelten Kategorienschemas können die Inhalte der Kategorien in den Texten gesucht und verschlüsselt (codiert) werden. Dabei sind die Texte zufällig auf die verschiedenen Codierer zu verteilen.

Auswertungsphase Die Daten einer Inhaltsanalyse werden aufbereitet und in einem Rohdatensatz gespeichert. In unserem Beispiel entspricht jeder codierte Bericht einem Fall im Datensatz. Die einzelnen Kategorien sind die Variablen. Die quantitative Auswertung erfolgt dann mit einem Statistikprogramm (Früh 2017, S. 191–193). Eine ausführliche Darstellung der Inhaltsanalyse mit vielen praktischen Hinweisen bietet auch Krippendorff (2013).

Software für Inhaltsanalyse

Für alle Formen der Inhaltsanalyse gibt es mittlerweile Softwareprogramme. In der quantitativen Inhaltsanalyse werden Programme wie Textpack, Wordfish und Wordscore genutzt, die kostenlos verfügbar sind. Qualitative Forscher nutzen Programme wie ATLAS.ti und MAXQDA. Weitere Software für die computergestützte Inhaltsanalyse (sogenannte CUI-Programme) lassen sich im Internet mit dem Suchbegriff „software for content analysis“ schnell recherchieren.

12 Datenaufbereitung und -analyse

Bei der Datenerhebung werden von den Untersuchungsobjekten (z. B. Personen, Gemeinden, Staaten) die interessierenden Merkmale (z. B. politisches Interesse, Demokratiezufriedenheit, Nettoeinkommen) erfasst. Diese Merkmale werden allgemein als Variablen bezeichnet. Die quantitative Datenanalyse setzt neben einer sorgfältigen Datenerhebung auch eine bestimmte Datenorganisation voraus, die Datenmatrix genannt wird. Nach der Erstellung der Datenmatrix werden die Daten für die Analyse vorbereitet. Diese Phase wird Datenaufbereitung genannt und kann je nach Zahl der Untersuchungsobjekte Tage, Wochen oder Monate dauern. Bei der anschließenden Datenanalyse werden die formulierten Hypothesen empirisch geprüft.

12.1 Datenaufbereitung

Für eine quantitative Datenanalyse werden die Daten in einer Datenmatrix organisiert. Eine Datenmatrix entspricht einer Tabelle, die alle erhobenen Merkmale für alle Untersuchungsobjekte enthält. In einer solchen Datenmatrix werden in der Regel die Untersuchungsobjekte (z. B. Personen) in Zeilen und deren Merkmale (z. B. Geschlecht, Alter) in Spalten aufgelistet. In den einzelnen Zellen finden sich die Merkmalsausprägungen des jeweiligen Untersuchungsobjekts.

In Tabelle 16 ist eine Beispiel-Datenmatrix dargestellt. In der linken Hälfte wird exemplarisch der Inhalt der einzelnen Zellen beschrieben und in der rechten Hälfte sind Ziffern eingetragen. Jede Zeile entspricht einem Befragten. In der Spalte „Variable 1“ sind die Werte (inhaltliche Information) der einzelnen Untersuchungsobjekte für die Variable 1 eingetragen. Die Zahlen haben jeweils eine inhaltliche Bedeutung, die sich allerdings nur aus dem konkreten Forschungs-

© Springer Fachmedien Wiesbaden GmbH, ein Teil von Springer Nature 2018

M. Tausendpfund, *Quantitative Methoden in der Politikwissenschaft*,
Grundwissen Politik, https://doi.org/10.1007/978-3-658-20698-7_12

Tabelle 16 Datenmatrix

	Variable 1	Variable 2
Fall 1	Wert von Fall 1 auf Variable 1	Wert von Fall 1 auf Variable 2
Fall 2	Wert von Fall 2 auf Variable 1	Wert von Fall 2 auf Variable 2
Fall 3	Wert von Fall 3 auf Variable 1	Wert von Fall 3 auf Variable 2
Fall 4	Wert von Fall 4 auf Variable 1	Wert von Fall 4 auf Variable 2

	v1	v2
1	1	2
2	2	4
3	2	5
4	1	99

kontext erschließt. Die Variable v1 könnte beispielsweise das Geschlecht der Befragten erfassen. Die Ausprägung 1 könnte dann für „weiblich" und die Ausprägung 2 für „männlich" stehen.

Die inhaltlichen Informationen der Variablen sind in der Regel in einem Codebuch dokumentiert. Ein solches Codebuch enthält eine Variablenübersicht, die inhaltliche Beschreibung der Variablen sowie eine Erläuterung der Ziffern. Solche Codebücher werden auch als Variablenreports bezeichnet und können mit dem Datensatz heruntergeladen werden.

Prinzipien der Datensammlung Mit Kromrey et al. (2016, S. 217–218) lassen sich drei Prinzipien der Datensammlung unterscheiden, die unmittelbar aus einer Datenmatrix ersichtlich sind: das Prinzip der Vergleichbarkeit, das Prinzip der Klassifizierbarkeit und das Prinzip der Vollständigkeit.

Das Prinzip der Vergleichbarkeit bedeutet, dass das Merkmal einer Variable für alle Untersuchungsobjekte identisch ist. Die Variable v1 erfasst beispielsweise das Geschlecht und die Variable v2 das politische Interesse der Befragten. Diese inhaltliche Bedeutung gilt für alle Befragten im Datensatz.

Das Prinzip der Klassifizierbarkeit besagt, dass für jedes Untersuchungsobjekt (z. B. Personen) und jedes Merkmal (z. B. politisches Interesse) genau ein Wert existieren muss. Ein Befragter darf auf einer Variablen nicht mehrere Werte aufweisen.

Das Prinzip der Vollständigkeit bedeutet, dass in einer Datenmatrix keine Zellen leer bleiben dürfen. Für jedes Untersuchungsobjekt muss für jede Variable eine empirische Information vorliegen. Bei einer Datenerhebung (z. B. Befragung) werden allerdings immer Situationen auftreten, bei denen ein Befragter die Frage nicht beantworten kann („weiß nicht") oder nicht beantworten will (Item-Non-

response).[48] Häufig deuten Ziffern wie 9 oder 99 auf „keine Angabe“ oder „weiß nicht“ hin. Bei der Datenanalyse können solche Angaben häufig nicht berücksichtigt werden, weshalb sie auch als fehlende Werte (missing values) bezeichnet werden. Der Befragte hat zwar einen formalen Wert erhalten (z. B. 99 für „keine Angabe“), dieser Wert enthält allerdings keine „gültige“ Information. In Tabelle 16 deutet die „99“ bei Variable v2 beim vierten Untersuchungsobjekt auf einen solchen missing value hin. Bei der statistischen Analyse werden solche missing values bei den Berechnungen (z. B. Mittelwert) in der Regel ausgeschlossen.

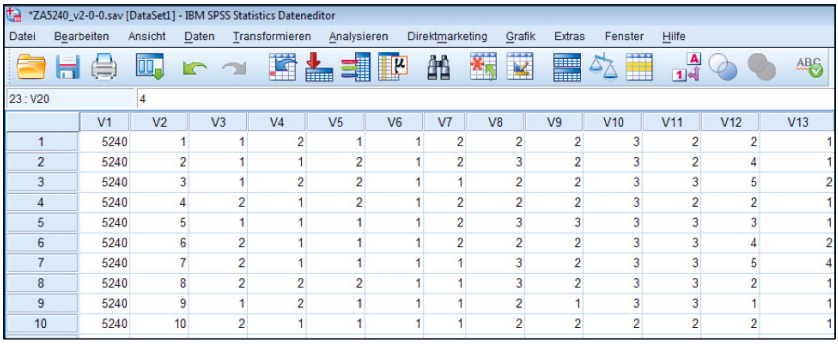
GESIS-Datenbestandskatalog

Der Datenbestandskatalog von GESIS – Leibniz-Institut für Sozialwissenschaften enthält neben den Datensätzen auch Studienbeschreibungen zu allen im Datenarchiv langfristig gesicherten Studien (z. B. ALLBUS, GLES). Dort können Studien recherchiert und Dokumente der Studien (z. B. Fragebögen und Codebücher) sowie die Datensätze heruntergeladen werden

Datenansicht in SPSS In Abbildung 33 findet sich ausschnittsweise die SPSS-Datenansicht der ALLBUS 2014. In V1 ist die Studiennummer (5240) eingetragen, die für alle Befragten identisch ist. Diese Studiennummer wird von GESIS – Leibniz-Institut für Sozialwissenschaften vergeben. Jeder bei GESIS archivierte Datensatz hat eine individuelle Studiennummer. Die Variable V2 ist die individuelle Identifikationsnummer der Befragten. In Variable V7 findet sich beispielsweise die Information zum Erhebungsgebiet. 1 steht für alte Bundesländer und 2 für neue Bundesländer. Offensichtlich wurden die Interviews der ersten zwei Befragten in den neuen Bundesländern und das Interview des dritten Befragten in den alten Bundesländern durchgeführt. Insgesamt liegen Informationen von 3471 Personen (Zeilen) zu 861 Variablen (Spalten) vor.

48 Die Differenzierung zwischen „weiß nicht“ und „keine Angabe“ ist inhaltlich bedeutsam. Während sich ein Befragter bei „weiß nicht“ aus kognitiven Gründen nicht in der Lage sieht, die gestellte Frage zu beantworten, lässt „keine Angabe“ auf eine explizite Verweigerung schließen (z. B. Gabriel und Thaidigsmann 2009). Nach Borg und Treder (2003) geben Befragte vor allem zu solchen Bereichen keine Auskunft („weiß nicht“), in denen sie sich weniger gut auskennen. Explizite Antwortverweigerungen („keine Angabe“) sind dagegen besonders bei sensiblen Fragen zu beobachten (z. B. Einkommen).

Abbildung 33 SPSS-Datenansicht der ALLBUS 2014



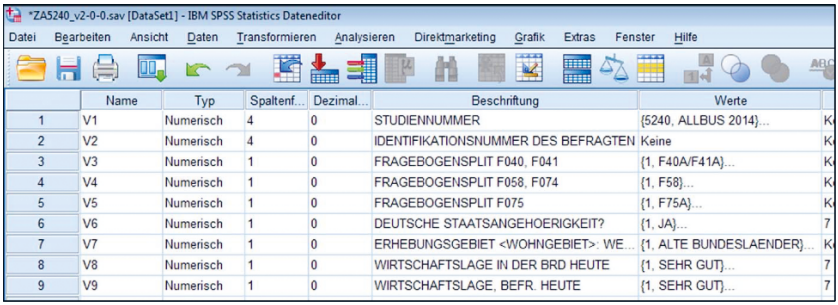
	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12	V13
1	5240	1	1	2	1	1	2	2	2	3	2	2	1
2	5240	2	1	1	2	1	2	3	2	3	2	4	1
3	5240	3	1	2	2	1	1	2	2	3	3	5	2
4	5240	4	2	1	2	1	2	2	2	3	2	2	1
5	5240	5	1	1	1	1	2	3	3	3	3	3	1
6	5240	6	2	1	1	1	2	2	2	3	3	4	2
7	5240	7	2	1	1	1	1	3	2	3	3	5	4
8	5240	8	2	2	2	1	1	3	2	3	3	2	1
9	5240	9	1	2	1	1	1	2	1	3	3	1	1
10	5240	10	2	1	1	1	1	2	2	2	2	2	1

Quelle: Eigene Darstellung

Die SPSS-Datenansicht ist ohne weitere Angaben nicht sonderlich informativ. Ohne Kenntnis der inhaltlichen Bedeutung der einzelnen Variablen handelt es sich bei der Datenansicht um eine kryptische Präsentation von Zahlen.

Variablenansicht in SPSS Informationen zur inhaltlichen Bedeutung der einzelnen Variablen finden sich zum einen im Codebuch, das meist als gesondertes Dokument zum Datensatz verfügbar ist, und zum anderen im Datensatz selbst. Die SPSS-Variablenansicht (Abbildung 34) enthält Basisinformationen zu den einzelnen Variablen. Den einzelnen Zeilen der SPSS-Variablenansicht sind Angaben zum Inhalt der Variablen zu entnehmen. Besonders interessant ist dabei die Spalte „Beschriftung", die inhaltliche Informationen zur Variable enthält. Die Varia-

Abbildung 34 SPSS-Variablenansicht der ALLBUS 2014



	Name	Typ	Spaltenf...	Dezimal...	Beschriftung	Werte	
1	V1	Numerisch	4	0	STUDIENNUMMER	{5240, ALLBUS 2014}...	K
2	V2	Numerisch	4	0	IDENTIFIKATIONSNUMMER DES BEFRAGTEN	Keine	K
3	V3	Numerisch	1	0	FRAGEBOGENSPLIT F040, F041	{1, F40A/F41A}...	K
4	V4	Numerisch	1	0	FRAGEBOGENSPLIT F058, F074	{1, F58}...	K
5	V5	Numerisch	1	0	FRAGEBOGENSPLIT F075	{1, F75A}...	K
6	V6	Numerisch	1	0	DEUTSCHE STAATSANGEHOERIGKEIT?	{1, JA}...	7
7	V7	Numerisch	1	0	ERHEBUNGSGEBIET <WOHNGEBIET>: WE...	{1, ALTE BUNDESLAENDER}...	K
8	V8	Numerisch	1	0	WIRTSCHAFTSLAGE IN DER BRD HEUTE	{1, SEHR GUT}...	7
9	V9	Numerisch	1	0	WIRTSCHAFTSLAGE, BEFR. HEUTE	{1, SEHR GUT}...	7

Quelle: Eigene Darstellung

ble V1 hat beispielsweise die Beschriftung „Studiennummer“, die Variable V2 die Beschriftung „Identifikationsnummer des Befragten“ und die Variable V7 die Beschriftung „Erhebungsgebiet ‚Wohngebiet‘: West – Ost“. Die Beschriftung einer Variable wird auch Variablenlabel genannt. Das Variablenlabel enthält Angaben, welche inhaltlichen Informationen in der Variable erfasst sind. Für ausführlichere Informationen (z. B. die konkrete Frageformulierung bei einer Befragung) ist das Codebuch erforderlich. Neben der inhaltlichen Bedeutung der einzelnen Variablen sind in der SPSS-Variablenansicht auch Informationen zu den jeweiligen Ausprägungen (Werte) einer Variable nützlich. Diese Beschriftung wird Wertelabel genannt. Bei V7 bedeuten 1 = Alte Bundesländer und 2 = Neue Bundesländer.

Von der Datenmatrix zum Datensatz

Wie kommen die Daten in den Datensatz? Wer mit einschlägigen Sekundärdaten arbeitet (z. B. ALLBUS), muss keine Daten eingeben. Über den GESIS-Datenbestandskatalog stehen „fertige“ Datensätze zur Verfügung. Den Prozess vom Fragenbogen zum Datensatz beschreiben Lück und Baur (2011). Für die Dateneingabe bietet das kostenlose Programm EpiData im Vergleich zu Excel viele Vorteile (www.epidata.dk), da es zahlreiche Kontrollfunktionen unterstützt (z. B. Definition von Wertebereichen).

Die Ziffern in den einzelnen Zellen der SPSS-Datenansicht (siehe Abbildung 33) haben eine konkrete inhaltliche Bedeutung. Ohne die Kenntnis der Variablen- und Wertelabels ist eine Datenaufbereitung und Datenanalyse aber nicht möglich.

Datenkontrolle Insbesondere bei der eigenständigen Dateneingabe, aber auch bei der Arbeit mit Sekundärdaten, sollte vor der eigentlichen Datenanalyse eine sorgfältige Fehlerkontrolle vorgenommen werden. Mit Diekmann (2011, S. 666; siehe auch Lück und Landrock 2014) lassen sich drei typische Fehler unterscheiden, die relativ einfach identifiziert werden können:

1) Werte außerhalb des definierten Wertebereichs (wild code)
2) Unplausible Werte
3) Inkonsistente Werte

wild code Mit folgender Frage werden bei der ALLBUS 2014 die Befragten gebeten, die wirtschaftliche Lage in Deutschland zu beurteilen:

Wie beurteilen Sie ganz allgemein die heutige wirtschaftliche Lage in Deutschland?

- *sehr gut (1)*
- *gut (2)*
- *teils gut/teils schlecht (3)*
- *schlecht (4)*
- *sehr schlecht (5)*
- *weiß nicht (8)*
- *keine Angabe (9)*

In Klammern sind die Werte der einzelnen Antwortmöglichkeiten dargestellt. Im ALLBUS-Datensatz kann ein Befragter nur den Wert 1, 2, 3, 4, 5, 8 oder 9 erhalten. Alle anderen Werte sind fehlerhafte Angaben (wild code) und müssen bei der Datenkontrolle identifiziert und korrigiert werden.

Unplausible Werte Bei der Befragung der Kommunalpolitiker im Rahmen des Projekts „Europa im Kontext" (z. B. Tausendpfund 2013c) wurde den Mandatsträgern eine Frage nach ihrem zeitlichen (ehrenamtlichen) Engagement gestellt:

Wenn nicht gerade Wahlkampf ist, wie viele Stunden wenden Sie in einem durchschnittlichen MONAT für Ihre ehrenamtlichen kommunalpolitischen Aktivitäten auf? Tragen Sie bitte in das Feld die volle Stundenzahl ein.

Ein Befragter hat als Stundenzahl „300" eingetragen. Bei dieser Stundenzahl muss der Kommunalpolitiker, der zudem berufstätig und verheiratet ist, jeden Tag rund zehn Stunden für die ehrenamtliche Tätigkeit aufwenden. Diese Angabe ist zumindest auffällig und sollte bei der Datenkontrolle (soweit möglich) überprüft werden. Vermutlich handelt es sich um einen Eingabefehler (30 statt 300 Stunden).

Inkonsistente Werte In jeder Datenmatrix liegen zahlreiche Informationen für jedes Untersuchungsobjekt vor, bei Personen beispielsweise das Geburtsjahr oder die Wohndauer in einer Gemeinde. Diese Angaben können genutzt werden, um inkonsistente Werte zu entdecken. So kann die Wohndauer in einer Gemeinde oder die Ehedauer nicht größer als das Alter des Befragten sein. Offensichtlich liegen dann inkonsistente Werte vor.

Wild codes, unplausible und inkonsistente Werte können mit einem Statistikprogramm leicht herausgefiltert werden. In vielen Fällen kann der Fehler nachträglich korrigiert werden (Vergleich mit dem Originalfragebogen). Falls eine solche Angabe nicht korrigiert werden kann, dann wird in der Regel ein Missing-Value-Code (z. B. 77 für „unplausible Angabe") zugewiesen.

Bei Lück und Landrock (2014, S. 403–407) finden sich weitere Hinweise, um Fehler in einem Datensatz zu identifizieren. Grundsätzlich sollte jeder Datensatz vor der Datenanalyse sorgfältig geprüft werden. Dies gilt auch für einschlägige Datensätze wie ALLBUS oder ESS. Auch bei solchen qualitativ hochwertigen Datensätzen kann es unplausible und/oder inkonsistente Werte geben. Deshalb werden auch korrigierte Datensätze und Fehlerverzeichnisse (Errata) veröffentlicht. Schließlich muss jede Fehlerkorrektur dokumentiert werden. Exemplarisch finden sich in Abbildung 35 die Errata-Angaben zur ALLBUS 2014.

Abbildung 35 Errata-Angaben zur ALLBUS 2014

Quelle: Eigene Darstellung

12.2 Datenanalyse

Bei der Datenanalyse lassen sich drei Phasen unterscheiden: Die deskriptive Beschreibung der Daten, die empirische Prüfung der Hypothesen und die Durchführung von Signifikanztests. Für diese einzelnen Phasen gibt es zahlreiche Analyseinstrumente; ausgewählte Verfahren werden bei Wolf und Best (2010a) vorgestellt. An dieser Stelle werden die drei Phasen knapp beschrieben

Deskriptive Beschreibung Bei der quantitativen Datenanalyse ist die Forscherin in der Regel mit vielen Untersuchungsobjekten konfrontiert. Ein ALLBUS-Datensatz umfasst mehr als 3000 befragte Personen, der Datensatz der sechsten Welle des European Social Survey enthält Informationen von über 50 000 Befragten.

Sozialwissenschaftliche Datenanalyse

Dieses Kapitel bietet einen Überblick über typische Phasen der sozialwissenschaftlichen (quantitativen) Datenanalyse. Für die Anwendung statistischer Verfahren und die angemessene Interpretation empirischer Resultate ist das Studium weiterführender Literatur erforderlich. Eine gelungene Darstellung der deskriptiven Statistik bietet Benninghaus (2007); weiterführende Verfahren der sozialwissenschaftlichen Datenanalyse (z. B. Regressionsanalyse) werden bei Wolf und Best (2010a), Urban und Mayerl (2011) sowie Kopp und Lois (2014) dargestellt.

Dies können demographische Angaben der Befragten (z. B. Geschlecht, Alter, Einkommen), Informationen zum politischen Interesse und sozialen Vertrauen oder auch zur Mediennutzung und Lebenszufriedenheit sein. Solche Datenmengen überfordern die visuelle Wahrnehmungsfähigkeit der meisten Menschen. Deshalb wird bei der deskriptiven Beschreibung die große Datenmenge auf wichtige Informationen verdichtet. Dabei lassen sich Lage- und Streuungsmaße unterscheiden.

Lagemaße informieren über das „Zentrum" einer Verteilung (z. B. Einkommen der Befragten). Besonders bekannte Lagemaße sind Modus (auch Modalwert genannt), Median und (arithmetischer) Mittelwert. Der Modus ist der häufigste Wert einer Variable in einem Datensatz. Der Median repräsentiert die „Mitte" einer geordneten Variable. 50 Prozent der Werte liegen „vor" und 50 Prozent liegen „nach" dem Median. Die Bezeichnung „vor" und „nach" haben aber nur dann

Warum sind elementare Statistikkenntnisse wichtig?

Welche Rolle spielt das politische Interesse bei der Wahlbeteiligung? Warum unterscheidet sich das Einkommen zwischen Männern und Frauen? Welche Auswirkungen hat eine nicht-eheliche Lebensgemeinschaft auf das spätere Scheidungsrisiko? Welcher sozialwissenschaftlichen Fragestellung auch immer das Interesse gilt, spätestens bei der Lektüre empirischer Studien wird der Leser mit statistischen Verfahren konfrontiert. Wer aber eine empirische Studie verstehen und kritisch bewerten möchte, muss die Prinzipien, Anwendungsvoraussetzungen und Probleme der wichtigsten statistischen Verfahren kennen. Für einen Sozialwissenschaftler sind deshalb elementare Statistikkenntnisse unverzichtbar (z. B. Diekmann 2011, S. 658).

eine sinnvolle Bedeutung, wenn die Daten in eine Reihenfolge gebracht werden können. Deshalb kann der Median nur bei ordinalskalierten Merkmalen angegeben werden. Der (arithmetische) Mittelwert trifft eine Aussage über den „Durchschnitt" des Merkmals. Er wird wie folgt berechnet: Summe der einzelnen Beobachtungswerte dividiert durch die Fallzahl. Die einzelnen Lagemaße haben jeweils Vor- und Nachteile, die bei Gehring und Weins (2009) dargestellt werden. Wichtig ist insbesondere das erforderliche Skalenniveau eines Merkmals. Die Berechnung des Mittelwerts fordert mindestens Intervallskalenniveau, der Modus kann bereits bei nominalskalierten Merkmalen berechnet werden. Beim Median muss, wie oben beschrieben, eine Ordinalskala vorliegen (siehe auch Abschnitt 7.3).

Trotz gleicher Lagemaße (z. B. gleiche Mittelwerte, gleicher Median) können Verteilungen von Merkmalen (z. B. das Einkommen der Befragten) völlig unterschiedlich aussehen. In Tabelle 17 ist das wöchentliche (fiktive) Taschengeld von jeweils elf Kindern in Euro dargestellt. Auf den ersten Blick ist offensichtlich, dass sich die Angaben der Stadt- und Landkinder sehr stark unterscheiden. Bei den

Tabelle 17 Taschengeld von Stadt- und Landkindern (fiktive Daten)

	Stadt	**Land**
Kind 1	10	9
Kind 2	8	9
Kind 3	13	9
Kind 4	9	9
Kind 5	11	9
Kind 6	14	10
Kind 7	6	8
Kind 8	4	8
Kind 9	12	14
Kind 10	9	8
Kind 11	3	6
Modus	9	9
Median	9	9
Mittelwert	9	9
(Empirische) Varianz	11,45	3,45
(Empirische) Standardabweichung	3,38	1,86

Stadtkindern variiert das wöchentliche Taschengeld deutlich stärker als bei den Landkindern. Trotz dieser offensichtlichen Unterschiede sind die Lagemaße allerdings in beiden Datensätzen völlig identisch. Bei einer überschaubaren Anzahl an Untersuchungsobjekten ist diese Variation unmittelbar ersichtlich, bei größeren Datensätzen mit mehreren 100 oder 1000 Befragten ist dies allerdings nicht erkennbar.

Deshalb sind Streuungsmaße erforderlich, die die Variation einer Verteilung beschreiben. Bekannte Streuungsmaße sind die Varianz und die Standardabweichung. Je höher Varianz und Standardabweichung, desto größer die Streuung (Dispersion). In unserem Beispiel (Tabelle 17) ist die Streuung der Daten bei den Stadtkindern deutlich größer als bei den Landkindern.

Hypothesen prüfen Quantitative Sozialforscherinnen wollen Beziehungen zwischen Merkmalen untersuchen. Besteht ein Zusammenhang zwischen Bildung und Einkommen? Fördert politisches Interesse die Wahlbeteiligung? Unterscheiden sich die Mathematikleistungen zwischen Jungen und Mädchen? Für die Untersuchung solcher Zusammenhänge gibt es zahlreiche statistische Verfahren (für eine Übersicht siehe z. B. Benninghaus 2007; Diekmann 2011, S. 703), die bekanntesten sind sicherlich Cramer's V, Spearman's rho und Pearson's r. Welches Zusammenhangsmaß jeweils geeignet ist, hängt vor allem vom Skalenniveau der Merkmale ab. Wie aus Tabelle 18 zu entnehmen, ist jeweils das niedrigste Skalenniveau entscheidend. Bei zwei nominalskalierten Merkmalen bietet sich Cramer's V an,

Tabelle 18 Zusammenhangsmaße in Abhängigkeit des Skalenniveaus

		Skalenniveau von Merkmal A		
		Nominal	**Ordinal**	**Metrisch**
Skalenniveau von Merkmal B	Nominal	Cramer's V	Cramer's V	Cramer's V
	Ordinal	Cramer's V	Spearman's rho	Spearman's rho
	Metrisch	Cramer's V	Spearman's rho	Pearson's r

bei zwei ordinalskalierten Merkmalen ist Spearman's rho das geeignete Zusammenhangsmaß und bei zwei metrischskalierten Merkmalen kann Pearson's r berechnet werden.

In den meisten Fällen wird die abhängige Variable (z. B. Einkommen) nicht nur von einer unabhängigen Variable (z. B. Bildung) abhängig sein, sondern auch von anderen Merkmalen (z. B. Geschlecht und Alter). Die möglichen relevanten

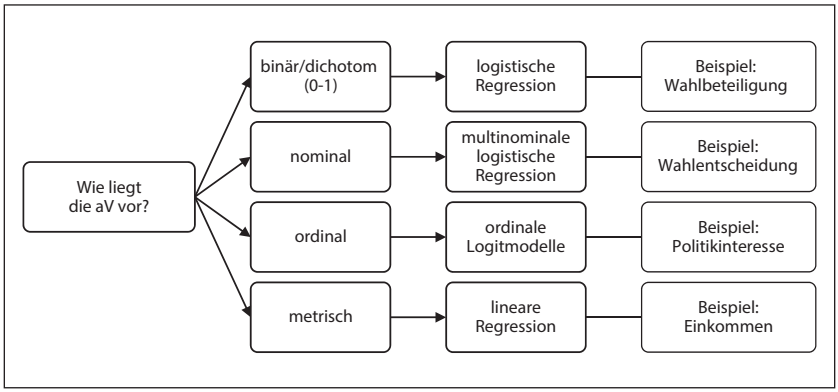
unabhängigen Variablen müssen auf Basis theoretischer Überlegungen und der intensiven Auseinandersetzung mit dem Forschungsstand identifiziert werden. In den Sozialwissenschaften steht unter dem Stichwort „Multivariate Analyseverfahren" eine große Zahl von Analysetechniken zur Verfügung, um den Einfluss mehrerer unabhängiger Variablen (z. B. Bildung, Geschlecht, Alter) auf die abhängige Variable (z. B. Einkommen) schätzen zu können.

Datenanalyse mit einem Statistikprogramm

Bei überschaubaren Datensätzen können statistische Maßzahlen natürlich einfach „per Hand" oder mit einem Taschenrechner berechnet werden, bei Datensätzen mit mehreren 100 Untersuchungsobjekten ist allerdings selbst die Berechnung des Mittelwerts mühsam, zeitaufwändig und fehleranfällig. Diese Routinetätigkeiten können Programme wie Excel oder Statistikprogramme wie SPSS, Stata oder R deutlich schneller und effizienter durchführen. Über die Angemessenheit eines statistischen Verfahrens und die Interpretation der empirischen Befunde kann allerdings kein Statistikprogramm entscheiden. Diese Verantwortung liegt ausschließlich beim Anwender. Deshalb: Wer mit einem Statistikprogramm arbeitet, sollte erstens mit den verwendeten statistischen Verfahren vertraut sein, zweitens die Voraussetzungen der Anwendbarkeit eines statistischen Verfahrens kennen und sich drittens den Grenzen der Interpretation der Resultate bewusst sein.

Das wichtigste multivariate Analyseverfahren in den Sozialwissenschaften ist die Regressionsanalyse. Die Bezeichnung „Regressionsanalyse" ist ein Oberbegriff für eine große Anzahl an regressionsanalytischen Verfahren (z. B. Wolf und Best 2010a). Grundsätzlich wird bei einer Regression der Effekt einer oder mehrerer unabhängiger Variablen auf eine abhängige Variable geschätzt (für eine Einführung in die Regressionsanalyse siehe Urban und Mayerl 2011; Kopp und Lois 2014). Welches Regressionsverfahren für die jeweilige Fragestellung geeignet ist, hängt (unter anderem) vom Skalenniveau der abhängigen Variable ab. In Abbildung 36 sind wichtige regressionsanalytische Verfahren in einem Entscheidungsbaum angeordnet. Ausgangspunkt ist immer die Frage: Wie liegt die abhängige Variable vor? Bei einer binären bzw. dichotomen abhängigen Variable wäre die logistische Regression eine mögliche Analysestrategie. Ein typischer Anwendungsfall ist die Wahlbeteiligung, die nur zwei Ausprägungen hat: „Ja" versus „Nein". Bei einer nominalen abhängigen Variable bieten sich multinominale logistische Regressions-

Abbildung 36 Auswahl von regressionsanalytischen Verfahren



Quelle: Eigene Darstellung

verfahren an. Ein Beispiel ist die Wahl einer Partei, da das Messniveau der aV (z. B. SPD, CDU, Grüne, Linke) nominal ist.

Bei ordinalen Merkmalen (z. B. politisches Interesse) können ordinale Logitmodelle geschätzt werden, bei metrischen Merkmalen (z. B. Einkommen) bietet die lineare Regression eine geeignete Analysestrategie.

Die Regressionsanalyse stellt in den Sozialwissenschaften die dominante Analysestrategie dar. Bei sorgfältiger Anwendung handelt es sich um ein „mächtiges" Instrument der Datenanalyse. Die Anwendung der Regressionsanalyse und die angemessene Interpretation der Ergebnisse ist allerdings an bestimmte Voraussetzungen gebunden. Diese werden in einschlägigen Lehrbüchern der sozialwissenschaftlichen Datenanalyse ausführlich dargestellt (z. B. Wolf und Best 2010a; Urban und Mayerl 2011; Kopp und Lois 2014).

Signifikanztests Aus zeitlichen, finanziellen und forschungspraktischen Gründen dominieren in den Sozialwissenschaften Teilerhebungen. Aus einer Grundgesamtheit werden zufällig Untersuchungsobjekte (z. B. Personen) „gezogen" (siehe Kapitel 10). Diese Stichprobe bildet die Grundlage für die empirische Analyse. Es werden Lage- und Streuungsmaße berechnet sowie Zusammenhänge zwischen den Merkmalen auf Basis der Stichprobendaten überprüft.

Sozialforscherinnen interessieren sich allerdings weniger für die Ergebnisse und Zusammenhänge in der Stichprobe, sondern vielmehr für die Resultate und Beziehungen in der entsprechenden Grundgesamtheit. Wie zufrieden sind die Bürger mit der Leistung der Bundesregierung? Besteht ein Zusammenhang

zwischen der unabhängigen Variable (z. B. Bildung) und der abhängigen Variable (z. B. Einkommen)? Unterscheidet sich das Wahlverhalten (aV) in Abhängigkeit des Geschlechts (uV)?

Bei der ALLBUS stehen die knapp 3500 Befragten stellvertretend für die deutsche Bevölkerung ab 18 Jahren. Die Auswahl der Personen bei der ALLBUS basiert auf einem mehrstufigen Zufallsverfahren (siehe Kapitel 10). Bei solchen Zufallsstichproben ist der Rückschluss von der Stichprobe auf die Grundgesamtheit durch Anwendung der sogenannten Inferenzstatistik möglich. Die Inferenzstatistik stellt „Werkzeuge" bereit, um die Sicherheit der Übertragbarkeit eines Stichprobenergebnisses auf die Grundgesamtheit berechnen zu können. Eine wichtige Frage ist beispielsweise, ob ein in der Stichprobe beobachteter Zusammenhang zwischen zwei Merkmalen (z. B. Bildung und Einkommen) zufällig entstanden sein kann, wenn diese Beziehung in der Grundgesamtheit überhaupt nicht existiert. Verfahren zur Beantwortung dieser Fragen werden Signifikanztests genannt (Schnell et al. 2013, S. 437).

! Bei einem Signifikanztest wird – vereinfacht formuliert – geprüft, ob ein in der Stichprobe gefundener Zusammenhang (sehr) wahrscheinlich auch in der Grundgesamtheit existiert. Häufig werden Signifikanztests falsch interpretiert. Mit Schnell et al. (2013, S. 442) lassen sich drei typische Anwendungsfehler unterscheiden. Ein „signifikantes" Ergebnis bedeutet nicht, dass

- es sich um einen „wichtigen" Effekt handelt.
- die „Existenz" eines Effekts bewiesen ist.
- es sich um einen „starken" Effekt handelt.

Bei einem Signifikanztest wird lediglich geprüft, ob ein in der Stichprobe gefundenes Ergebnis (sehr) wahrscheinlich auf die Grundgesamtheit übertragen werden kann. Auf Basis eines Signifikanztests lässt sich beispielsweise argumentieren, dass der in der Stichprobe gefundene Zusammenhang zwischen Bildung und Einkommen (sehr) wahrscheinlich auch in der Grundgesamtheit existiert. Über die theoretische Bedeutung des Effekts bzw. des Zusammenhangs von zwei Merkmalen kann ein Signifikanztest keine Aussagen treffen (zur Interpretation von Signifikanztests siehe z. B. Schnell et al. 2013, S. 704–723).

Fazit Die Datenaufbereitung und -analyse ist eine wichtige Phase in einem (quantitativen) Forschungsprojekt. In den vergangenen Jahren hat sich insbesondere die quantitative Datenanalyse immer weiter ausdifferenziert. Zu dieser Entwicklung hat zweifellos der technische Fortschritt beigetragen; immer leistungsfähigere Computer erlauben die Berechnung von komplexen statistischen Model-

len. Auch die praktische Umsetzung in den klassischen Statistikprogrammen (z. B. SPSS, Stata, R) ist heute wesentlich komfortabler als noch vor wenigen Jahren. Bei aller Begeisterung für die Anwendung fortgeschrittener statistischer Verfahren ist an dieser Stelle explizit auf die Anwendungsvoraussetzungen und die Grenzen dieser Verfahren hinzuweisen. Kein Statistikprogramm kann prüfen, ob die Daten für das Verfahren geeignet und/oder die empirische Interpretation angemessen ist. Die Anwendung der Verfahren setzt die (menschliche) Kenntnis der Verfahren voraus.

Publikation 13

Forschungsergebnisse gelten erst als anerkannt, „wenn sie veröffentlicht und damit der Kritik und Überprüfung zugänglich gemacht worden sind“ (Deutsche Forschungsgemeinschaft 2013, S. 43). Die Publikation der Forschungsergebnisse ist damit ein integraler Bestandteil im Forschungsprozess (siehe auch Döring und Bortz 2016, S. 787–804). In diesem Kapitel werden klassische Publikationsformen vorgestellt (Abschnitt 13.1). Das wichtigste Medium für die Veröffentlichung von Forschungsergebnissen sind mittlerweile Fachzeitschriften. In Abschnitt 13.2 werden ausgewählte Fachzeitschriften aufgeführt.

13.1 Publikationsformen

Für die Darstellung und Publikation von Forschungsergebnissen existieren verschiedene Möglichkeiten. Zu den wichtigsten Formen zählen Bericht, Vortrag, Buchveröffentlichung und Artikel in einer Fachzeitschrift. Die einzelnen Darstellungsformen schließen sich nicht aus, sondern können sich ergänzen. Bei einem Vortrag geäußerte Kritikpunkte können beispielweise beim Verfassen eines Fachaufsatzes berücksichtigt werden (z. B. Bauer et al. 2013, S. 137–154).

Bericht Der Bericht als Publikationsform erfüllt in erster Linie die Rechenschaftspflicht der Wissenschaftler gegenüber den Geldgebern. Ein Arbeits- und Ergebnisbericht im Rahmen eines DFG-Projekts umfasst beispielsweise maximal zehn Seiten und enthält unter anderem die Ausgangsfragen bzw. Zielsetzung des Projekts, die Entwicklung der durchgeführten Arbeiten einschließlich Abweichungen vom ursprünglichen Konzept, ggf. wissenschaftliche Fehlschläge, Probleme in der Projektorganisation oder technischen Durchführung sowie die Darstellung

© Springer Fachmedien Wiesbaden GmbH, ein Teil von Springer Nature 2018

M. Tausendpfund, *Quantitative Methoden in der Politikwissenschaft*,
Grundwissen Politik, https://doi.org/10.1007/978-3-658-20698-7_13

der erreichten Ergebnisse und Diskussion im Hinblick auf den relevanten Forschungsstand.[49]

Ein solcher Rechenschaftsbericht ist keine Publikation im eigentlichen Sinne. Vielmehr sind einem solchen Bericht Verzeichnisse der Publikationen beizulegen, die zum Zeitpunkt der Berichterstattung erschienen oder bei Fachzeitschriften angenommen sind. Je nach Förderinstitution und Dauer des Projekts sind auch Zwischenberichte einzureichen, um die weitere Förderung sicherzustellen.

Vortrag Erste Forschungsergebnisse werden meist bei nationalen und internationalen Konferenzen präsentiert. Solche Vorträge informieren über das Projekt und vorläufige Ergebnisse, bringen Anregungen bzw. Verbesserungsvorschläge und können ein erster Schritt zu einer Publikation in einem Tagungsband oder einer Fachzeitschrift sein. Häufig umfasst ein Vortrag auch ein Konferenzpaper, das als Grundlage für eine spätere Veröffentlichung genutzt werden kann.

Buch Die Veröffentlichung der Forschungsergebnisse in einem Buch bietet die Möglichkeit, ausführlich über das Forschungsprojekt und die Ergebnisse zu berichten. Qualifikationsarbeiten (z. B. Promotionen, Habilitationen) werden häufig in Buchform veröffentlicht.

Fachzeitschrift Die Publikation eines Aufsatzes in einer Fachzeitschrift ist mittlerweile die wichtigste Form der Veröffentlichung von Forschungsergebnissen. Aufsätze in Fachzeitschriften sind heute das zentrale Kommunikationsmedium der Wissenschaft. In den Naturwissenschaften gilt dies schon seit vielen Jahren, aber auch in den Sozialwissenschaften gelten Fachzeitschriften heute als unverzichtbar. Die Bedeutung von Fachzeitschriften beschreibt Bär (2002, S. 11):

> *„Was für den Bäcker das Brot, für den Metzger die Wurst, das ist für den Wissenschaftler das Paper. Ein Paper ist ein Aufsatz in einer wissenschaftlichen Zeitschrift, der neue Forschungsergebnisse beschreibt."*

Neuere Forschungsergebnisse finden sich praktisch ausschließlich in wissenschaftlichen Fachzeitschriften. Die Veröffentlichung von Forschungsergebnissen in Fachzeitschriften erfüllt zwei Funktionen. Zum einen werden neue Ergebnisse in das Wissenschaftssystem eingespeist und können dann von anderen Wissenschaftlern kritisiert werden. Zum anderen können die Erkenntnisse dazu bei-

49 Der DFG-Leitfaden für Abschlussberichte und Regeln guter wissenschaftlicher Praxis steht als pdf-Dokument unter www.dfg.de/formulare/2_01/ zum Download zur Verfügung (Datum des Abrufs: 15. 11. 2017)

tragen, neue Forschungsprojekte zu initiieren. Deshalb stellt die systematische Aufarbeitung des Forschungsstands zu einem konkreten Thema – besser: einer konkreten Fragestellung – meist den Beginn eines Forschungsprojekts dar. Die Auseinandersetzung mit Fachaufsätzen verdeutlicht den kumulativen Prozess von Wissenschaft. Jede einzelne wissenschaftliche Publikation zielt darauf ab, das „Wissen" ein kleines Stück zu erweitern.

Tages- und Wochenzeitungen sind keine Fachzeitschriften

Tageszeitungen wie die Süddeutsche Zeitung oder die Frankfurter Allgemeine Zeitung und auch Wochenzeitschriften wie der Spiegel oder die Zeit sind keine Fachliteratur. Es handelt sich dabei um journalistische Produkte, die ggf. über wissenschaftliche Forschungsergebnisse berichten.

Die Veröffentlichung in sichtbaren Fachzeitschriften sowie die Einwerbung von sogenannten Drittmitteln gelten im Wissenschaftssystem als zentrale Indikatoren für wissenschaftliche Leistungen (z. B. Plümper und Schimmelfennig 2007, S. 99). Deshalb spielen Publikationen in Fachzeitschriften auch für akademische Karrieren eine bedeutsame Rolle.

13.2 Sozialwissenschaftliche Fachzeitschriften

Die Auswahl an sozialwissenschaftlichen Fachzeitschriften ist groß. In der elektronischen Zeitschriftenbibliothek finden sich über 4500 politikwissenschaftliche und knapp 4200 soziologische Fachzeitschriften. Die Aufnahme in die elektronische Zeitschriftenbibliothek ist allerdings kein Kriterium für die Qualität einer Fachzeitschrift. Zur Bewertung von Fachzeitschriften sind zwei Kriterien heranzuziehen: Erstens die Qualitätssicherung eines Aufsatzes durch ein Begutachtungsverfahren und zweitens die Aufnahme der Zeitschrift in den Social Science Citation Index (SSCI).

Begutachtung von Fachaufsätzen Die Begutachtung (review) eines Fachaufsatzes vor der Veröffentlichung stellt ein wichtiges Element der Qualitätssicherung wissenschaftlicher Ergebnisse dar. Nach Stykow et al. (2010, S. 192) gelten nur Fachzeitschriften, die ihre Beiträge bereits vor der Veröffentlichung einer Kontrolle durch Fachexperten unterziehen, als wissenschaftliche Fachzeitschriften. Eine

ähnliche Einschätzung findet sich auch in der Denkschrift der Deutschen Forschungsgemeinschaft zur „Sicherung guter wissenschaftlicher Praxis“:

> *„Alle angesehenen wissenschaftlichen Zeitschriften veröffentlichen heute nur solche Arbeiten, die von sachverständigen Gutachterinnen und Gutachtern auf ihre Validität und Originalität geprüft worden sind.“ (Deutsche Forschungsgemeinschaft 2013, S. 43)*

Ein solches Begutachtungsverfahren wird auch als review system bezeichnet. Es gibt verschiedene Varianten der Begutachtungsverfahren, das Wichtigste ist sicherlich das Peer-Review-Verfahren. Beim peer-review-Verfahren erhalten fachspezifische Gutachter den (anonymisierten) Aufsatz und geben den Herausgebern einer Fachzeitschrift eine Empfehlung, ob der eingereichte Beitrag in der Fachzeitschrift veröffentlicht werden soll oder nicht. So wird einerseits die Qualität der einzelnen Aufsätze und zum anderen die Qualität der Zeitschrift als Ganzes gewährleistet.

Wie funktioniert ein Review-Verfahren?

Die Autoren reichen einen (meist anonymisierten) Aufsatz bei einer Fachzeitschrift ein. Die Herausgeber wählen bzw. suchen geeignete Gutachter aus, die die Qualität eines Aufsatzes prüfen. Diese verfassen ein Gutachten zum eingereichten Aufsatz und sprechen eine Empfehlung aus, ob ein Artikel veröffentlicht werden sollte. Dabei lassen sich grob vier zusammenfassende Urteile der Gutachter unterscheiden: Das Manuskript sollte erstens wie vorliegend ohne Änderungen angenommen werden zweitens, nach geringfügigen Überarbeitungen angenommen werden, drittens nach gründlicher Überarbeitung erneut begutachtet werden und viertens abgelehnt werden. Die dritte Kategorie „Überarbeitung und erneute Begutachtung“ (auch als „revise and resubmit“ bezeichnet) stellt dabei den Regelfall dar. Zwischen Ersteinreichung eines Fachaufsatzes und Publikation können 12 bis 18 Monate vergehen. Eine lesenswerte Darstellung des Begutachtungsverfahrens bietet Plümper (2012, S. 153–157; siehe auch Bauer et al. 2013, S. 152–153).

Selbstverständlich wird nicht jeder eingereichte Aufsatz in einer Fachzeitschrift veröffentlicht. Die Annahmequote bei deutschsprachigen Fachzeitschriften liegt etwa bei 30 bis 40 Prozent. Von etwa zehn eingereichten Artikeln werden etwa

drei bis vier auch wirklich in der Zeitschrift veröffentlicht. Bei hochkarätigen englischsprachigen Fachzeitschriften ist die Ablehnungsquote deutlich höher (zum Teil über 90 Prozent). Begutachtungsverfahren erhöhen zwar die Wahrscheinlichkeit einen qualitativ hochwertigen Beitrag lesen zu können, bieten allerdings keine Garantie für die Güte einer Publikation.

Deutschsprachige Fachzeitschriften In Tabelle 19 finden sich ausgewählte deutschsprachige sozialwissenschaftliche Fachzeitschriften. Im Zusammenhang

Tabelle 19 Ausgewählte deutschsprachige Fachzeitschriften

Titel	Reviewverfahren
Aus Politik und Zeitgeschichte (APuZ)	Nein
Berliner Journal für Soziologie (BJS)	Herausgebergremium
Der moderne Staat (dms)	Double Blind Peer Review
Kölner Zeitschrift für Soziologie und Sozialpsychologie (KZfSS)	Double Blind Peer Review
Leviathan. Zeitschrift für Sozialwissenschaft	Begutachtungsverfahren
Politische Vierteljahresschrift (PVS)	Double Blind Peer Review
Soziale Welt	Begutachtungsverfahren
Zeitschrift für Familienforschung (ZfF)	Double Blind Peer Review
Zeitschrift für Internationale Beziehungen (ZIB)	Double Blind Peer Review
Zeitschrift für Parlamentsfragen (ZParl)	Redaktion
Zeitschrift für Politik (ZfP)	Double Blind Peer Review
Zeitschrift für Politikwissenschaft (ZPol)	Double Blind Peer Review
Zeitschrift für Soziologie (ZfS)	Double Blind Peer Review
Zeitschrift für Staats- und Europawissenschaften (ZSE)	Double Blind Peer Review
Zeitschrift für Vergleichende Politikwissenschaft (ZfVP)	Double Blind Peer Review

Quelle: Eigene Zusammenstellung auf Angaben der jeweiligen Homepage der Fachzeitschrift.

mit Begutachtungsverfahren ist dabei anzumerken, dass die Beiträge der – nicht nur bei Studierenden – wohl bekanntesten deutschsprachigen Zeitschrift „Aus Politik und Zeitgeschichte“ (APuZ) kein Begutachtungsverfahren durchlaufen. Natürlich finden sich auch in der APuZ gelungene Beiträge, allerdings handelt es sich

streng genommen nicht um eine wissenschaftliche Fachzeitschrift.[50] Dies sollte bei der Aufarbeitung des Forschungsstands berücksichtigt werden.

In Zeiten von Globalisierung und Europäisierung endet die Veröffentlichung von wissenschaftlichen Ergebnissen allerdings nicht an nationalen (Sprach-)Grenzen. Vielmehr werden Forschungsergebnisse (auch von deutschen Sozialwissenschaftlern) zunehmend in internationalen Fachzeitschriften und somit meist in englischer Sprache publiziert. Besonders sichtbar sind Zeitschriften, die im Social Sciences Citation Index (SSCI) enthalten sind (Plümper und Schimmelfennig 2007, S. 116).

Was ist der Social Sciences Citation Index?

Beim Social Sciences Citation Index (SSCI) handelt es sich um eine Zusammenstellung von ausgewählten sozialwissenschaftlichen Fachzeitschriften. Der SSCI umfasst aktuell mehr als 3200 Fachzeitschriften in 50 sozialwissenschaftlichen Disziplinen – von Anthropologie über Geographie bis hin zur Soziologie. Die Politische Vierteljahresschrift (PVS), die Kölner Zeitschrift für Soziologie und Sozialpsychologie (KZfSS) und die Zeitschrift für Soziologie (ZfS) sind im SSCI enthalten. Die meisten SSCI-Fachzeitschriften erscheinen allerdings ausschließlich in englischer Sprache. Über das Web of Science lassen sich die in den SSCI-Fachzeitschriften veröffentlichten Aufsätze elegant recherchieren.

Für die Darstellung des Forschungsstands ist die Auseinandersetzung mit englischsprachigen Publikationen obligatorisch. Wer den Forschungsstand zu einer (konkreten) Forschungsfrage nur auf Basis deutschsprachiger Literatur aufbereitet, der „fischt“ an einem kleinen See. Die großen „Fische“ finden sich aber nur im (englischsprachigen) Ozean.

50 Auf der Homepage der Bundeszentrale für politische Bildung heißt es: „Sie [Aus Politik und Zeitgeschichte] veröffentlicht wissenschaftlich fundierte, allgemein verständliche Beiträge zu zeitgeschichtlichen und sozialwissenschaftlichen Themen sowie zu aktuellen politischen Fragen. Sie ist ein Forum kontroverser Diskussion, eine Einführung in komplexe Wissensgebiete und bietet eine ausgewogene Mischung aus grundsätzlichen und aktuellen Analysen.“ Quelle: www.bpb.de/shop/zeitschriften/apuz/. Zugriff am 11. 11. 2017.

Politikwissenschaft Der SSCI enthält rund 160 politikwissenschaftliche Fachzeitschriften. In Tabelle 20 sind die – auf Basis des 5-Jahres-Impact-Faktors – zehn wichtigsten Fachzeitschriften für die Politikwissenschaft aufgeführt.

Tabelle 20 Politikwissenschaftliche Fachzeitschriften im SSCI

Rang	Titel	5-Jahres-Impact-Faktor
1	American Political Science Review	6,341
2	Political Analysis	6,098
3	American Journal of Political Science	5,424
4	Annual Review of Political Science	4,288
5	Journal of Peace Research	3,283
6	Perspectives on Politics	3,257
7	Political Geography	3,020
8	Global Evironmental Politics	2,977
9	Journal of Politics	2,903
10	Governance	2,866

Quelle: Eigene Zusammenstellung auf Basis der Angaben des SSCI (Stand: 2017).

Der in Tabelle 20 ausgewiesene Impact-Faktor ist ein Indikator für die Bedeutung einer Fachzeitschrift. Vereinfacht formuliert erfasst der Impact-Faktor, wie oft Beiträge aus einer Fachzeitschrift in anderen Fachzeitschriften zitiert werden. Je größer der Impact-Faktor, desto „angesehener" ist eine Fachzeitschrift. Der Impact-Faktor ist als Indikator für die Bedeutung einer Fachzeitschrift durchaus umstritten (z. B. Fleck 2013; siehe auch Bauer et al. 2013, S. 145), aber er bietet zumindest eine grobe Orientierung im Hinblick auf die Relevanz einer Fachzeitschrift für die jeweilige Disziplin.

Soziologie Der SSCI für die Soziologie umfasst aktuell rund 140 Fachzeitschriften. In der Tabelle 21 sind die zehn soziologischen Fachzeitschriften mit dem höchsten 5-Jahres-Impact-Faktor aufgeführt.

Tabelle 21 Soziologische Fachzeitschriften im SSCI

Rang	Titel	5-Jahres-Impact-Faktor
1	Annual Review of Sociology	8,381
2	American Sociological Review	6,920
3	American Journal of Sociology	5,811
4	Sociological Methodology	5,622
5	Social Networks	4,113
6	Journal of Consumer Culture	3,952
7	Annals of Tourism Research	3,882
8	Sociological Theory	3,864
9	Journal of Marriage and Family	3,725
10	Sociology of Education	3,564

Quelle: Eigene Zusammenstellung auf Basis der Angaben des SSCI (Stand: 2017).

Verwaltungswissenschaft In der Kategorie „Public Administration" finden sich im SSCI rund 45 Fachzeitschriften mit verwaltungswissenschaftlicher Ausrichtung. In Tabelle 22 sind die zehn verwaltungswissenschaftlichen Fachzeitschriften mit dem höchsten 5-Jahres-Impact-Faktor dargestellt.

Tabelle 22 Verwaltungswissenschaftliche Fachzeitschriften im SSCI

Rang	Titel	5-Jahres-Impact-Faktor
1	Journal of Public Administration Research and Theory	4,545
2	Journal of Policy Analysis and Management	3,135
3	Governance	2,866
4	Public Administration Review	2,707
5	Journal of European Social Policy	2,364
6	Policy Study Journal	2,331
7	Policy Sciences	2,322
8	Regulation & Governance	2,290
9	Public Administration	2,273
10	Environment and Planning C: Politics and Space	2,161

Quelle: Eigene Zusammenstellung auf Basis der Angaben des SSCI (Stand: 2017).

Publikationen als Start- und Endpunkt Publikationen bilden den Start- und Endpunkt eines Forschungsprojekts. Auf Basis existierender Veröffentlichungen werden der Forschungsstand zu einem Thema aufgearbeitet und Forschungslücken aufgezeigt, die das Forschungsprojekt schließen soll. Die Ergebnisse eines Forschungsprojekts werden veröffentlicht und können dazu beitragen, das gesammelte Wissen ein kleines Stück zu erweitern. Im Anschluss an Forschungsprojekte ergeben sich häufig neue Fragen, die wiederum in neuen Forschungsvorhaben untersucht werden können. Die Sichtung und die Lektüre (relevanter) Veröffentlichungen stellt daher eine Daueraufgabe im Forschungsprozess dar.

14 Weiterführende Literaturhinweise

Das vorliegende Buch bietet eine Einführung in die (quantitativen) Methoden der empirischen Sozialforschung und die Logik sozialwissenschaftlicher Forschung. Es existiert umfangreiche Literatur zu den Grundlagen und fortgeschrittenen Methoden empirischer Sozialforschung. In diesem Kapitel werden Bücher und Sammelbände zu den Methoden empirischer Sozialforschung vorgestellt. Darüber hinaus werden Fachzeitschriften dargestellt, die sich schwerpunktmäßig mit den Methoden der empirischen Sozialforschung beschäftigen.

14.1 Monographien und Sammelbände

In diesem Abschnitt werden einzelne Bücher und Sammelbände vorgestellt, die einzelne Themen dieser Einführung vertiefen und/oder einen allgemeinen Überblick zur empirischen Sozialforschung geben.

Baur und Blasius (2014) Baur, Nina und Jörg Blasius, Hrsg. 2014. Handbuch Methoden der empirischen Sozialforschung. Wiesbaden: Springer VS.
In 88 Beiträgen bietet das Handbuch einen umfangreichen Überblick zu den quantitativen und qualitativen Methoden der empirischen Sozialforschung (Baur und Blasius 2014a). Die quantitativen und qualitativen Methoden werden gleichberechtigt dargestellt und es werden Gemeinsamkeiten und Unterschiede der beiden Forschungstraditionen herausgearbeitet. Ein besonderer Schwerpunkt liegt dabei auf der Datenerhebung, insbesondere auf standardisierten und offenen Befragungen, aber es werden auch weitere Datentypen vorgestellt (z. B. Paradaten, Daten der amtlichen Statistik, Dokumente, Bilder, Artefakte, Videodaten, Aggregat-, Netzwerk- und Paneldaten).

© Springer Fachmedien Wiesbaden GmbH, ein Teil von Springer Nature 2018

M. Tausendpfund, *Quantitative Methoden in der Politikwissenschaft*,
Grundwissen Politik, https://doi.org/10.1007/978-3-658-20698-7_14

Diaz-Bone und Weischer (2015) Diaz-Bone, Rainer und Christoph Weischer, Hrsg. 2015. Methoden-Lexikon für die Sozialwissenschaften. Wiesbaden: Springer VS.
Das Methoden-Lexikon für die Sozialwissenschaften präsentiert Grundkonzepte, Verfahren und Ansätze aus dem Bereich der sozialwissenschaftlichen Methoden (Diaz-Bone und Weischer 2015). Das Lexikon bietet in alphabetischer Folge verschiedene Typen von Artikeln: Mehr oder weniger kurze Beiträge zu einzelnen Begriffen, längere Artikel, die wichtige Schlüsselkonzepte der empirischen Forschung behandeln, und schließlich Überblicksartikel, die sich jenseits der begrifflichen Ordnung mit den Praktiken, Strukturen und Diskursen der empirischen Forschung befassen. Daher eignet sich das Lexikon in erster Linie als Nachschlagewerk und richtet sich an Personen mit Grundkenntnissen.

Diekmann (2006) Diekmann, Andreas, Hrsg. 2006. Methoden der Sozialforschung. Kölner Zeitschrift für Soziologie und Sozialpsychologie. Sonderheft 44/2004. Wiesbaden: VS Verlag für Sozialwissenschaften.
Dieser Sonderband der Kölner Zeitschrift für Soziologie und Sozialpsychologie bietet einen Überblick über aktuellere Entwicklungen der empirischen Sozialforschung (Diekmann 2006). Die einzelnen Beiträge sind sechs verschiedenen Bereichen zugeordnet: „Wissenschaftstheoretische Grundlagen", „Stichproben, Gewinnung von Daten, Datenqualität", „Sozialer Kontext und Netzwerke", „Zeitbezogene Daten", „Spezielle Probleme der Datenanalyse" sowie „Modellbildung und Simulation". Der Band eignet sich für Leser, die bereits Grundkenntnisse der empirischen Sozialforschung haben und diese erweitern möchten.

Diekmann (2011) Diekmann, Andreas. 2011. Empirische Sozialforschung. Grundlagen, Methoden, Anwendungen. Reinbek: Rowohlt.
Die äußerst lesenswerte Einführung behandelt grundlegende Methoden der empirischen Sozialforschung. In vier Abschnitten – Grundlagen, Untersuchungsplanung, Datenerhebung und Datenauswertung – bietet der Band auf mehr als 700 Seiten eine gelungene Darstellung der Thematik. Dabei werden die Vor- und Nachteile der einzelnen Untersuchungstechniken herausgearbeitet und an zahlreichen Beispielen aus der Forschungspraxis illustriert.

Dillman et al. (2014) Dillman, Don A., Jolene D. Smyth und Leah Melani Christian. 2014. Internet, Phone, Mail, and Mixed-Mode Surveys. The Tailored Design Method. Hoboken: Wiley.
Unter dem Titel „Mail and Telephone Surveys: The Total Design Method" ist 1978 die erste Auflage des Buchs erschienen, die vierte Auflage mit dem Titel „Internet, Phone, Mail, and Mixed-Mode Surveys. The Tailored Design Method" kennzeich-

net den Wandel der Surveyforschung in den vergangenen 30 Jahren. Die Total Design Method zielt darauf ab, die Beteiligung bei standardisierten Befragungen zu optimieren. Die Bezeichnung „Total Design" bringt dabei zum Ausdruck, dass der komplette Prozess der Befragung – von der Kontaktierung der Personen über die Gestaltung der Erhebungsinstrumente bis hin zu Nachfassaktionen – in den Blick zu nehmen ist. Die Tailored Design Method baut auf den Ergebnissen der Total Design Method auf und erweitert diese Erkenntnisse mit Blick auf die Bedürfnisse unterschiedlicher Zielgruppen. Befragungen müssen an die jeweilige Zielgruppen angepasst werden, um eine optimale Ausschöpfung zu erzielen. Das Buch zählt zu den Klassikern der Surveyforschung und ist äußerst zugänglich verfasst.

Döring und Bortz (2016) Döring, Nicola, und Jürgen Bortz. 2016. Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler. Heidelberg: Springer.
Das Lehrbuch richtet sich an Studierende, die sich das Handwerkszeug empirischer Forschung aneignen und einen Überblick gewinnen wollen (Döring und Bortz 2016). Es werden die quantitativen und qualitativen Methoden vorgestellt; die Beispiele sind weitgehend der Psychologie entnommen. Ein ausführliches Sach- und Namensverzeichnis ermöglicht den selektiven Zugriff auf konkrete Fragestellungen und ein Glossar hilft beim Verständnis der Fachsprache. Im Band finden sich auch zahlreiche Übungsaufgaben und Musterlösungen.

Engel et al. (2015) Engel, Uwe, Ben Jann, Peter Lynn, Annette Scherpenzeel und Patrick Sturgis, Hrsg. 2015. Improving Survey Methods: Lessons from Recent Research. New York: Routledge.
Der von Engel et al. (2015) herausgegebene Sammelband bietet eine Übersicht über neuere Entwicklungen der Umfragemethoden. In acht Teilen und 33 Kapiteln beschäftigen sich die Autoren mit den Herausforderungen der unterschiedlichen Befragungsformen, Interviewereffekte, sensiblen Fragen, der Durchführung von Internetbefragungen und Panelerhebungen, der Zusammenführung von unterschiedlichen Datenquellen, Strategien zur Verringerung von Nonresponse sowie Ersetzungsverfahren für fehlende Daten. Der Band bündelt zentrale Ergebnisse des DFG-Schwerpunktprogramms „Survey Methodology", das von der Universität Bremen koordiniert wird. Auf der Homepage unter www.survey-methodology.de finden sich umfangreiche Informationen zum Band sowie ein Verzeichnis der Publikationen, die im Kontext des DFG-Schwerpunktprogramms entstanden sind.

Groves et al. (2009) Groves, Robert M., Floyd J. Fowler, Mick P. Couper, James M. Lepkowski, Eleanor Singer und Roger Tourangeau, Hrsg. 2009. Survey Methodology. Hoboken: Wiley.
Das Buch von Groves et al. (2009) bietet eine ausgezeichnete Einführung in die Surveyforschung. In zwölf Kapiteln werden die Grundlagen (z. B. Stichprobenziehung, Datenerhebung) und Herausforderungen (z. B. Nonresponse) der Umfrageforschung thematisiert. Neuere Entwicklungen wie der Dual-Frame-Ansatz sowie Internetbefragungen werden ebenfalls berücksichtigt. Die Autoren sensibilisieren für die zahlreichen Herausforderungen, die bei der Durchführung einer Befragung zu berücksichtigen sind (z. B. Coverage-Fehler). Praktische Beispiele und zahlreiche Übungsaufgaben am Ende jedes Kapitels helfen, die vermittelten Inhalte zu verarbeiten und den eigenen Lernfortschritt zu überprüfen.

Häder (2015) Häder, Michael. 2015. Empirische Sozialforschung. Eine Einführung. Wiesbaden: Springer VS.
Das Buch behandelt die theoretischen Grundlagen der empirischen Sozialforschung, die Schritte bei der Konzipierung und Umsetzung eines Forschungsprojekts sowie die vielfältigen Varianten der Datenerhebung. Außerdem werden die bei der Auswahl der Untersuchungseinheiten einzusetzenden Methoden ebenso betrachtet wie die Prinzipien, die bei der Auswertung und Dokumentation der Befunde zu beachten sind.

Mallock et al. (2016) Mallock, Wolfgang, Udo Riege, und Matthias Stahl. 2016. Informationsressourcen für die Sozialwissenschaften. Datenbanken – Längsschnittuntersuchungen – Portale – Institutionen. Wiesbaden: Springer VS.
Das Buch bietet eine Zusammenstellung nützlicher Informationen für die Sozialwissenschaften. Auf 260 Seiten werden sozialwissenschaftliche Institutionen, Forschungsdatenzentren, Portale, Längsschnittuntersuchungen sowie Datenbanken vorgestellt. Damit ist das Buch ein nützliches Nachschlagewerk, das als Wegweiser der Informationsressourcen in den Sozialwissenschaften dient.

Schnell (2012) Schnell, Rainer. 2012. Survey-Interviews. Methoden standardisierter Befragungen. Wiesbaden: VS Verlag für Sozialwissenschaften.
Das Lehrbuch behandelt die Durchführung standardisierter Befragungen. Zunächst werden allgemeine Probleme der Frageformulierung und des Antwortverhaltens behandelt. Besondere Aufmerksamkeit wird den praktischen Problemen der Fragebogenkonstruktion und der Durchführung von Pretests gewidmet. Für alle Formen standardisierter Interviews (persönlich, schriftlich, telefonisch, online) werden die jeweiligen Besonderheiten der Befragungsinstrumente, der Stichprobenverfahren und der Durchführung dargestellt. Hierbei werden vor allem

Nonresponse-Probleme ausführlich diskutiert und Möglichkeiten zu ihrer Vermeidung dargestellt.

Schnell et al. (2013) Schnell, Rainer, Paul B. Hill und Elke Esser. 2013. Methoden der empirischen Sozialforschung. München: Oldenbourg.
In neun Kapiteln – von der historischen Entwicklung der empirischen Sozialforschung über Fragen der Konzeptspezifikation und Operationalisierung bis hin zur Datenaufbereitung und Analyse – bietet das Werk auf mehr als 500 Seiten einen einführenden Einblick in die empirische Sozialforschung. Der Band ist klar strukturiert, verständlich geschrieben und verknüpft Theorie und Praxis. Das Buch bietet eine gelungene Einführung und Vertiefung der Grundlagen der empirischen Sozialforschung.

Wolf et al. (2016) Wolf, Christof, Dominique Joye, Tom E. C. Smith, und Yang-Chih Fu, Hrsg. 2016. The Sage Handbook of Survey Methodology. London: SAGE.
Das „SAGE Handbook of Survey Methodology" (Wolf et al. 2016) bietet in 43 Kapiteln einen kenntnisreichen Überblick zur Umfrageforschung. Das Handbuch ist in neun größere Abschnitte unterteilt, die alle Phasen eines Forschungsprojekts abdecken – von der Konzeption, Gestaltung und Planung einer Befragung über die Stichprobenziehung, Messung und Datenerhebung bis hin zu Fragen der Datenaufbereitung und Prüfung der Datenqualität. Dabei werden insbesondere die Herausforderungen internationaler Forschungsprojekte berücksichtigt (z. B. Vergleichbarkeit von Messinstrumenten).

14.2 Fachzeitschriften

Neuere Forschungsergebnisse zu den Methoden empirischer Sozialforschung werden praktisch ausschließlich in (englischsprachigen) Fachzeitschriften veröffentlicht. Diese Fachzeitschriften können über die elektronische Zeitschriftenbibliothek recherchiert werden. Dort finden sich auch allgemeine Angaben zur Zeitschrift.

Die Zugriffsmöglichkeiten auf die Aufsätze einer Fachzeitschrift sind institutionsabhängig und werden durch verschiedenfarbige Punkte angezeigt. Falls eine Zeitschrift nicht im Volltext verfügbar ist stehen in der Regel Inhaltsverzeichnisse und oftmals auch Zusammenfassungen (Abstracts) kostenlos zur Verfügung. Für die Recherche von Aufsätzen zu einem spezifischen Thema (z. B. Nonresponse) bietet sich die Recherche in Literaturdatenbanken an (z. B. Web of Science).

Im Folgenden findet sich eine alphabetische Übersicht methodischer Fachzeitschriften. Die Fachzeitschriften behandeln unterschiedliche Fragestellungen

zu den Methoden empirischer Sozialforschung. Ein Blick in die Fachzeitschriften ist hilfreich, um einerseits eine Übersicht zu den aktuellen Themen zu erhalten und andererseits sich von der Komplexität methodischer Fragestellungen selbst zu überzeugen.

Field Methods Die Fachzeitschrift „Field Methods" richtet sich an Personen, die selbst Daten erheben. Sie informiert über methodische Weiterentwicklungen, die Anwendung spezifischer Erhebungstechniken sowie praktische Tipps und Tricks. Die Fachzeitschrift berücksichtigt dabei die qualitativen und quantitativen Methoden der empirischen Sozialforschung.

International Journal of Public Opinion Research Das „International Journal of Public Opinion Research" veröffentlicht Fachaufsätze, die sich mit der Verbesserung von Umfragemethoden beschäftigen. Außerdem wird auch die Rolle von Bevölkerungsumfragen für politische Entscheidungsprozesse und Wahlverhalten betrachtet.

Journal of Survey Statistics and Methodology Das „Journal of Survey Statistics and Methodology" behandelt statistische und methodische Fragen zu Stichprobenerhebungen, Volkszählungen sowie anderen Datenformen. Die Themen der Zeitschrift lassen sich in drei Bereiche gliedern. Im Bereich „Survey Statistics" werden unter anderem Fragen der Stichprobenziehung, Imputationsverfahren und neuere Methoden der Datenanalyse diskutiert. Der Bereich „Survey Methodology" betrachtet die methodische Weiterentwicklung. Innerhalb der Sektion werden methodische Experimente, Datenerhebungsverfahren und die Nutzung von sogenannten Paradaten diskutiert. Schließlich konzentriert sich der Bereich „Applications" auf innovative Ansätze und praktische Beiträge der Datenerhebung.

Journal of Official Statistics Das „Journal of Official Statistics" wird vom schwedischen Statistikamt (Statistics Sweden) herausgegeben. Die Fachzeitschrift veröffentlicht in erster Linie Forschungsbeiträge zu amtlichen Daten und deckt dabei die gesamte Breite der Surveyforschung ab – von Stichprobenplänen über die Qualitätskontrolle der amtlichen Statistik bis hin zur internationalen Zusammenarbeit.

MDA. Methods, Data, Analyses Die MDA deckt alle Fragestellungen der empirischen Sozialforschung ab, insbesondere aus dem Bereich der Umfragemethodik. Im Mittelpunkt stehen Beiträge, die die methodischen und/oder statistischen Kenntnisse erweitern, sowie Artikel, die sich mit der Anwendung der Methoden der empirischen Sozialforschung in der Forschungspraxis beschäftigen, oder sol-

che, in denen ein statistisches Verfahren exemplarisch angewandt wird. Seit 2014 erscheint die Zeitschrift in englischer Sprache; von 2007 bis 2013 ist die Zeitschrift unter dem Titel „Methoden – Daten – Analysen. Zeitschrift für Empirische Sozialforschung" erschienen. Die MDA ist 2007 aus den beiden nunmehr eingestellten GESIS-Zeitschriften „ZA-Informationen" und „ZUMA-Nachrichten" hervorgegangen. Alle Ausgaben der MDA sowie die Vorläuferzeitschriften sind online bei GESIS (www.gesis.org) verfügbar. Die MDA kann auch kostenfrei abonniert werden.

Survey Methods: Insights from the Field Bei „Survey Methods" (http://surveyinsights.org) handelt es sich um eine Online-Zeitschrift. Die Zeitschrift zielt darauf ab, den fachlichen Austausch über praktische Fragen der Umfrageforschung und die Diskussion über neue und vielversprechende Wege in der Umfrageforschung zu fördern. Dabei behandelt die Zeitschrift unter anderem Fragen zu Fragebogendesign, Sampling, Interviewerschulung, Feldverwaltung, Datenaufbereitung und Dokumentation. Die Publikation als Online-Journal soll dazu beitragen, die schnelle Veröffentlichung sowie Diskussionen zu fördern.

Survey Research Methods Bei der „Survey Research Methods" (SRM) handelt es sich um eine Open-Access-Zeitschrift der europäischen Vereinigung der Surveyforscher (European Survey Research Association). Die Zeitschrift deckt ein umfangreiches Themenspektrum ab: Erhebungsdesign, Fragenbogenkonstruktion, Datenerhebung, Gewichtungsverfahren und Analysetechniken.

Sociological Methods & Research Die „Sociological Methods & Research" (SMR) ist eine vierteljährlich erscheinende Fachzeitschrift, die insbesondere methodische Fragen und Herausforderungen zur Datenerhebung, Messung und Datenanalyse in den Blick nimmt.

Survey Methodology Die Zeitschrift „Survey Methodology" der kanadischen Statistikbehörde (Statistics Canada) veröffentlicht Beiträge, die sich schwerpunktmäßig mit methodischen Fragen rund um die amtliche Statistik beschäftigen. Dabei werden Datenquellen und Erhebungstechniken diskutiert, neuere Entwicklungen der Datenintegration vorgestellt und aktuelle Datenanalyseverfahren präsentiert. Die Zeitschrift erscheint als Open-Access; alle Aufsätze stehen als pdf-Dokumente zur Verfügung.

Public Opinion Quarterly Die seit 1937 erscheinende „Public Opinion Quarterly" (POQ) gehört zu den einflussreichsten Fachzeitschriften im Bereich der empirischen Sozialforschung. In der Zeitschrift werden theoretische und empirische

Beiträge zur Umfrageforschung, Fragebogenkonstruktion sowie Stichprobenziehung veröffentlicht. Auch Effekte der Befragungsform (Mode-Effekte) werden in der Fachzeitschrift diskutiert. Ausgewählte Aufsätze sind frei zugänglich verfügbar.

Political Science Research and Methods Die „Political Science Research and Methods" (PSRM) ist eine jüngere politikwissenschaftliche Fachzeitschrift, die Aufsätze zu allen Teilgebieten der Politikwissenschaft veröffentlicht. Der Schwerpunkt liegt auf der methodischen Umsetzung theoretischer und empirischer Fragestellungen.

Quality and Quantity „Quality and Quantity" ist eine interdisziplinäre Fachzeitschrift zu sozialwissenschaftlichen Methoden. Der Schwerpunkt liegt auf Beiträgen zur Datenanalyse, aber es werden auch Aufsätze zur Datenerhebung publiziert. Das Themenspektrum reicht von Simulationsstudien über mathematische Modelle des Wahlverhaltens bis hin zu allgemeineren Diskussionen der sozialwissenschaftlichen Forschungslogik.

Literatur

AAPOR. 2011. *Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys.* Deerfield: American Association for Public Opinion Research.

Abendschön, Simone. 2010. *Die Anfänge demokratischer Bürgerschaft. Sozialisation politischer und demokratischer Werte und Normen im jungen Kindesalter.* Baden-Baden: Nomos.

Abendschön, Simone, Hrsg. 2013. *Growing into Politics. Contexts and Timing of Political Socialization.* Colchester: ECPR Press.

Abendschön, Simone, und Sigrid Roßteutscher. 2011. Jugend und Politik: Verliert die Demokratie ihren Nachwuchs? In *Der unbekannte Wähler? Mythen und Fakten über das Wahlverhalten der Deutschen,* Hrsg. Evelyn Bytzek und Sigrid Roßteutscher, 59–80. Frankfurt: Campus.

Abendschön, Simone, und Markus Tausendpfund. 2017. Political Knowledge of Children and the Role of Sociostructural Factors. *American Behavioral Scientist* 61 (2): 204–221.

ADM. 2013. *Jahresbericht 2013.* Frankfurt: Arbeitskreis Deutscher Markt- und Sozialforschungsinstitute e. V.

ADM. 2015. *Jahresbericht 2015.* Frankfurt: Arbeitskreis Deutscher Markt- und Sozialforschungsinstitute e. V.

ADM. 2016. *Jahresbericht 2016.* Frankfurt: Arbeitskreis Deutscher Markt- und Sozialforschungsinstitute e. V.

Akremi, Leila. 2014. Stichprobenziehung in der qualitativen Sozialforschung. In *Handbuch Methoden der empirischen Sozialforschung,* Hrsg. Nina Baur und Jörg Blasius, 265–282. Wiesbaden: Springer VS.

Albers, Ines. 1997. Einwohnermelderegister-Stichproben in der Praxis. Ein Erfahrungsbericht. In *Stichproben in der Umfragepraxis,* Hrsg. Siegfried Gabler und Jürgen H. P. Hoffmeyer-Zlotnik, 117–126. Wiesbaden: Westdeutscher Verlag.

Allison, Paul D. 2002. *Missing Data.* Thousand Oaks: Sage.

Almond, Gabriel A., und Sidney Verba. 1963. *The Civic Culture. Political Attitudes and Democracy in Five Nations.* Boston: Little, Brown and Company.

© Springer Fachmedien Wiesbaden GmbH, ein Teil von Springer Nature 2018

M. Tausendpfund, *Quantitative Methoden in der Politikwissenschaft*,
Grundwissen Politik, https://doi.org/10.1007/978-3-658-20698-7

Alpheis, Hannes. 1988. *Kontextanalyse. Die Wirkung des sozialen Umfeldes untersucht am Beispiel der Eingliederung von Ausländern.* Wiesbaden: Deutscher Universitäts Verlag.

Amoo, Taiwo, und Hershey H. Friedman. 2001. Do Numeric Values Influence Subjects' Responses to Rating Scales? *Journal of International Marketing and Marketing Research* 26 (1): 41–46.

Aronson, Elliot, Timothy D. Wilson, und Robin M. Akert. 2004. *Sozialpsychologie.* München: Pearson.

Arzheimer, Kai. 2002. *Politikverdrossenheit. Bedeutung, Verwendung und empirische Relevanz eines politikwissenschaftlichen Begriffes.* Wiesbaden: Westdeutscher Verlag.

Arzheimer, Kai. 2006. Jung, dynamisch, Nichtwähler? Der Einfluss von Lebensalter und Kohortenzugehörigkeit auf die Wahlbereitschaft. In *Jugend und Politik: „Voll normal!" Der Beitrag der politischen Soziologie zur Jugendforschung,* Hrsg. Edeltraud Roller, Frank Brettschneider und Jan W. van Deth, 317–335. Wiesbaden: VS Verlag für Sozialwissenschaften.

Arzheimer, Kai. 2009. Ideologien. In *Politische Soziologie. Ein Studienbuch,* Hrsg. Viktoria Kaina und Andrea Römmele, 83–108. Wiesbaden: VS Verlag für Sozialwissenschaften.

Atteslander, Peter. 2010. *Methoden der empirischen Sozialforschung.* Berlin: Erich Schmidt Verlag.

Aust, Folkert, und Helmut Schröder. 2009. Sinkende Stichprobenausschöpfung in der Umfrageforschung – ein Bericht aus der Praxis. In *Umfrageforschung. Herausforderungen und Grenzen,* Hrsg. Martin Weichbold, Johann Bacher und Christof Wolf, 195–273. Wiesbaden: VS Verlag für Sozialwissenschaften.

Bakker, Ryan, Catherine de Vries, Erica Edwards, Liesbet Hooghe, Seth Jolly, Gary Marks, Jonathan Polk, Jan Rovny, Marco Steenbergen, und Milada Anna Vachudova. 2015. Measuring party positions in Europe. *Party Politics* 21 (1): 143–152.

Bär, Siegfried. 2002. *Forschen auf Deutsch.* Frankfurt: Deutsch.

Bauer, Gerrit. 2015. Natürliche Experimente in der Sozialforschung. Ein Überblicksartikel mit ausgewählten Beispielen. In *Experimente in den Sozialwissenschaften. Soziale Welt – Sonderband 22,* Hrsg. Marc Keuschnigg und Tobias Wolbring, 268–288. Baden-Baden: Nomos.

Bauer, Thomas, Gerd Gigerenzer, und Walter Krämer. 2014. *Warum dick nicht doof macht und Genmais nicht tötet. Über Risiken und Nebenwirkungen der Unstatistik.* Frankfurt: Campus.

Bauer, Waldemar, Jörn Bleck-Neuhaus, Rainer Dombois, und Ingo S. Wehrtmann. 2013. *Forschungsprojekte entwickeln – von der Idee bis zur Publikation.* Baden-Baden: Nomos.

Baumert, Jürgen. 1991. Langfristige Auswirkungen der Bildungsexpansion. *Unterrichtswissenschaft* 19 (4): 333–349.

Baumert, Jürgen, Petra Stanat, und Anke Demmrich. 2001. PISA 2000: Untersuchungsgegenstand, theoretische Grundlagen und Durchführung der Studie. In *Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich,* Hrsg. Deutsches PISA-Konsortium, 15–68. Opladen: Leske+Budrich.

Baur, Nina. 2011. Das Ordinalskalenproblem. In *Datenanalyse mit SPSS für Fortgeschrittene 1,* Hrsg. Leila Akremi, Nina Baur und Sabine Fromm, 211–221. Wiesbaden: VS Verlag für Sozialwissenschaften.

Baur, Nina, und Jörg Blasius, Hrsg. 2014a. *Handbuch Methoden der empirischen Sozialforschung.* Wiesbaden: Springer VS.

Baur, Nina, und Jörg Blasius. 2014b. Methoden der empirischen Sozialforschung. In *Handbuch Methoden der empirischen Sozialforschung,* Hrsg. Nina Baur und Jörg Blasius, 41–62. Wiesbaden: Springer VS.

Becker, Oliver Arránz. 2015. Determinanten und Konsequenzen von Trennung und Scheidung. In *Handbuch Familiensoziologie,* Hrsg. Paul B. Hill und Johannes Kopp, 527–561. Wiesbaden: Springer VS.

Behnke, Joachim. 2005. Lassen sich Signifikanztests auf Vollerhebungen anwenden? Einige essayistische Anmerkungen. *Politische Vierteljahresschrift. Online-Beitrag* 46 (1): 1–15.

Behnke, Joachim. 2015. *Logistische Regressionsanalyse. Eine Einführung.* Wiesbaden: Springer VS.

Behnke, Joachim, Nina Baur, und Nathalie Behnke. 2010. *Empirische Methoden der Politikwissenschaft.* Paderborn: Schöningh.

Behnke, Joachim, und Nathalie Behnke. 2006. *Grundlagen der statistischen Datenanalyse. Eine Einführung für Politikwissenschaftler.* Wiesbaden: VS Verlag für Sozialwissenschaften.

Behnke, Joachim, Thomas Gschwend, Delia Schindler, und Kai-Uwe Schnapp, Hrsg. 2006. *Methoden der Politikwissenschaft. Neuere qualitative und quantitative Analyseverfahren.* Baden-Baden: Nomos.

Behnke, Nathalie. 2009. Datenarchive. In *Datenwelten. Datenerhebung und Datenbestände in der Politikwissenschaft,* Hrsg. Kai-Uwe Schnapp, Nathalie Behnke und Joachim Behnke, 392–408. Baden-Baden: Nomos.

Benninghaus, Hans. 2007. *Deskriptive Statistik. Eine Einführung für Sozialwissenschaftler.* Wiesbaden: VS Verlag für Sozialwissenschaften.

Berenskoetter, Felix, Hrsg. 2016. *Concepts in world politics.* Los Angeles: SAGE.

Bergmann, Michael. 2015. *Panel Conditioning. Wirkungsmechanismen und Konsequenzen wiederholter Befragungen.* Baden-Baden: Nomos.

Berninger, Ina, Katrin Botzen, Christian Kolle, Dominikus Vogl, und Oliver Watteler. 2017. *Grundlagen sozialwissenschaftlichen Arbeitens. Eine anwendungsorientierte Einführung.* Stuttgart: UTB.

Best, Henning, und Clemens Kroneberg. 2012. Die Low-Cost-Hypothese. Theoretische Grundlagen und empirische Implikationen. *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 64 (3): 535–561.

Best, Henning, und Christof Wolf. 2010. Logistische Regression. In *Handbuch der sozialwissenschaftlichen Datenanalyse,* Hrsg. Christof Wolf und Henning Best, 827–854. Wiesbaden: VS Verlag für Sozialwissenschaften.

Binder, Tanja, und Andreas M. Wüst. 2004. Inhalte der Europawahlprogramme deutscher Parteien 1979–1999. *Aus Politik und Zeitgeschichte* 54 (17): 38–45.

Blais, André. 2006. What affects voter turnout? *Annual Review of Political Science* 9: 111–125.

Blais, André, und Kees Aarts. 2006. Electoral Systems and Turnout. *Acta Politica* 41 (2): 180–196.

Bläser, Karl-Alois. 2013. Europa im Spiegel der öffentlichen Meinung. Bilanz und Perspektiven des Eurobarometers nach 40 Jahren. *Leviathan* 41 (3): 351–357.

Blasius, Jörg. 2014. Fälschungen von Interviews. In *Handbuch Methoden der empirischen Sozialforschung,* Hrsg. Nina Baur und Jörg Blasius, 323–329. Wiesbaden: Springer VS.

Blasius, Jörg, und Maurice Brandt. 2009. Repräsentativität in Online-Befragungen. In *Umfrageforschung. Herausforderungen und Grenzen,* Hrsg. Martin Weichbold, Johann Bacher und Christof Wolf, 157–177. Wiesbaden: VS Verlag für Sozialwissenschaften.

Blatter, Joachim, Phil C. Langer, und Claudius Wagemann. 2018. *Qualitative Methoden in der Politikwissenschaft. Eine Einführung.* Wiesbaden: Springer VS.

Blohm, Michael, und Achim Koch. 2013. Respondent Incentives in a National Face-to-Face Survey. Effects on Outcome Rates, Sample Composition and Fieldwork Efforts. *methods, data, analyses* 7 (1): 89–122.

Blom, Annelies G., Christina Gathmann, und Ulrich Krieger. 2015. Setting Up an Online Panel Representative of the General Population:The German Internet Panel. *Field Methods* 27 (4): 391–408.

Bogumil, Jörg, Lars Holtkamp, Martin Junkernheinrich, und Uwe Wagschal. 2014. Ursachen kommunaler Haushaltsdefizite. *Politische Vierteljahresschrift* 55 (4): 614–647.

Boomgaarden, Hajo G., Andreas R. T. Schuck, Matthijs Elenbaas, und Claes H. de Vreese. 2011. Mapping EU attitudes: Conceptual and empirical dimensions of Euroscepticism and EU support. *European Union Politics* 12 (2): 241–266.

Borenstein, Michael, Larry V. Hedges, Julian P. T. Higgins, und Hannah R. Rothstein. 2009. *Introduction to Meta-Analysis.* Chichester: Wiley.

Borg, Ingwer, und Christian Treder. 2003. Item-Nonresponse in Mitarbeiterbefragungen. *ZUMA-Nachrichten* 27 (53): 77–95.

Bortz, Jürgen, und Nicola Döring. 2006. *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler.* Heidelberg: Springer.

Bosnjak, Michael, Tanja Dannwolf, Tobias Enderle, Ines Schaurer, Bella Struminskaya, Angela Tanner, und Kai W. Weyandt. 2017. Establishing an Open Probability-Based Mixed-Mode Panel of the General Population in Germany: The GESIS Panel. *Social Science Computer Review:* doi: 10.1177/0894439317697949.

Bourdieu, Pierre. 1983. Ökonomisches Kapital, kulturelles Kapital, soziales Kapital. In *Soziale Ungleichheiten. Soziale Welt. Sonderband 2,* Hrsg. Reinhard Kreckel, 183–198. Göttingen: Schwartz.

Brace, Ian. 2013. *Questionnaire Design. How to plan, structure and write survey material for effective market research.* London: Kogan.

Bradburn, Norman M., Seymour Sudman, und Brian Wansink. 2004. *Asking Questions. The Definitive Guide to Questionnaire Design. For Market Research, Political Polls, and Social and Health Questionnaires.* San Francisco: John Wiley & Sons.

Braun, Daniela, und Hermann Schmitt. 2009. Politische Legitimität. In *Politische Soziologie. Ein Studienbuch,* Hrsg. Viktoria Kaina und Andrea Römmele, 53–81. Wiesbaden: VS Verlag für Sozialwissenschaften.

Braun, Daniela, Nicole Seher, Markus Tausendpfund, und Ansgar Wolsing. 2010. Einstellungen gegenüber Immigranten und die Zustimmung zur Europäischen Integration. Eine Mehrebenenanalyse. Mannheim. Mannheimer Zentrum für Europäische Sozialforschung (Arbeitspapier Nr. 136).

Braun, Daniela, und Markus Tausendpfund. 2013. Immigration als Erklärung für die Zustimmung zur weiteren Europäischen Integration. Eine europaweite empirische Analyse. *Zeitschrift für Vergleichende Politikwissenschaft* 7 (3): 205–226.

Braun, Daniela, und Markus Tausendpfund. 2014. The Impact of the Euro Crisis on Citizens's Support for the European Union. *Journal of European Integration* 36 (3): 231–245.

Braunecker, Claus. 2016. *How to do Empirie, how to do SPSS. Eine Gebrauchsanleitung.* Stuttgart: UTB.

Bräuninger, Thomas, und Marc Debus. 2012. *Parteienwettbewerb in den deutschen Bundesländern.* Wiesbaden: VS Verlag für Sozialwissenschaften.

Brennan, Jason. 2011. The Right to a Competent Electorate. *The Philosophical Quarterly* 61 (245): 700–724.

Brennan, Jason. 2017. *Gegen Demokratie. Warum wir die Politik nicht den Unvernünftigen überlassen dürfen.* Berlin: Ullstein.

Brettschneider, Frank, Anikar M. Haseloff, und Jan Kercher. 2009. Kann man Wahlaussagen verstehen? Über die Sprache der Parteiprogramme zur Bundestagswahl 2009. *Forschung und Lehre* (9): 668–670.

Brettschneider, Frank, und Markus Rettich. 2005. Europa – (k)ein Thema für die Medien. In *Wahl-Kampf um Europa. Analysen aus Anlass der Wahlen zum Europäischen Parlament 2004,* Hrsg. Jens Tenscher, 136–156. Wiesbaden: VS Verlag für Sozialwissenschaften.

Brinegar, Adam, und Seth Jolly. 2004. Integration: Using the Eurobarometer to Measure Support. In *Public opinion and polling around the world,* Hrsg. John G. Geer, 497–504. Santa Barbara: ABC-CLIO.

Broscheid, Andreas, und Thomas Gschwend. 2005. Zur statistischen Analyse von Vollerhebungen. *Politische Vierteljahresschrift. Online-Beitrag* 46 (1): 16–26.

Brüderl, Josef. 2010. Kausalanalyse mit Paneldaten. In *Handbuch der sozialwissenschaftlichen Datenanalyse,* Hrsg. Christof Wolf und Henning Best, 963–994. Wiesbaden: VS Verlag für Sozialwissenschaften.

Bryman, Alan. 2016. *Social Research Methods.* Oxford: Oxford University Press.

Bühlmann, Marc. 2013. Verbundenheit mit der Gemeinde. In *Politik im Kontext: Ist alle Politik lokale Politik? Individuelle und kontextuelle Determinanten politischer Orientierungen,* Hrsg. Jan W. van Deth und Markus Tausendpfund, 329–358. Wiesbaden: Springer VS.

Bühlmann, Marc, und Markus Freitag. 2006. Individual and Contextual Determinants of Electoral Participation. *Swiss Political Science Review* 12 (4): 13–47.

Bürklin, Wilhelm, und Markus Klein. 1998. *Wahlen und Wählerverhalten. Eine Einführung.* Opladen: Leske+Budrich.

Caballero, Claudio. 2014. Nichtwahl. In *Handbuch Wahlforschung,* Hrsg. Jürgen W. Falter und Harald Schoen, 437–488. Wiesbaden: Springer VS.

Campbell, Donald T. 1957. Factors relevant to the validity of experiments in social settings. *Psychological Bulletin* 54 (4): 297–312.

Canache, Damarys, Jeffery J. Mondak, und Mitchell A. Seligson. 2001. Meaning and measurement in cross-national research on satisfaction with democracy. *Public Opinion Quarterly* 65 (4): 506–528.

Castiglione, Dario, Jan W. van Deth, und Guglielmo Wolleb, Hrsg. 2008. *The Handbook of Social Capital.* Oxford: Oxford University Press.

Chalmers, Alan F. 2006. *Wege der Wissenschaft. Einführung in die Wissenschaftstheorie.* Berlin: Springer.

Clark, Nicholas. 2014. Explaining Low Turnout in European Elections: The Role of Issue Salience and Institutional Perceptions in Elections to the European Parliament. *Journal of European Integration* 36 (4): 339–356.

Converse, Jean M., und Stanley Presser. 1986. *Survey Questions. Handcrafting the Standardized Questionnaire.* Beverly Hills: Sage.

Cortina, Jose M. 1993. What Is Coefficient Alpha? An Examination of Theory and Applications. *Journal of Applied Psychology* 78 (1): 98–104.

Cronbach, Lee J. 1951. Coefficient Alpha and the internal structure of tests. *Psychometrika* 16 (3): 297–334.

Dahl, Robert A. 2006. *Politische Gleichheit – ein Ideal?* Hamburg: Hamburger Edition.

Dalton, Russell J. 1984. Cognitive Mobilization and Partisan Dealignment in Advanced Industrial Democracies. *Journal of Politics* 46 (1): 264–284.

Dalton, Russell J. 2004. *Democratic Challenges, Democratic Choices. The Erosion of Political Support in Advanced Industrial Democracies.* Oxford: Oxford University Press.

Dalton, Russell J., und Christopher J. Anderson. 2011. Citizens, Context, and Choice. In *Citizens, Context and Choice. How Context Shapes Citizens' Electoral Choices,* Hrsg. Russell J. Dalton und Christopher J. Anderson, 3–30. Oxford: Oxford University Press.

de Souto Barreto, Philipe, Matteo Cesari, Sandrine Andrieu, Bruno Vellas, und Yves Rolland. 2017. Physical Activity and Incident Chronic Diseases: A Longitudinal Observational Study in 16 European Countries. *American Journal of Preventive Medicine* 52 (3): 373–378.

Debus, Marc. 2016. Parlamentswahl in der Diktatur: Eine Analyse des Wählerverhaltens auf Ebene der Land- und Stadtkreise bei der Reichstagswahl vom 12. November 1933. *Zeitschrift für Parlamentsfragen* 47 (3): 693–708.

Dedrick, Robert F., John M. Ferron, Melinda R. Hess, Kristine Y. Hogarty, Jeffrey D. Kromrey, Thomas R. Lang, John D. Niles, und Reginald S. Lee. 2009. Multilevel Modeling: A Review of Methodological Issues and Applications. *Review of Educational Research* 79 (1): 69–102.

Deutsche Forschungsgemeinschaft. 1999. *Qualitätskriterien der Umfrageforschung.* Berlin: Akademie-Verlag.

Deutsche Forschungsgemeinschaft. 2013. *Sicherung guter wissenschaftlicher Praxis.* Weinheim: Wiley.

Deutsche Forschungsgemeinschaft. 2014. *Leitfaden für die Antragstellung. Projektanträge. DFG-Vordruck 54.01.* Bonn: Deutsche Forschungsgemeinschaft.

Deutscher Bundestag. 2002. *Bericht der Enquete-Kommission „Zukunft des Bürgerschaftlichen Engagements". Bürgerschaftliches Engagement: auf dem Weg in eine zukunftsfähige Bürgergesellschaft. Drucksache 14/8900.* Berlin: Deutscher Bundestag.

Deutscher Ethikrat. 2012. *Intersexualität.* Berlin: Deutscher Ethikrat.

Diaz-Bone, Rainer, und Christoph Weischer, Hrsg. 2015. *Methoden-Lexikon für die Sozialwissenschaften.* Wiesbaden: Springer VS.

Diekmann, Andreas. 1998. Die Bedeutung der Sekundäranalyse zur Kontrolle von Forschungsergebnissen. In *Statistik im Dienste der Öffentlichkeit,* Hrsg. Werner Haug, 135–142. Bern: Bundesamt für Statistik.

Diekmann, Andreas, Hrsg. 2006. *Methoden der Sozialforschung. Kölner Zeitschrift für Soziologie und Sozialpsychologie. Sonderheft 44/2004.* Wiesbaden: VS Verlag für Sozialwissenschaften.

Diekmann, Andreas. 2011. *Empirische Sozialforschung. Grundlagen, Methoden, Anwendungen.* Reinbek: Rowohlt.

Dillman, Don A., und Michelle L. Edwards. 2016. Designing a Mixed-Mode Survey. In *The Sage Handbook of Survey Methodology,* Hrsg. Christof Wolf, Dominique Joye, Tom E. C. Smith und Yang-Chih Fu, 255–268. London: SAGE.

Dillman, Don A., Jolene D. Smyth, und Leah Melani Christian. 2014. *Internet, Phone, Mail, and Mixed-Mode Surveys. The Tailored Design Method.* Hoboken: Wiley.

Döring, Herbert, und Christoph Hönnige. 2008. Parlament, Regierung, Staatsoberhaupt. In *Die EU-Staaten im Vergleich. Strukturen, Prozesse, Politikinhalte,* Hrsg. Oscar W. Gabriel und Sabine Kropp, 451–481. Wiesbaden: VS Verlag für Sozialwissenschaften.

Döring, Holger. 2013. The collective action of data collection: A data infrastructure on parties, elections and cabinets. *European Union Politics* 14 (1): 161–178.

Döring, Nicola, und Jürgen Bortz. 2016. *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler.* Heidelberg: Springer.

Dow, Jay K. 2009. Gender Differences in Political Knowledge: Distinguishing Characteristics-Based and Returns-Based Differences. *Political Behavior* 31 (1): 117–136.

Druckman, James N., Donald P. Green, James H. Kuklinski, und Arthur Lupia, Hrsg. 2011. *Cambridge Handbook of Experimental Political Science.* Cambridge: Cambridge University Press.

Easton, David. 1965. *A Systems Analysis of Political Life.* New York: Wiley.

Easton, David. 1975. A Re-Assessment of the Concept of Political Support. *British Journal of Political Science* 5 (4): 435–457.

Ebbinghaus, Bernhard. 2009. Vergleichende Politische Soziologie: Quantitative Analyse- oder qualitative Fallstudiendesigns? In *Politische Soziologie. Ein Studienbuch,* Hrsg. Viktoria Kaina und Andrea Römmele, 481–501. Wiesbaden: VS Verlag für Sozialwissenschaften.

Ebster, Claus, und Lieselotte Stalzer. 2017. *Wissenschaftliches Arbeiten für Wirtschafts- und Sozialwissenschaftler.* Stuttgart: UTB.

Eder, Christina. 2011. Direkte Demokratie als Allheilmittel? Möglichkeiten und Grenzen der Volksgesetzgebung. *Politische Psychologie* 1 (1): 83–97.

Eifler, Stefanie. 2014. Experiment. In *Handbuch Methoden der empirischen Sozialforschung,* Hrsg. Nina Baur und Jörg Blasius, 195–209. Wiesbaden: Springer VS.

Eifler, Stefanie, und Frank Faulbaum, Hrsg. 2017. *Methodische Probleme von Mixed-Mode-Ansätzen in der Umfrageforschung.* Wiesbaden: Springer VS.

Engel, Uwe, Hrsg. 2002. *Praxisrelevanz der Methodenausbildung.* Bonn: Informationszentrum Sozialwissenschaften.

Engel, Uwe, Ben Jann, Peter Lynn, Annette Scherpenzeel, und Patrick Sturgis, Hrsg. 2015. *Improving Survey Methods: Lessons from Recent Research.* New York: Routledge.

Engelhardt, Henriette. 2000. Untersuchungsdesigns in der Bevölkerungswissenschaft. In *Handbuch der Demographie 1. Modelle und Methoden,* Hrsg. Ulrich Mueller, Bernhard Nauck und Andreas Diekmann, 524–561. Berlin: Springer.

Esser, Hartmut. 1986. Über die Teilnahme an Befragungen. *ZUMA-Nachrichten* 18: 38–47.

Esser, Hartmut. 1987. Warum die Routine nicht weiterhilft – Überlegungen zur Kritik an der „Variablen-Soziologie". In *Problemlösungsoperator Sozialwissenschaft. Anwendungsorientierte Modelle der Sozial- und Planungswissenschaften in ihrer Wirksamkeitsproblematik,* Hrsg. Norbert Müller und Herbert Stachowiak, 230–245. Stuttgart: Ferdinand Enke.

Esser, Hartmut. 1999a. *Soziologie – Spezielle Grundlagen. Band 1: Situationslogik und Handeln.* Frankfurt: Campus.

Esser, Hartmut. 1999b. *Soziologie. Allgemeine Grundlagen.* Frankfurt: Campus.

Esser, Hartmut. 2000a. *Soziologie – Spezielle Grundlagen. Band 2: Die Konstruktion der Gesellschaft.* Frankfurt: Campus.

Esser, Hartmut. 2000b. *Soziologie – Spezielle Grundlagen. Band 4: Opportunitäten und Restriktionen.* Frankfurt: Campus.

Esser, Hartmut. 2008. The Two Meanings of Social Capital. In *The Handbook of Social Capital,* Hrsg. Dario Castiglione, Jan W. van Deth und Guglielmo Wolleb, 22–49. Oxford: Oxford University Press.

Esser, Hartmut, Klaus Klenovits, und Helmut Zehnpfennig. 1977. *Wissenschaftstheorie. Grundlagen und Analytische Wissenschaftstheorie.* Stuttgart: Teubner.

Faas, Thorsten. 2009. Das Experiment – ein unbekanntes Wesen? In *Datenwelten. Datenerhebung und Datenbestände in der Politikwissenschaft,* Hrsg. Kai-Uwe Schnapp, Nathalie Behnke und Joachim Behnke, 72–93. Baden-Baden: Nomos.

Faas, Thorsten. 2010. *Arbeitslosigkeit und Wählerverhalten. Direkte und indirekte Wirkungen auf Wahlbeteiligung und Parteipräferenzen in Ost- und Westdeutschland.* Baden-Baden: Nomos.

Faas, Thorsten. 2013. Wahlbeteiligung. In *Politik im Kontext: Ist alle Politik lokale Politik? Individuelle und kontextuelle Determinanten politischer Orientierungen,* Hrsg. Jan W. van Deth und Markus Tausendpfund, 413–440. Wiesbaden: Springer VS.

Faas, Thorsten, und Sascha Huber. 2010. Experimente in der Politikwissenschaft: Vom Mauerblümchen zum Mainstream. *Politische Vierteljahresschrift* 51 (4): 721–749.

Falter, Jürgen W. 1991. *Hitlers Wähler.* München: Beck.
Falter, Jürgen W. 2013. *Zur Soziographie des Nationalsozialismus. Studien zu den Wählern und Mitgliedern der NSDAP. HSR Supplement 25.* Köln: GESIS.
Falter, Jürgen W., und Uwe W. Gehring. 1998. Alter – ein neues Cleavage? In *Wahlen und Wähler. Analysen aus Anlaß der Bundestagswahl 1994,* Hrsg. Max Kaase und Hans-Dieter Klingemann, 463–503. Opladen: Westdeutscher Verlag.
Falter, Jürgen W., Andreas Link, Jan-Bernd Lohmöller, Johann de Rijke, und Siegfried Schumann. 1983. Arbeitslosigkeit und Nationalsozialismus. Eine empirische Analyse des Beitrags der Massenerwerbslosigkeit zu den Wahlerfolgen der NSDAP 1932 und 1933. *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 35 (3): 525–554.
Falter, Jürgen W., und Harald Schoen, Hrsg. 2014. *Handbuch Wahlforschung.* Wiesbaden: Springer VS.
Fatke, Matthias. 2015. Participation and Political Equality in Direct Democracy: Educative Effect or Social Bias. *Swiss Political Science Review* 21 (1): 99–118.
Faulbaum, Frank, Peter Prüfer, und Margrit Rexroth. 2009. *Was ist eine gute Frage? Die systematische Evaluation der Fragenqualität.* Wiesbaden: VS Verlag für Sozialwissenschaften.
Feist, Ursula. 1992. Niedrige Wahlbeteiligung – Normalisierung oder Krisensymptom der Demokratie in Deutschland? In *Protestwähler und Wahlverweigerer. Krise der Demokratie?,* Hrsg. Karl Starzacher, Konrad Schacht, Bernd Friedrich und Thomas Leif, 40–57. Köln: Bund-Verlag.
Fischer, Ernst Peter. 2003. *Die andere Bildung. Was man von den Naturwissenschaften wissen sollte.* München: Ullstein.
Fisher, Ronald A. 1951. *The design of experiments.* Edinburgh: Oliver & Boyd.
Fleck, Christian. 2013. Der Impact Faktor-Fetischismus. *Leviathan* 41 (4): 611–646.
Flick, Uwe. 2010. *Qualitative Sozialforschung. Eine Einführung.* Reinbek: Rowohlt.
Flick, Uwe, Ernst von Kardorff, und Ines Steinke, Hrsg. 2013. *Qualitative Forschung. Ein Handbuch.* Reinbek: Rowohlt.
Fogt, Helmut. 1982. *Politische Generationen.* Opladen: Westdeutscher Verlag.
Fox, Stuart. 2014. Is it Time to Update the Definition of Political Participation? *Parliamentary Affairs* 67 (2): 495–505.
Franklin, Mark N. 2001. How structural factors cause turnout variations at European Parliament Elections. *European Union Politics* 2 (3): 309–328.
Franzen, Axel. 2014. Antwortskalen in standardisierten Befragungen. In *Handbuch Methoden der empirischen Sozialforschung,* Hrsg. Nina Baur und Jörg Blasius, 701–711. Wiesbaden: Springer VS.
Franzen, Axel, und Markus Freitag. 2007a. Aktuelle Themen und Diskussionen der Sozialkapitalforschung. In *Sozialkapital. Grundlagen und Anwendungen,* Hrsg. Axel Franzen und Markus Freitag, 7–22. Wiesbaden: VS Verlag für Sozialwissenschaften.
Franzen, Axel, und Markus Freitag, Hrsg. 2007b. *Sozialkapital. Grundlagen und Anwendungen.* Wiesbaden: VS Verlag für Sozialwissenschaften.

Franzen, Axel, und Sonja Pointner. 2007. Sozialkapital: Konzeptualisierungen und Messungen. In *Sozialkapital. Grundlagen und Anwendungen,* Hrsg. Axel Franzen und Markus Freitag, 66–90. Wiesbaden: VS Verlag für Sozialwissenschaften.

Freitag, Markus. 1996. Wahlbeteiligung in westlichen Demokratien. Eine Analyse zur Erklärung von Niveauunterschieden. *Swiss Political Science Review* 2 (4): 101–134.

Frey, Bruno S., und Hannelore Weck. 1981. Hat Arbeitslosigkeit den Aufstieg des Nationalsozialismus bewirkt? *Jahrbücher für Nationalökonomie und Statistik* 196 (1): 1–31.

Friedrichs, Jürgen. 1990. *Methoden empirischer Sozialforschung.* Opladen: Westdeutscher Verlag.

Friedrichs, Jürgen, und Alexandra Nonnenmacher, Hrsg. 2014. *Soziale Kontexte und Soziale Mechanismen. Kölner Zeitschrift für Soziologie und Sozialpsychologie. Sonderheft 54/2014.* Wiesbaden: Springer VS.

Früh, Werner. 2017. *Inhaltsanalyse.* Stuttgart: UTB.

Fuchs, Dieter. 1989. *Die Unterstützung des politischen Systems der Bundesrepublik Deutschland.* Opladen: Westdeutscher Verlag.

Fuchs, Dieter. 2002a. Die politische Theorie der Systemanalyse: David Easton. In *Politische Theorien der Gegenwart I. Eine Einführung,* Hrsg. André Brodocz und Gary Schaal, 345–369. Opladen: Leske+Budrich.

Fuchs, Dieter. 2002b. Politikverdrossenheit. In *Handwörterbuch zur politischen Kultur der Bundesrepublik Deutschland,* Hrsg. Martin Greiffenhagen und Sylvia Greiffenhagen, 338–343. Wiesbaden: Westdeutscher Verlag.

Gabel, Matthew. 1998a. Public Support for European Integration: An Empirical Test of Five Theories. *The Journal of Politics* 60 (2): 333–354.

Gabel, Matthew J. 1998b. *Interests and Integration. Market Liberalization, Public Opinion, and European Union.* Ann Arbor: University of Michigan Press.

Gabel, Matthew, und Harvey D. Palmer. 1995. Understanding variation in public support for European Integration. *European Journal of Political Research* 27 (1): 3–19.

Gabler, Siegfried, und Sabine Häder. 1997. Überlegungen zu einem Stichprobendesign für Telefonumfragen in Deutschland. *ZUMA-Nachrichten* 21 (41): 7–18.

Gabler, Siegfried, und Sabine Häder. 2009. Die Kombination von Mobilfunk- und Festnetzstichproben in Deutschland. In *Umfrageforschung. Herausforderungen und Grenzen,* Hrsg. Martin Weichbold, Johann Bacher und Christof Wolf, 239–252. Wiesbaden: VS Verlag für Sozialwissenschaften.

Gabler, Siegfried, Sabine Häder, und Jürgen H. P. Hoffmeyer-Zlotnik, Hrsg. 1998. *Telefonstichproben in Deutschland.* Opladen: Westdeutscher Verlag.

Gabler, Siegfried, und Andreas Quatember. 2012. Das Problem mit der Repräsentativität von Stichprobenerhebungen. In *Jahrbuch 2012. Markt- und Sozialforschung,* Hrsg. Verband Schweizer Markt- und Sozialforschung, 17–19. Alpnach: Verband Schweizer Markt- und Sozialforschung.

Gabriel, Oscar. 2012. Wahlen in der Demokratie. In *Wählerverhalten in der Demokratie. Eine Einführung,* Hrsg. Oscar Gabriel und Bettina Westle, 13–42. Stuttgart: UTB.

Gabriel, Oscar W., und S. Isabell Thaidigsmann. 2009. Item Nonresponse: Ausprägung und Ursachen. In *Vom Interview zur Analyse. Methodische Aspekte der Einstellungs- und Wahlforschung,* Hrsg. Harald Schoen, Hans Rattinger und Oscar W. Gabriel, 283–320. Baden-Baden: Nomos.

Gehring, Uwe W., und Cornelia Weins. 2009. *Grundkurs Statistik für Politologen und Soziologen.* Wiesbaden: VS Verlag für Sozialwissenschaften.

Geißler, Rainer. 2014. *Die Sozialstruktur Deutschlands.* Wiesbaden: VS Verlag für Sozialwissenschaften.

Gerring, John. 2001. *Social Science Methodology. A Criterial Framework.* Cambridge: Cambridge University Press.

Gerring, John, und Craig W. Thomas. 2011. Quantitative versus Qualitative Methods. In *International Encyclopedia of Political Science. Volume 7,* Hrsg. Bertrand Badie, Dirk Berg-Schlosser und Leonardo Morlino, 2189–2196. Thousand Oaks: Sage.

Geys, Benny. 2006. Explaining voter turnout: A review of aggregate-level research. *Electoral Studies* 25 (4): 637–663.

Gherghina, Sergiu, und Alexia Katsanidou. 2013. Data Availability in Political Science Journals. *European Political Science* 12 (3): 333–349.

Giesselmann, Marco, und Michael Windzio. 2012. *Regressionsmodelle zur Analyse von Paneldaten.* Wiesbaden: Springer VS.

Giesselmann, Marco, und Michael Windzio. 2014. Paneldaten in der Soziologie: Fixed Effects Paradigma und empirische Praxis in Panelregression und Ereignisanalyse. *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 66 (1): 95–113.

Glemser, Axel, Gerd Meier, und Christiane Heckel. 2014. Dual-Frame: Stichprobendesign für CATI-Befragungen im mobilen Zeitalter. In *Stichproben-Verfahren in der Umfrageforschung. Eine Darstellung für die Praxis,* Hrsg. ADM Arbeitskreis Deutscher Markt- und Sozialforschungsinstitute e. V., 167–190. Wiesbaden: Springer VS.

Goertz, Gary. 2006. *Social Science Concepts. A User's Guide.* Princeton: Princeton University Press.

Göthlich, Stephan E. 2009. Zum Umgang mit fehlenden Daten in großzahligen empirischen Erhebungen. In *Methodik der empirischen Forschung,* Hrsg. Sänke Albers, Daniel Klapper, Udo Konradt, Achim Walter und Joachim Wolf, 119–135. Wiesbaden: Gabler.

Götzmann, Anke. 2015. *Entwicklung politischen Wissens in der Grundschule.* Wiesbaden: Springer VS.

Graeff, Peter. 2014. Aggregatdaten. In *Handbuch Methoden der empirischen Sozialforschung,* Hrsg. Nina Baur und Jörg Blasius, 915–924. Wiesbaden: Springer VS.

Graham, J. W. 2009. Missing Data Analysis: Making It Work in the Real World. In *Annual Review of Psychology,* 549–576. Palo Alto: Annual Reviews.

Green, Donald P., und Alan S. Gerber. 2003. The Underprovision of Experiments in Political Science. *The ANNALS of the American Academy of Political and Social Science* 589: 94–112.

Greiffenhagen, Sylvia. 2002. Bildung. In *Handwörterbuch zur politischen Kultur der Bundesrepublik Deutschland,* Hrsg. Martin Greiffenhagen und Sylvia Greiffenhagen, 52–57. Wiesbaden: Westdeutscher Verlag.

Groves, Robert M., Floyd J. Fowler, Mick P. Couper, James M. Lepkowski, Eleanor Singer, und Roger Tourangeau, Hrsg. 2009. *Survey Methodology.* Hoboken: Wiley.

Groves, Robert M., Eleanor Singer, und Amy Corning. 2000. Leverage-Saliency Theory of Survey Participation: Description and an Illustration. *Public Opinion Quarterly* 64 (3): 299–308.

Häder, Michael. 2015. *Empirische Sozialforschung. Eine Einführung.* Wiesbaden: Springer VS.

Häder, Michael, und Sabine Häder, Hrsg. 2009. *Telefonbefragungen über das Mobilfunknetz. Konzept, Design und Umsetzung einer Strategie zur Datenerhebung.* Wiesbaden: VS Verlag für Sozialwissenschaften.

Häder, Michael, und Sabine Häder. 2014. Stichprobenziehung in der quantitativen Sozialforschung. In *Handbuch Methoden der empirischen Sozialforschung,* Hrsg. Nina Baur und Jörg Blasius, 283–297. Wiesbaden: Springer VS.

Häder, Sabine, und Siegfried Gabler, Hrsg. 2002. *Telefonstichproben. Methodische Innovationen und Anwendungen in Deutschland.* Münster: Waxmann.

Häder, Sabine, Matthias Ganninger, und Siegfried Gabler. 2009. Die Stichprobenziehung für den European Social Survey: Prinzipien und Ergebnisse. In *Umfrageforschung. Herausforderungen und Grenzen,* Hrsg. Martin Weichbold, Johann Bacher und Christof Wolf, 181–193. Wiesbaden: VS Verlag für Sozialwissenschaften.

Hadjar, Andreas, und Michael Beck. 2010. Who does not participate in Elections in Europe and why is this? *European Societies* 12 (4): 521–542.

Hadjar, Andreas, und Rolf Becker. 2006. Bildungsexpansion und Wandel des politischen Interesses in Westdeutschland zwischen 1980 und 2002. *Politische Vierteljahresschrift* 47 (1): 12–34.

Hanmer, Michael J., Antoine J. Banks, und Ismail K. White. 2014. Experiments to Reduce the Over-Reporting of Voting: A Pipeline to the Truth. *Political Analysis* 22 (1): 130–141.

Hartmann, Martin. 2004. Vertrauen. In *Politische Theorie. 22 umkämpfte Begriffe zur Einführung,* Hrsg. Gerhard Göhler, Mattias Iser und Ina Kerner, 385–401. Wiesbaden: VS Verlag für Sozialwissenschaften.

Haug, Sonja. 1997. *Soziales Kapital. Ein kritischer Überblick über den aktuellen Forschungsstand.* Mannheim: Mannheimer Zentrum für Europäische Sozialforschung (Arbeitspapier Nr. 15).

Haunberger, Sigrid. 2011. *Teilnahmeverweigerung in Panelstudien.* Wiesbaden: VS Verlag für Sozialwissenschaften.

Heckel, Christiane, Axel Glemser, und Gerd Meier. 2014. Das ADM-Telefonstichproben-System. In *Stichproben-Verfahren in der Umfrageforschung. Eine Darstellung für die Praxis,* Hrsg. ADM Arbeitskreis Deutscher Markt- und Sozialforschungsinstitute e. V., 137–166. Wiesbaden: Springer VS.

Heckel, Christiane, und Oliver Hofmann. 2014. Das ADM-Stichproben-System (F2F) ab 1997. In *Stichproben-Verfahren in der Umfrageforschung. Eine Darstellung für die Praxis,* Hrsg. ADM Arbeitskreis Deutscher Markt- und Sozialforschungsinstitute e. V., 85–116. Wiesbaden: Springer VS.

Heinsohn, Till, und Markus Freitag. 2016. Vergleichende Regionenforschung in der Politikwissenschaft. In *Handbuch Vergleichende Politikwissenschaft,* Hrsg. Hans-Joachim Lauth, Marianne Kneuer und Gert Pickel, 385–397. Wiesbaden: Springer VS.

Heiser, Patrick. 2018. *Meilensteine der qualitativen Sozialforschung. Eine Einführung entlang klassischer Studien.* Wiesbaden: Springer VS.

Hempel, Carl G. 1977. *Aspekte wissenschaftlicher Erklärung.* Berlin: de Gruyter.

Hempel, Carl G., und Paul Oppenheim. 1948. Studies in the Logic of Explanation. *Philosophy of Science* 15 (2): 135–175.

Henrich, Joseph, Steven J. Heine, und Ara Norenzayan. 2010. The weirdest people in the world? *Behavioral and Brain Sciences* 33 (2-3): 61–83.

Hensel, Paul R. 2012. Review of Available Data Sets. In *Guide to the Scientific Study of International Processes,* Hrsg. Sara McLaughlin Mitchell, Paul F. Diehl und James D. Morrow, 43–62. Chichester: Wiley-Blackwell.

Hierlemann, Dominik, und Ulrich Sieberer. 2014. *Sichtbare Demokratie. Debatten und Fragestunden im Deutschen Bundestag.* Gütersloh: Bertelsmann Stiftung.

Hill, Austin Bradford. 1965. The Environment and Disease: Association or Causation? *Journal of the Royal Society of Medicine* 58 (5): 295–300.

Hill, Paul Bernhard, und Johannes Kopp, Hrsg. 2015. *Handbuch Familiensoziologie.* Wiesbaden: Springer VS.

Hoffmeyer-Zlotnik, Jürgen H. P., und Uwe Warner. 2014. Soziodemographische Standards. In *Handbuch Methoden der empirischen Sozialforschung,* Hrsg. Nina Baur und Jörg Blasius, 733–743. Wiesbaden: Springer VS.

Holtz-Bacha, Christina, Hrsg. 2010. *Die Massenmedien im Wahlkampf. Das Wahljahr 2009.* Wiesbaden: VS Verlag für Sozialwissenschaften.

Hooghe, Marc. 2014. Defining political participation: How to pinpoint an elusive target? *Acta Politica* 49 (3): 337–348.

Hoops, Christian, Alexander Glantz, und Tobias Michael. 2013. Innovationen im Rolling-Cross-Section-Design. Neue Ansätze zur Analyse des Wahlverhaltens. *Zeitschrift für Politikwissenschaft* 23 (2): 213–240.

Höpner, Martin, und Bojan Jurczyk. 2012. Kritik des Eurobarometers. Über die Verwischung der Grenze zwischen seriöser Demoskopie und interessengeleiteter Propaganda. *Leviathan* 40 (3): 326–349.

Höpner, Martin, und Bojan Jurczyk. 2015. How the Eurobarometer Blurs the Line between Research and Propaganda. Köln. Max-Planck-Institut für Gesellschaftsforschung (MPIfG Discussion Paper 15/6).

Hornig, Eike-Christian. 2011. Direkte Demokratie und Parteienwettbewerb – Überlegungen zu einem obligatorischen Referendum als Blockadelöser auf Bundesebene. *Zeitschrift für Parlamentsfragen* 42 (3): 475–492.

Hosch-Dayican, Bengü. 2014. Online political activities as emerging forms of political participation: How do they fit in the conceptual map? *Acta Politica* 49 (3): 337–348.

Hox, Joop J., Edith D. De Leeuw, und Eva A. O. Zijlmans. 2015. Measurement equivalence in mixed modes surveys. *Frontiers in Psychology* 6: 1–11.

Hox, Joop J., und J. Kyle Roberts. 2011. *Handbook of Advanced Multilevel Analysis.* New York: Routledge.

Huber, Oswald. 2013. *Das psychologische Experiment. Eine Einführung.* Bern: Huber.

Huemer, Birgit, Markus Rheindorf, und Helmut Gruber. 2012. *Abstract, Exposé und Förderantrag.* Stuttgart: UTB.

Hüfken, Volker. 2014. Telefonische Befragung. In *Handbuch Methoden der empirischen Sozialforschung,* Hrsg. Nina Baur und Jörg Blasius, 631–641. Wiesbaden: Springer VS.

Huinink, Johannes, Josef Brüderl, Bernhard Nauck, Sabine Walper, Laura Castiglioni, und Michael Feldhaus. 2011. Panel Analysis of Intimate Relationships and Family Dynamics (pairfam): Conceptual framework and design. *Zeitschrift für Familienforschung* 23 (1): 77–100.

Hummell, Hans J. 1972. *Probleme der Mehrebenenanalyse.* Stuttgart: Teubner.

Hunsicker, Stefan, und Yvonne Schroth. 2007. Die Kombination von Mobilfunk- und Festnetzstichproben. Eine praktische Anwendung des Dual-Frame-Ansatzes. *Methoden – Daten – Analysen* 1 (2): 161–182.

Hunsicker, Stefan, und Yvonne Schroth. 2014. Dual-Frame-Ansatz in politischen Umfragen. Mannheim. Forschungsgruppe Wahlen e. V. (Arbeitspapier Nr. 2).

Huntington, Samuel P. 1996. *The clash of civilizations and the remaking of world order.* New York: Simon & Schuster.

Hyde, Janet Shibley. 2005. The Gender Similarities Hypothesis. *American Psychologist* 60 (6): 581–592.

Immerfall, Stefan, und Andreas Sobisch. 1997. Europäische Integration und europäische Identität: Die Europäische Union im Bewußtsein ihrer Bürger. *Aus Politik und Zeitgeschichte* 47 (10): 25–37.

Inglehart, Ronald. 1970. Cognitive Mobilization and European Identity. *Comparative Politics* 3 (1): 45–70.

Inglehart, Ronald. 1971. Changing Value Priorities and European Integration. *Journal of Common Market Studies* 10 (1): 1–36.

Inglehart, Ronald. 1977. *The Silent Revolution. Changing Values and Political Styles Among Western Publics.* Princeton: Princeton University Press.

Inglehart, Ronald. 1995. *Kultureller Umbruch. Wertwandel in der westlichen Welt.* Frankfurt: Campus.

Inglehart, Ronald. 1998. *Modernisierung und Postmodernisierung. Kultureller, wirtschaftlicher und politischer Wandel in 43 Gesellschaften.* Frankfurt: Campus.

Inglehart, Ronald. 2007. Postmaterialist Values and the Shift from Survival to Self-Expression Values. In *The Oxford Handbook of Political Behavior,* Hrsg. Russell J. Dalton und Hans-Dieter Klingemann, 223–239. Oxford: Oxford University Press.

Inglehart, Ronald, und Paul R. Abramson. 1999. Measuring Postmaterialism. *American Political Science Review* 93 (3): 665–677.

Inglehart, Ronald, und Hans-Dieter Klingemann. 1996. Dimensionen des Wertewandels. Theoretische und methodische Reflexionen anläßlich einer neuerlichen Kritik. *Politische Vierteljahresschrift* 37 (2): 319–340.

Inglehart, Ronald, und Jacques-René Rabier. 1978. Economic Uncertainty and European Solidarity: Public Opinion Trends. *The Annals of the American Academy of Political and Social Science* 440: 66–97.

Jäckle, Sebastian. 2015. Experimente. In *Methodologie, Methoden, Forschungsdesign. Ein Lehrbuch für fortgeschrittene Studierende der Politikwissenschaft,* Hrsg. Achim Hildebrandt, Sebastian Jäckle, Frieder Wolf und Andreas Heindl, 13–35. Wiesbaden: Springer VS.

Jahn, Detlef. 2009. Die Aggregatdatenanalyse in der vergleichenden Politikwissenschaft. In *Methoden der vergleichenden Politik- und Sozialwissenschaft. Neue Entwicklungen und Anwendungen,* Hrsg. Susanne Pickel, Detlef Jahn, Hans-Joachim Lauth und Gert Pickel, 173–196. Wiesbaden: VS Verlag für Sozialwissenschaften.

Jahn, Detlef. 2011. *Vergleichende Politikwissenschaft.* Wiesbaden: VS Verlag für Sozialwissenschaften.

Jahoda, Marie, Paul Felix Lazarsfeld, und Hans Zeisel. 1975 [1933]. *Die Arbeitslosen von Marienthal. Ein soziographischer Versuch über die Wirkungen langandauernder Arbeitslosigkeit.* Frankfurt: Suhrkamp.

Kaase, Max. 1995. Partizipation. In *Wörterbuch Staat und Politik,* Hrsg. Dieter Nohlen, 521–527. Bonn: Bundeszentrale für politische Bildung.

Kailitz, Steffen, Hrsg. 2007. *Schlüsselwerke der Politikwissenschaft.* Wiesbaden: VS Verlag für Sozialwissenschaften.

Kaina, Viktoria. 2009. *Wir in Europa. Kollektive Identität und Demokratie in der Europäischen Union.* Wiesbaden: VS Verlag für Sozialwissenschaften.

Kallus, K. Wolfgang. 2016. *Erstellung von Fragebogen.* Stuttgart: UTB.

Karp, Jeffrey A., Susan A. Banducci, und Shaun Bowler. 2003. To know it is to love it? Satisfaction with democracy in the European Union. *Comparative Political Studies* 36 (3): 271–292.

Kaspar, Hanna, und Jürgen W. Falter. 2007. Entstehung neuer Konfliktlinien: Geschlechterkonflikt und Alterskonflikt? In *Der gesamtdeutsche Wähler. Stabilität und Wandel des Wählerverhaltens im wiedervereinigten Deutschland,* Hrsg. Hans Rattinger, Oscar W. Gabriel und Jürgen W. Falter, 113–140. Baden-Baden: Nomos.

Keil, Silke I. 2009. Die Datengrundlage der Politischen Soziologie in Forschung und Lehre. In *Politische Soziologie. Ein Studienbuch,* Hrsg. Viktoria Kaina und Andrea Römmele, 421–445. Wiesbaden: VS Verlag für Sozialwissenschaften.

Keil, Silke I., und Jan W. van Deth, Hrsg. 2012. *Deutschlands Metamorphosen. Ergebnisse des European Social Survey 2002 bis 2008.* Baden-Baden: Nomos.

Kelle, Udo. 2014. Mixed Methods. In *Handbuch Methoden der empirischen Sozialforschung,* Hrsg. Nina Baur und Jörg Blasius, 153–166. Wiesbaden: Springer VS.

Keuschnigg, Marc, und Tobias Wolbring, Hrsg. 2015. *Experimente in den Sozialwissenschaften. Soziale Welt – Sonderband 22.* Baden-Baden: Nomos.

Kielmansegg, Peter Graf. 2006. Über direkte Demokratie – sechs Anmerkungen zu einer unbefriedigenden Debatte. In *Jahrbuch Extremismus & Demokratie,* Hrsg. Uwe Backes und Eckhard Jesse, 57–80. Baden-Baden: Nomos.

King, Gary. 1995. Replication, Replication. *PS: Political Science & Politics* 28 (3): 444–452.

King, Gary, James Honaker, Anne Joseph, und Kenneth Scheve. 2001. Analyzing incomplete political science data: An alternative algorithm for multiple imputation. *American Political Science Review* 95 (1): 49–69.

King, Gary, Robert O. Keohane, und Sidney Verba. 1994. *Designing Social Inquiry. Scientific Inference in Qualitative Research.* Princeton: Princeton University Press.

Kish, Leslie. 1949. A Procedure for Objective Respondent Selection within the Household. *Journal of the American Statistical Association* 44 (247): 380–387.

Klein, Markus. 2009. Die Entwicklung der grünen Wählerschaft im Laufe dreier Jahrzehnte – eine empirische APK-Analyse. In *Politik – Wissenschaft – Medien. Festschrift für Jürgen W. Falter zum 65. Geburtstag,* Hrsg. Hanna Kaspar, Harald Schoen, Siegfried Schumann und Jürgen R. Winkler, 391–401. Wiesbaden: VS Verlag für Sozialwissenschaften.

Klingemann, Hans-Dieter, Andrea Volkens, Judith L. Bara, Ian Budge, und Michael McDonald. 2006. *Mapping policy preferences II. Estimates for parties, electors, and governments in Eastern Europe, European Union, and OECD 1990–2003.* Oxford: Oxford University Press.

Klingemann, Hans-Dieter, und Christian Welzel. 2002. Ideologie. In *Handwörterbuch zur politischen Kultur der Bundesrepublik Deutschland,* Hrsg. Martin Greiffenhagen und Sylvia Greiffenhagen, 200–203. Wiesbaden: Westdeutscher Verlag.

Klöckner, Jennifer, und Jürgen Friedrichs. 2014. Gesamtgestaltung des Fragebogens. In *Handbuch Methoden der empirischen Sozialforschung,* Hrsg. Nina Baur und Jörg Blasius, 675–685. Wiesbaden: Springer VS.

Kluckhohn, Clyde. 1967. Values and Value-Orientations in the Theory of Action: An Exploration in Definition and Classification. In *Toward a General Theory of Action,* Hrsg. Talcott Parsons und Edward A. Shils, 388–433. Cambridge: Harvard University Press.

Koch, Achim. 2002. 20 Jahre Feldarbeit im ALLBUS: Ein Blick in die Blackbox. *ZUMA-Nachrichten* 26: 9–37.

Koch, Achim, und Martina Wasmer. 2004. Der ALLBUS als Instrument zur Untersuchung sozialen Wandels: Eine Zwischenbilanz nach 20 Jahren. In *Sozialer und politischer Wandel in Deutschland. Analysen mit ALLBUS-Daten aus zwei Jahrzehnten,* Hrsg. Rüdiger Schmitt-Beck, Martina Wasmer und Achim Koch, 13–41. Wiesbaden: VS Verlag für Sozialwissenschaften.

Koch, Wolfgang, und Beate Frees. 2017. ARD/ZDF-Onlinestudie 2017: Neun von zehn Deutschen online. *Media Perspektiven* (9): 434–446.

Kohler, Ulrich. 2008. Assessing the Quality of European Surveys. Towards an Open Method of Coordination for Survey Data. In *Handbook of Quality of Life in the Enlarged European Union,* Hrsg. Jens Alber, Tony Fahey und Chiara Saraceno, 405–423. London: Routledge.

Kohler, Ulrich. 2016. Anforderungen an Hochschulabsolventen. Oder: was Mitarbeiter in einem empirisch ausgerichteten Forschungsprojekt können sollten? In *Human Resources. Qualitätsaspekte der Ausbildung in der empirischen Forschung,* Hrsg. Christian König, Matthias Stahl und Erich Wiegand, 43–77. Wiesbaden: Springer VS.

Kohler, Ulrich, und Frauke Kreuter. 2017. *Datenanalyse mit Stata. Allgemeine Konzepte der Datenanalyse und ihre praktische Anwendung.* Berlin: de Gruyter.

König, René. 1974. Praktische Sozialforschung. In *Das Interview. Formen. Technik. Auswertung,* Hrsg. René König, 13–33. Köln: Kiepenheuer & Witsch.

Kopp, Johannes, und Daniel Lois. 2014. *Sozialwissenschaftliche Datenanalyse. Eine Einführung.* Wiesbaden: Springer VS.

Kornberg, Allan, und Harold D. Clarke. 2011. *Citizens and Community. Political Support in a Representative Democracy.* Cambridge: Cambridge University Press.

Krämer, Walter. 1994. *Statistik verstehen. Eine Gebrauchsanweisung.* Frankfurt: Campus.

Krämer, Walter. 2009. *So lügt man mit Statistik.* München: Piper.

Krebs, Dagmar, und Jürgen H. P. Hoffmeyer-Zlotnik. 2010. Positive First or Negative First? Effects of the Order of Answering Categories on Response Behavior. *Methodology* 6 (3): 118–127.

Krebs, Dagmar, und Natalja Menold. 2014. Gütekriterien quantitativer Sozialforschung. In *Handbuch Methoden der empirischen Sozialforschung,* Hrsg. Nina Baur und Jörg Blasius, 425–438. Wiesbaden: Springer VS.

Kriesi, Hanspeter. 2007. Sozialkapital. Eine Einführung. In *Sozialkapital. Grundlagen und Anwendungen,* Hrsg. Axel Franzen und Markus Freitag, 23–46. Wiesbaden: VS Verlag für Sozialwissenschaften.

Krimmel, Iris. 2000. Politische Beteiligung in Deutschland – Strukturen und Erklärungsfaktoren. In *Wirklich ein Volk? Die politischen Orientierungen von Ost- und Westdeutschen im Vergleich,* Hrsg. Jürgen Falter, Oscar W. Gabriel und Hans Rattinger, 609–639. Opladen: Leske+Budrich.

Krippendorff, Klaus. 2013. *Content Analysis. An Introduction to Its Methodology.* Thousand Oaks: Sage.

Kromrey, Helmut, Jochen Roose, und Jörg Strübing. 2016. *Empirische Sozialforschung.* Stuttgart: UTB.

Krosnick, Jon A. 1991. Response Strategies for Coping with the Cognitive Demands of Attitude Measures in Surveys. *Applied Cognitive Psychology* 5 (3): 213–236.

Krosnick, Jon A. 1999. Survey Research. *Annual Review of Psychology* 50: 537–567.

Krosnick, Jon, und Leandre R. Fabrigar. 1997. Designing Rating Scales for Effective Measurement in Surveys. In *Survey Measurement and Process Quality,* Hrsg. Lars E. Lyberg, Paul P. Biemer, Martin Collins, Edith D. De Leeuw, Cathryn Dippo, Norbert Schwarz und Dennis Trewin, 141–164. New York: Wiley.

Krumm, Thomas. 2009. Grundgedanken und Grundelemente qualitativer Forschung. In *Methoden der Politikwissenschaft,* Hrsg. Bettina Westle, 96–109. Baden-Baden: Nomos.

Krumm, Thomas, und Bettina Westle. 2009. Der Forschungsprozess im Überblick. In *Methoden der Politikwissenschaft,* Hrsg. Bettina Westle, 115–123. Baden-Baden: Nomos.

Kruskal, William, und Frederick Mosteller. 1979a. Representative Sampling, I: Non-Scientific Literature. *International Statistical Review/Revue Internationale de Statistique* 47 (1): 13–24.

Kruskal, William, und Frederick Mosteller. 1979b. Representative Sampling, II: Scientific Literature, Excluding Statistics. *International Statistical Review/Revue Internationale de Statistique* 47 (2): 111–127.

Kruskal, William, und Frederick Mosteller. 1979c. Representative Sampling, III: The Current Statistical Literature. *International Statistical Review/Revue Internationale de Statistique* 47 (3): 245–265.

Kubbe, Ina. 2016. *Experimente in der Politikwissenschaft. Eine methodische Einführung.* Wiesbaden: Springer VS.

Kuckartz, Udo. 2009. Inhaltsanalyse. In *Methoden der Politikwissenschaft,* Hrsg. Bettina Westle, 334–344. Baden-Baden: Nomos.

Kuckartz, Udo. 2014. *Mixed Methods. Methodologie, Forschungsdesigns und Analyseverfahren.* Wiesbaden: Springer VS.

Kühne, Rinaldo. 2013. Konzeptspezifikation und Messung. In *Handbuch standardisierte Erhebungsverfahren in der Kommunikationswissenschaft,* Hrsg. Wiebke Möhring und Daniela Schlütz, 23–40. Wiesbaden: Springer VS.

Kühnel, Steffen, und André Dingelstedt. 2014. Kausalität. In *Handbuch Methoden der empirischen Sozialforschung,* Hrsg. Nina Baur und Jörg Blasius, 1017–1028. Wiesbaden: Springer VS.

Kunter, Mareike, Gundel Schümer, Cordula Artelt, Jürgen Baumert, Eckard Klieme, Michael Neubrand, Manfred Prenzel, Ulrich Schiefele, Wolfang Schneider, Petra Stanat, Klaus-Jürgen Tillmann, und Manfred Weiß. 2002. *PISA 2000: Dokumentation der Erhebungsinstrumente.* Berlin: Max-Planck-Institut für Bildungsforschung.

Lamers, Patrick, und Sigrid Roßteutscher. 2014. Die Wahlbeteiligung. In *Zwischen Fragmentierung und Konzentration: Die Bundestagswahl 2013,* Hrsg. Rüdiger Schmitt-Beck, Hans Rattinger, Sigrid Roßteutscher, Bernhard Weßels und Christof Wolf, 119–131. Baden-Baden: Nomos.

Lamnek, Siegfried. 2010. *Qualitative Sozialforschung.* Weinheim: Beltz.

LaPiere, Richard T. 1934. Attitudes vs. actions. *Social Forces* 13 (2): 230–237.

Lauth, Hans-Joachim. 2008. Demokratieentwicklung und demokratische Qualität. In *Die EU-Staaten im Vergleich. Strukturen, Prozesse, Politikinhalte,* Hrsg. Oscar W. Gabriel und Sabine Kropp, 33–61. Wiesbaden: VS Verlag für Sozialwissenschaften.

Lauth, Hans-Joachim, Gert Pickel, und Susanne Pickel. 2015. *Methoden der vergleichenden Politikwissenschaft. Eine Einführung.* Wiesbaden: Springer VS.

Lazarsfeld, Paul F., Bernard Berelson, und Hazel Gaudet. 1968. *The People's Choice. How the Voter makes up his mind in a Presidential Campaign.* New York: Columbia University Press.

Lehnert, Matthias, Bernhard Miller, und Arndt Wonka. 2007. Na Und? Überlegungen zur theoretischen und gesellschaftlichen Relevanz in der Politikwissenschaft. In *Forschungsdesign in der Politikwissenschaft. Probleme – Strategien – Anwendungen,* Hrsg. Thomas Gschwend und Frank Schimmelfennig, 39–60. Frankfurt: Campus.

Lehrer, David, Janine Leschke, Stefan Ihachimi, Ana Vasilio, und Brigitte Weiffen. 2007. Negative results in social science. *European Political Science* 6 (1): 51–68.

Lenzner, Timo, Cornelia Neuert, und Wanda Otto. 2015. *Kognitives Pretesting.* Mannheim: GESIS.

Lijphart, Arend. 1971. Comparative Politics and the Comparative Method. *American Political Science Review* 65 (3): 682–693.

Lijphart, Arend. 1975. The Comparable-Cases Strategy in Comparative Research. *Comparative Political Studies* 8 (2): 158–177.

Lijphart, Arend. 1999. *Patterns of Democray. Government Forms and Performance in Thirty-Six Countries.* New Haven: Yale University Press.

Lippe, Peter von der, und Andreas Kladroba. 2002. Repräsentativität von Stichproben. *Marketing. Zeitschrift für Forschung und Praxis* 24 (2): 139–146.

Lipset, Seymour Martin, Martin A. Trow, und James Samuel Coleman. 1956. *Union Democracy. The internal politics of the International Typographical Union.* Glencoe: Free Press.

Loer, Kathrin, Renate Reiter, und Annette Elisabeth Töller. 2015. Was ist ein Politikfeld und warum entsteht es? *dms – der moderne staat* 8 (1): 7–28.

Löffler, Ute, und Christian von der Heyde. 2014. Ein Kurz-Überblick über die gebräuchlichsten Auswahl-Verfahren in der Marktforschung. In *Stichproben-Verfahren in der Umfrageforschung. Eine Darstellung für die Praxis,* Hrsg. ADM Arbeitskreis Deutscher Markt- und Sozialforschungsinstitute e. V., 19–24. Wiesbaden: Springer VS.

Lois, Daniel. 2013. *Wenn das Leben religiös macht. Altersabhängige Veränderungen der kirchlichen Religiosität im Lebensverlauf.* Wiesbaden: Springer VS.

Lück, Detlev, und Nina Baur. 2011. Wie kommen die Daten in den Datensatz? Arbeitsschritte vom Fragebogen zum fertigen Datensatz. In *Datenanalyse mit SPSS für Fortgeschrittene 1,* Hrsg. Leila Akremi, Nina Baur und Sabine Fromm, 22–58. Wiesbaden: VS Verlag für Sozialwissenschaften.

Lück, Detlev, und Uta Landrock. 2014. Datenaufbereitung und Datenbereinigung in der quantitativen Sozialforschung. In *Handbuch Methoden der empirischen Sozialforschung,* Hrsg. Nina Baur und Jörg Blasius, 397–409. Wiesbaden: Springer VS.

Lynn, Peter. 2009. The Problem of Nonresponse. In *International Handbook of Survey Methodology,* Hrsg. Edith D. de Leeuw, J. Joop Hox und Don A. Dillman, 35–55. New York: Psychology Press.

Maier, Jürgen. 2000. *Politikverdrossenheit in der Bundesrepublik Deutschland. Dimensionen – Determinanten – Konsequenzen.* Opladen: Leske+Budrich.

Mallock, Wolfgang, Udo Riege, und Matthias Stahl. 2016. *Informationsressourcen für die Sozialwissenschaften. Datenbanken – Längsschnittuntersuchungen – Portale – Institutionen.* Wiesbaden: Springer VS.

Mannheim, Karl. 1928. Das Problem der Generationen. *Kölner Vierteljahrshefte für Soziologie 7*: 157–185; 309–330.

Marcus, Jan. 2009. Der Einfluss von Erhebungsformen auf den Postmaterialismus-Index. *Methoden – Daten – Analysen* 3 (2): 137–166.

Mayring, Philipp. 2002. *Einführung in die qualitative Sozialforschung.* Weinheim: Beltz.

McDermott, Rose. 2011. Internal and External Validity. In *Cambridge Handbook of Experimental Political Science,* Hrsg. James N. Druckman, Donald P. Green, James H. Kuklinski und Arthur Lupia, 27–40. Cambridge: Cambridge University Press.

McLaren, Lauren M. 2006. *Identity, interests and attitudes to European integration.* Houndmills: Palgrave Macmillan.

Meier, Gerd, Axel Glemser, und Christiane Heckel. 2014. Random-Telefonstichproben. In *Stichproben-Verfahren in der Umfrageforschung. Eine Darstellung für die Praxis,* Hrsg. ADM Arbeitskreis Deutscher Markt- und Sozialforschungsinstitute e. V., 117–135. Wiesbaden: Springer VS.

Meier, Gerd, und Jochen Hansen. 2014. Quotenverfahren. In *Stichproben-Verfahren in der Umfrageforschung. Eine Darstellung für die Praxis,* Hrsg. ADM Arbeitskreis Deutscher Markt- und Sozialforschungsinstitute e. V., 197–205. Wiesbaden: Springer VS.

Meier, Gerd, Michael Schneid, Yvonne Stegemann, und Angelika Stiegler. 2005. Steigerung der Ausschöpfungsquote von Telefonumfragen durch geschickte Einleitungstexte. *ZUMA-Nachrichten* 57: 37–55.

Merkel, Wolfang. 2015. *Nur schöner Schein? Demokratische Innovationen in Theorie und Praxis.* Frankfurt: Otto Brenner Stiftung.

Merkel, Wolfgang. 2011. Volksabstimmungen: Illusion und Realität. *Aus Politik und Zeitgeschichte* 61 (44-45): 47–55.

Merkel, Wolfgang, und Alexander Petring. 2012. Politische Partizipation und demokratische Inklusion. In *Demokratie in Deutschland. Zustand – Herausforderungen – Perspektiven,* Hrsg. Tobias Mörschel und Christian Krell, 93–119. Wiesbaden: Springer VS.

Merkel, Wolfgang, und Claudia Ritzi, Hrsg. 2017. *Die Legitimität direkter Demokratie. Wie demokratisch sind Volksabstimmungen?* Wiesbaden: Springer VS.

Meulemann, Heiner, Hrsg. 2008. *Social Capital in Europe: Similarity of Countries and Diversity of People? Multi-Level analyses of the European Social Survey 2002.* Leiden: Brill.

Michels, Robert. 1989 [1911]. *Zur Soziologie des Parteiwesens in der modernen Demokratie. Untersuchungen über die oligarchischen Tendenzen des Gruppenlebens.* Stuttgart: Kröner.

Milgram, Stanley, Leon Mann, und Susan Harter. 1965. The lost-letter Technnique: A Tool of social research. *Public Opinion Quarterly* 29 (3): 437–438.

Miller, Bernhard. 2007. Maßvoll Messen: Zur konzeptorientierten Entwicklung von Messinstrumenten. In *Forschungsdesign in der Politikwissenschaft. Probleme – Strategien – Anwendungen,* Hrsg. Thomas Gschwend und Frank Schimmelfennig, 123–148. Frankfurt: Campus.

Miller, George A. 1956. The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological Review* 63 (2): 81–97.

Mochmann, Ekkehard. 2014. Quantitative Daten für die Sekundäranalyse. In *Handbuch Methoden der empirischen Sozialforschung,* Hrsg. Nina Baur und Jörg Blasius, 233–244. Wiesbaden: Springer VS.

Morton, Rebecca B., und Joshua A. Tucker. 2014. Welcome to JEPS! *Journal of Experimental Political Science* 1 (1): 1–5.

Müller, Thomas, und Susanne Pickel. 2007. Wie lässt sich Demokratie am besten messen? Zur Konzeptqualität von Demokratie-Indizes. *Politische Vierteljahresschrift* 48 (3): 511–539.

Müller, Walter, und Reinhard Pollak. 2010. Weshalb gibt es so wenige Arbeiterkinder in Deutschlands Universitäten? In *Bildung als Privileg. Erklärungen und Befunde zu den Ursachen der Bildungsungleichheit,* Hrsg. Rolf Becker und Wolfgang Lauterbach, 303–342. VS Verlag für Sozialwissenschaften.

Neller, Katja. 2004. Politik und Lebenszufriedenheit. In *Deutschland in Europa. Ergebnisse des European Social Survey 2002–2003,* Hrsg. Jan W. van Deth, 27–53. Wiesbaden: VS Verlag für Sozialwissenschaften.

Neller, Katja. 2005. Kooperation und Verweigerung: Eine Non-Response Studie. *ZUMA-Nachrichten* 29 (57): 9–36.

Neundorf, Anja. 2012. Die Links-Rechts-Dimension auf dem Prüfstand: Ideologisches Wählen in Ost- und Westdeutschland 1990 bis 2008. In *Wählen in Deutschland. Sonderheft der Politischen Vierteljahresschrift 45/2011,* Hrsg. Rüdiger Schmitt-Beck, 227–250. Baden-Baden: Nomos.

Niedermayer, Oskar. 2005. *Bürger und Politik. Politische Orientierungen und Verhaltensweisen der Deutschen.* Wiesbaden: VS Verlag für Sozialwissenschaften.

Niedermayer, Oskar. 2008. Plädoyer für die Abschaffung der Links-Rechts-Dimension. *Neue Gesellschaft. Frankfurter Hefte* (5): 32–35.

Nieland, Jörg-Uwe. 2011. Nebenschauplätze des Wahlkampfs: Beobachtungen zur Inszenierung und Wahrnehmung von Europawahltagen 2009. In *30 Jahre Direktwahlen zum Europäischen Parlament (1979–2009),* Hrsg. Jürgen Mittag, 271–287. Baden-Baden: Nomos.

Noelle-Neumann, Elisabeth, und Thomas Petersen. 2005. *Alle, nicht jeder. Einführung in die Methoden der Demoskopie.* Heidelberg: Springer.

Norris, Pippa. 1999. Introduction: The Growth of Critical Citizens? In *Critical Citizens. Global Support for Democratic Government,* Hrsg. Pippa Norris, 1–27. Oxford: Oxford University Press.

Norris, Pippa. 2004. *Electoral Engineering. Voting Rules and Political Behavior.* Cambridge: Cambridge University Press.

Norris, Pippa. 2011. *Democratic Deficit. Critical Citizens Revisited.* Cambridge: Cambridge University Press.

Norris, Pippa. 2017. The conceptual framework of political support. In *Handbook on Political Trust,* Hrsg. Sonja Zmerli und Tom W. G. van der Meer, 19–32. Cheltenham: Edward Elgar.

Ohr, Dieter. 2010. Lineare Regression: Modellannahmen und Regressionsdiagnostik. In *Handbuch der sozialwissenschaftlichen Datenanalyse,* Hrsg. Christof Wolf und Henning Best, 639–676. Wiesbaden: VS Verlag für Sozialwissenschaften.

Opp, Karl-Dieter. 2014. *Methodologie der Sozialwissenschaften. Einführung in Probleme ihrer Theorienbildung und praktischen Anwendung.* Wiesbaden: Springer VS.

Papcke, Sven, und Georg W. Oesterdiekhoff, Hrsg. 2001. *Schlüsselwerke der Soziologie.* Wiesbaden: Westdeutscher Verlag.

Pappi, Franz Urban, Nicole Michaela Seher, und Anna-Sophie Kurella. 2011. Das Politikangebot deutscher Parteien bei den Bundestagswahlen seit 1976 im dimensionsweisen Vergleich. Gesamtskala und politikfeldspezifische Skalen. Mannheim. Mannheimer Zentrum für Europäische Sozialforschung (Arbeitspapier Nr. 142).

Pausch, Markus. 2009. Eurobarometer und die Konstruktion eines europäischen Bewusstseins. In *Umfrageforschung. Herausforderungen und Grenzen,* Hrsg. Martin Weichbold, Johann Bacher und Christof Wolf, 539–552. Wiesbaden: VS Verlag für Sozialwissenschaften.

Payne, Stanley L. 1958. *The Art of Asking Questions.* Princeton: Princeton University Press.

Pforr, Klaus, Michael Blohm, Annelies G. Blom, Barbara Erdel, Barbara Felderer, Mathis Fräßdorf, Kristin Hajek, Susanne Helmschrott, Corinna Kleinert, Achim Koch, Ulrich Krieger, Martin Kroh, Silke Martin, Denise Saßenroth, Claudia Schmiedeberg, Eva-Maria Trüdinger, und Beatrice Rammstedt. 2015. Are incentive effects on response rates and nonresponse bias in large-scale, face-to-face surveys generalizable to Germany? Evidence from ten experiments. *Public Opinion Quarterly* 79 (3): 740–768.

Pickel, Susanne. 2016. Methodologische Grundlagen des Vergleichs und Vergleichsdesigns. In *Handbuch Vergleichende Politikwissenschaft,* Hrsg. Hans-Joachim Lauth, Marianne Kneuer und Gert Pickel, 25–45. Wiesbaden: Springer VS.

Pickel, Susanne, und Gert Pickel. 2012. Die Messung von Indizes in der Vergleichenden Politikwissenschaft – methodologische Spitzfindigkeit oder substantielle Notwendigkeit. *Zeitschrift für Vergleichende Politikwissenschaft* 6 (2): 1–17.

Plischke, Thomas. 2014. *Wann Wähler entscheiden. Abläufe von Entscheidungsprozessen und der Zeitpunkt der Wahlentscheidung.* Baden-Baden: Nomos.

Plümper, Thomas. 2012. *Effizient schreiben. Leitfaden zum Verfassen von Qualifizierungsarbeiten und wissenschaftlichen Texten.* München: Oldenbourg.

Plümper, Thomas, und Frank Schimmelfennig. 2007. Wer wird Prof – und wann? Berufungsdeterminanten in der deutschen Politikwissenschaft. *Politische Vierteljahresschrift* 48 (1): 97–117.

Polk, Jonathan, Jan Rovny, Ryan Bakker, Erica Edwards, Liesbet Hooghe, Seth Jolly, Jelle Koedam, Filip Kostelka, Gary Marks, Gijs Schumacher, Marco Steenbergen, Milada Vachudova, und Marko Zilovic. 2017. Explaining the salience of anti-elitism and reducing political corruption for political parties in Europe with the 2014 Chapel Hill Expert Survey data. *Research & Politics* 4 (1): 1–9.

Popper, Karl R. 1969. Die Logik der Sozialwissenschaften. In *Der Positivismusstreit in der deutschen Soziologie,* Hrsg. Theodor W. Adorno, Hans Albert, Ralf Dahrendorf, Jürgen Habermas, Harald Pilot und Karl R. Popper, 103–123. Neuwied: Luchterhand.

Popper, Karl R. 1982 [1934]. *Logik der Forschung.* Tübingen: Mohr.

Porst, Rolf. 2001. *Wie man die Rücklaufquote bei postalischen Befragungen erhöht.* Mannheim: Zentrum für Umfragen, Methoden und Analysen.

Porst, Rolf. 2014a. *Fragebogen. Ein Arbeitsbuch.* Wiesbaden: VS Verlag für Sozialwissenschaften.

Porst, Rolf. 2014b. Sekundäranalyse und Zugang zu sozialwissenschaftlichen Daten. *Gesellschaft. Wirtschaft. Politik* 63 (4): 553–562.

Pötschke, Manuela. 2010. Datengewinnung und Datenaufbereitung. In *Handbuch der sozialwissenschaftlichen Datenanalyse,* Hrsg. Christof Wolf und Henning Best, 41–64. Wiesbaden: VS Verlag für Sozialwissenschaften.

Pötschke, Manuela. 2014. Mehrebenenanalyse. In *Handbuch Methoden der empirischen Sozialforschung,* Hrsg. Nina Baur und Jörg Blasius, 1101–1115. Wiesbaden: Springer VS.

Precht, Richard David. 2015. *Anna, die Schule und der liebe Gott. Der Verrat des Bildungssystems an unseren Kindern.* München: Goldmann.

Prexl, Lydia. 2017. *Mit der Literaturübersicht die Bachelorarbeit meistern.* Konstanz: UVK.

Proner, Hanna. 2011. *Ist keine Antwort auch eine Antwort? Die Teilnahme an politischen Umfragen.* Wiesbaden: VS Verlag für Sozialwissenschaften.

Prüfer, Peter, und Margrit Rexroth. 2005. *Kognitive Interviews.* Mannheim: Zentrum für Umfragen, Methoden und Analysen.

Przeworski, Adam, und Henry Teune. 1970. *The logic of comparative social inquiry.* New York: Wiley.

Przyborski, Aglaja, und Monika Wohlrab-Sahr. 2014a. Forschungsdesigns für die qualitative Sozialforschung. In *Handbuch Methoden der empirischen Sozialforschung,* Hrsg. Nina Baur und Jörg Blasius, 117–133. Wiesbaden: Springer VS.

Przyborski, Aglaja, und Monika Wohlrab-Sahr. 2014b. *Qualitative Sozialforschung. Ein Arbeitsbuch.* München: Oldenbourg.

Putnam, Robert D. 1993. *Making democracy work. Civic traditions in modern Italy.* Princeton: Princeton University Press.

Putnam, Robert D. 1995. Tuning in, Tuning out: the Strange Disappearance of Social Capital in America. *Political Science and Politics* 28 (4): 664–683.

Putnam, Robert D. 2000. *Bowling Alone. The Collapse and Revival of American Community.* New York: Simon & Schuster.

Quatember, Andreas. 2015. *Statistischer Unsinn. Wenn Medien an der Prozenthürde scheitern.* Berlin: Springer Spektrum.

Rammstedt, Beatrice. 2010. Reliabilität, Validität, Objektivität. In *Handbuch der sozialwissenschaftlichen Datenanalyse,* Hrsg. Christof Wolf und Henning Best, 239–258. Wiesbaden: VS Verlag für Sozialwissenschaften.

Rammstedt, Beatrice, und Débora B. Maehler. 2014. PIAAC and its Methodological Challenges. *methods, data, analyses* 8 (2): 125–136.

Rattinger, Hans. 2009. *Einführung in die Politische Soziologie.* München: Oldenbourg.
Reinecke, Jost. 2014. Grundlagen der standardisierten Befragung. In *Handbuch Methoden der empirischen Sozialforschung,* Hrsg. Nina Baur und Jörg Blasius, 601–617. Wiesbaden: Springer VS.
Reinisch, Kerstin. 2017. *Wahl ohne Wissen? Empirische Analyse zur Entpolitisierung der Wahlkampfberichterstattung deutscher Printmedien.* Baden-Baden: Nomos.
Ressing, Meike, Maria Blettner, und Stefanie J. Klug. 2009. Systematische Übersichtsarbeiten und Metaanalysen. *Deutsches Ärzteblatt* 106 (27): 456–463.
Reuband, Karl-Heinz. 2001. Möglichkeiten und Probleme des Einsatzes postalischer Befragungen. *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 53 (2): 307–333.
Reuband, Karl-Heinz. 2002. Über die Schwierigkeit, den Nutzen postalischer Befragungen zu erkennen. Antwort auf eine Kritik von Rainer Schnell. *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 54 (1): 157–162.
Reuband, Karl-Heinz. 2014. Schriftlich-postalische Befragung. In *Handbuch Methoden der empirischen Sozialforschung,* Hrsg. Nina Baur und Jörg Blasius, 643–660. Wiesbaden: Springer VS.
Rippl, Susanne. 2008. Politische Sozialisation. In *Handbuch Sozialisationsforschung,* Hrsg. Klaus Hurrelmann, Matthias Grundmann und Sabine Walper, 443–457. Weinheim: Beltz.
Robinson, William S. 1950. Ecological Correlations and the Behavior of Individuals. *American Sociological Review* 15 (3): 351–357.
Rohrmann, Bernd. 1978. Empirische Studien zur Entwicklung von Antwortskalen für die sozialwissenschaftliche Forschung. *Zeitschrift für Sozialpsychologie* 9: 222–245.
Rölle, Daniel. 2002. Nicht genaues weiß man nicht? Über die Perzeption von Wahlprogrammen in der Öffentlichkeit. *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 54 (2): 264–280.
Roller, Edeltraud, Frank Brettschneider, und Jan W. van Deth, Hrsg. 2006. *Jugend und Politik: „Voll normal!" Der Beitrag der politischen Soziologie zur Jugendforschung.* Wiesbaden: VS Verlag für Sozialwissenschaften.
Roose, Jochen. 2013. Fehlermultiplikation und Pfadabhängigkeit. *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 65 (4): 697–714.
Rössler, Patrick. 2010. *Inhaltsanalyse.* Konstanz: UVK.
Rössler, Patrick. 2011. *Skalenhandbuch Kommunikationswissenschaft.* Wiesbaden: VS Verlag für Sozialwissenschaften.
Roßteutscher, Sigrid. 2004. Die Rückkehr der Tugend? In *Deutschland in Europa. Ergebnisse des European Social Survey 2002–2003,* Hrsg. Jan W. van Deth, 175–200. Wiesbaden: VS Verlag für Sozialwissenschaften.
Roßteutscher, Sigrid. 2009. Soziale Partizipation und Soziales Kapital. In *Politische Soziologie. Ein Studienbuch,* Hrsg. Viktoria Kaina und Andrea Römmele, 163–180. Wiesbaden: VS Verlag für Sozialwissenschaften.
Roßteutscher, Sigrid, und Armin Schäfer. 2016. Asymmetrische Mobilisierung: Wahlkampf und ungleiche Wahlbeteiligung. *Politische Vierteljahresschrift* 57 (3): 455–483.

Roth, Dieter. 1992. Sinkende Wahlbeteiligung – eher Normalisierung als Krisensymptom. In *Protestwähler und Wahlverweigerer. Krise der Demokratie?,* Hrsg. Karl Starzacher, Konrad Schacht, Bernd Friedrich und Thomas Leif, 58–68. Köln: Bund-Verlag.

Rubin, Donald D. 1987. *Multiple Imputation for Nonresponse in Surveys.* New York: Wiley.

Rudi, Tatjana. 2010. Die Links-Rechts-Dimension in Mittel- und Osteuropa: ‚Super-Issue' oder bedeutungslos? In *Information – Wahrnehmung – Emotion. Politische Psychologie in der Wahl- und Einstellungsforschung,* Hrsg. Thorsten Faas, Kai Arzheimer und Sigrid Roßteutscher, 169–189. Wiesbaden: VS Verlag für Sozialwissenschaften.

Salzborn, Samuel, Hrsg. 2016. *Klassiker der Sozialwissenschaften. 100 Schlüsselwerke im Portrait.* Wiesbaden: Springer VS.

Sälzer, Christine. 2016. *Studienbuch Schulleistungsstudien. Das Rasch-Modell in der Praxis.* Heidelberg: Springer.

Saris, Willem E., und Irmtraud N. Gallhofer. 2014. *Design, Evaluation, and Analysis of Questionnaires for Survey Research.* Hoboken: John Wiley & Sons.

Schaeffer, Nora Cate, und Stanley Presser. 2003. The Science of Asking Questions. *Annual Review of Sociology* 29: 65–88.

Schäfer, Ansgar, Christian Adam, und Michael Schlichenmaier. 2012. Das Plebiszit über das Bahnhofsprojekt „Stuttgart 21". Eine Analyse der Abstimmungsergebnisse in den Stadt- und Landkreisen Baden-Württembergs. *Zeitschrift für Politikwissenschaft* 22 (2): 187–213.

Schäfer, Armin. 2011. Mehr Mitsprache, aber nur für wenige? Direkte Demokratie und politische Gleichheit. In *MPIfG Jahrbuch 2011/2012,* Hrsg. Max-Planck-Institut für Gesellschaftsforschung, 53–59. Köln: Max-Planck-Institut für Gesellschaftsforschung.

Schäfer, Armin. 2012. Beeinflusst die sinkende Wahlbeteiligung das Wahlergebnis? Eine Analyse kleinräumiger Wahldaten in deutschen Großstädten. *Politische Vierteljahresschrift* 53 (2): 240–264.

Schäfer, Armin. 2015. *Der Verlust politischer Gleichheit. Warum die sinkende Wahlbeteiligung der Demokratie schadet?* Frankfurt: Campus.

Schäfer, Armin, und Harald Schoen. 2013. Mehr Demokratie, aber nur für wenige? Der Zielkonflikt zwischen mehr Beteiligung und politischer Gleichheit. *Leviathan* 41 (1): 94–120.

Schedler, Andreas. 2011. Concept Formation. In *International Encyclopedia of Political Science. Volume 2,* Hrsg. Bertrand Badie, Dirk Berg-Schlosser und Leonardo Morlino, 370–382. Thousand Oaks: Sage.

Schendera, Christian FG. 2015. *Deskriptive Statistik verstehen.* Konstanz: UVK.

Schiller, Theo, und Volker Mittendorf. 2002. Neue Entwicklungen der direkten Demokratie. In *Direkte Demokratie,* Hrsg. Theo Schiller und Volker Mittendorf, 7–21. Wiesbaden: Westdeutscher Verlag.

Schimpl-Neimanns, Bernhard, und Felix Weiss. 2014. Zur Bereitstellung amtlicher Mikrodaten für die Wissenschaft aus sozialwissenschaftlicher Perspektive. *AStA Wirtschafts- und Sozialstatistisches Archiv* 8 (4): 205–219.

Schindler, Steffen. 2014. *Wege zur Studienberechtigung – Wege ins Studium? Eine Analyse sozialer Inklusions- und Ablenkungsprozesse.* Wiesbaden: Springer VS.

Schlichte, Klaus, und Julia Sievers. 2015. *Einführung in die Arbeitstechniken der Politikwissenschaft.* Wiesbaden: Springer VS.

Schmidt, Manfred G. 2010. *Demokratietheorien. Eine Einführung.* Bonn: Bundeszentrale für politische Bildung.

Schmied, Gerhard. 2002. *Friedhofsgespräche. Untersuchungen zum „Wohnort der Toten".* Opladen: Leske+Budrich.

Schmitt-Beck, Rüdiger. 1992. Wertewandel. In *Lexikon der Politik. Band 3: Die westlichen Länder,* Hrsg. Manfred G. Schmidt, 527–533. München: Beck.

Schmitt-Beck, Rüdiger, Hrsg. 2012. *Wählen in Deutschland. Sonderheft der Politischen Vierteljahresschrift 45/2011.* Baden-Baden: Nomos.

Schmitt-Beck, Rüdiger, Christian Mackenrodt, und Thorsten Faas. 2008. Hintergründe kommunaler Wahlbeteiligung. Eine Fallstudie zur Kommunalwahl 2004 in Duisburg. *Zeitschrift für Parlamentsfragen* 39 (3): 561–581.

Schmitt-Beck, Rüdiger, Hans Rattinger, Sigrid Roßteutscher, und Bernhard Weßels. 2010. Die deutsche Wahlforschung und die German Longitudinal Election Study (GLES). In *Gesellschaftliche Entwicklungen im Spiegel der empirischen Sozialforschung,* Hrsg. Frank Faulbaum und Christof Wolf, 141–172. Wiesbaden: VS Verlag für Sozialwissenschaften.

Schmitt, Annette. 2014. Die Rolle von Wahlen in der Demokratie. In *Handbuch Wahlforschung,* Hrsg. Jürgen W. Falter und Harald Schoen, 3–35. Wiesbaden: Springer VS.

Schmitt, Hermann. 2003. The Eurobarometers: Their Evolution, Obvious Merits, and Ways to Add Value to them. *European Union Politics* 4 (2): 243–251.

Schmitt, Hermann. 2005. The European Parliament Elections of June 2004: Still Second-Order? *West European Politics* 28 (3): 650–679.

Schnapp, Kai-Uwe, Nathalie Behnke, und Joachim Behnke. 2004. Methodenausbildung in der Politikwissenschaft – Oder: Wie aus dem Aschenputtel eine Prinzessin werden kann. *Rundbrief der Deutschen Vereinigung für Politische Wissenschaft 131:* 158–167.

Schnapp, Kai-Uwe, Delia Schindler, Thomas Gschwend, und Joachim Behnke. 2006. Qualitative und Quantitative Zugänge: Eine integrative Perspektive. In *Methoden der Politikwissenschaft. Neuere qualitative und quantitative Analyseverfahren,* Hrsg. Joachim Behnke, Thomas Gschwend, Delia Schindler und Kai-Uwe Schnapp, 11–26. Baden-Baden: Nomos.

Schnaudt, Christian. 2013. In search of reliable results: Socialist values and political trust in unified Germany – A comment on Campbell's analyses and conclusions. *European Journal of Political Research* 52 (3): 419–430.

Schnaudt, Christian, Michael Weinhardt, Rory Fitzgerald, und Stefan Liebig. 2014. The European Social Survey: Contents, Design, and Research Potential. *Schmollers Jahrbuch. Zeitschrift für Wirtschafts- und Sozialwissenschaften* 134 (4): 487–506.

Schnaudt, Christian, Michael Weinhardt, und Stefan Liebig. 2016. (Um-)Fragen über (Um-)Fragen: Die Entstehung von Umfragedaten am Beispiel des European Social Survey (ESS). *Gesellschaft. Wirtschaft. Politik* 65 (2): 189–198.

Schneekloth, Ulrich, und Ingo Leven. 2003. Woran bemisst sich eine „gute" allgemeine Bevölkerungsumfrage? Analysen zu Ausmaß, Bedeutung und zu den Hintergründen von Nonresponse in zufallsbasierten Stichprobenerhebungen am Beispiel des ALLBUS. *ZUMA-Nachrichten* 27 (53): 16–57.

Schneider, Silke L. 2008. *The International Standard Classification of Education (ISCED-97). An Evaluation of Content and Criterion Validity for 15 European Countries.* Mannheim: Mannheimer Zentrum für Europäische Sozialforschung.

Schnell, Rainer. 1997. *Nonresponse in Bevölkerungsumfragen. Ausmaß, Entwicklung und Ursachen.* Opladen: Leske+Budrich.

Schnell, Rainer. 2002. Anmerkungen zur Publikation „Möglichkeiten und Probleme des Einsatzes postalischer Befragungen" von Karl-Heinz Reuband in der KZfSS 2001, S. 307–333. *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 54 (1): 147–156.

Schnell, Rainer. 2012. *Survey-Interviews. Methoden standardisierter Befragungen.* Wiesbaden: VS Verlag für Sozialwissenschaften.

Schnell, Rainer, Paul B. Hill, und Elke Esser. 2013. *Methoden der empirischen Sozialforschung.* München: Oldenbourg.

Schoen, Harald. 2009. Wahlsoziologie. In *Politische Soziologie. Ein Studienbuch,* Hrsg. Viktoria Kaina und Andrea Römmele, 181–208. Wiesbaden: VS Verlag für Sozialwissenschaften.

Scholl, Armin. 2003. *Die Befragung.* Konstanz: UVK.

Schuck, Andreas R. T., Rachid Azrout, Hajo Boomgaarden, Matthijs Elenbaas, Joost van Spanje, Rens Vliegenthart, und Claes De Vreese. 2011. Media Visibility and Framing of the European Parliamentary Elections 2009: A Media Content Analysis in 27 Countries. In *Political Communication in European Parliamentary Elections,* Hrsg. Michaela Maier, Jesper Strömbäck und Lynda Lee Kaid, 175–196. Farnham: Ashgate.

Schülein, Johann August, und Simon Reitze. 2012. *Wissenschaftstheorie für Einsteiger.* Stuttgart: UTB.

Schumpeter, Joseph A. 1950. *Kapitalismus, Sozialismus und Demokratie.* Bern: Francke.

Schupp, Jürgen. 2009. 25 Jahre Sozio-oekonomisches Panel – Ein Infrastrukturprojekt der empirischen Sozial- und Wirtschaftsforschung in Deutschland. *Zeitschrift für Soziologie* 38 (5): 350–357.

Schupp, Jürgen. 2014. Paneldaten für die Sozialforschung. In *Handbuch Methoden der empirischen Sozialforschung,* Hrsg. Nina Baur und Jörg Blasius, 925–939. Wiesbaden: Springer VS.

Schupp, Jürgen, und Christof Wolf. 2015. *Nonresponse Bias. Qualitätssicherung sozialwissenschaftlicher Umfragen.* Wiesbaden: Springer VS.

Schurz, Gerhard. 2009. Wissenschaftliche Erklärung. In *Wissenschaftstheorie. Ein Studienbuch,* Hrsg. Andreas Bartels und Manfred Stöckler, 69–88. Paderborn: Mentis.

Schurz, Gerhard. 2014. *Einführung in die Wissenschaftstheorie.* Darmstadt: WBG.

Schwanitz, Dietrich. 2002. *Bildung. Alles, was man wissen muß.* München: Goldmann.

Schwarz, Norbert. 1991. *In welcher Reihenfolge fragen? Kontexteffekte in standardisierten Befragungen. ZUMA-Arbeitsbericht 1991/16.* Mannheim: Zentrum für Umfragen, Methoden und Analysen.

Schwarz, Norbert, und Herbert Bless. 1992. Scandals and the Public's Trust in Politicians: Assimilation and Contrast Effects. *Personality and Social Psychology Bulletin* 18 (5): 574–579.

Schwarz, Norbert, und Hans-J. Hippler. 1995. The Numeric Values of Rating Scales: A Comparison of their Impact in Mail Surveys and Telephone Interviews. *International Journal of Public Opinion Research* 7 (1): 72–74.

Schwarz, Norbert, Hans-Jürgen Hippler, Brigitte Deutsch, und Fritz Strack. 1985. Response Scales: Effects of Category Range on Reported Behavior and Comparative Judgments. *Public Opinion Quarterly* 49 (3): 388–395.

Schwarz, Norbert, Bärbel Knauper, Hans-J. Hippler, Elisabeth Noelle-Neumann, und Leslie Clark. 1991. Rating Scales. Numeric Values May Change the Meaning of Scale Labels. *Public Opinion Quarterly* 55 (4): 570–582.

Schwarz, Norbert, und Fritz Strack. 1991. Context Effects in Attitude Surveys: Applying Cognitive Theory to Social Research. *European Review of Social Psychology* 2 (1): 31–50.

Schwarzer, Gudrun. 2001. Forschungsbeiträge verfassen. *Zeitschrift für Internationale Beziehungen* 8 (1): 141–156.

Sciarini, Pascal, und Andreas C. Goldberg. 2016. Turnout Bias in Postelection Surveys: Political Involvement, Survey Participation, and Vote Overreporting. *Journal of Survey Statistics and Methodology* 4 (1): 110–137.

Sebe-Opfermann, Andreas. 2016. Die Fragen der Fragen – was ist eine gute Forschungsfrage? In *Wege durch den Forschungsdschungel,* Hrsg. Nina Dunker, Nina-Kathrin Joyce-Finnern und llka Koppel, 21–36. Wiesbaden: Springer VS.

Sedlmeier, Peter, und Frank Renkewitz. 2008. *Forschungsmethoden und Statistik in der Psychologie.* München: Pearson.

Seher, Nicole Michaela, und Franz Urban Pappi. 2011. Politikfeldspezifische Positionen der Landesverbände der deutschen Parteien. Mannheim. Mannheimer Zentrum für Europäische Sozialforschung (Arbeitspapier Nr. 139).

Shadish, William R., Thomas D. Cook, und Donald T. Campbell. 2002. *Experimental and Quasi-Experimental Designs for Generalized Causal Inference.* Boston: Houghton Mifflin.

Sieberer, Ulrich. 2008. Prinzipal Parlament. Die Bedeutung europäischer Parlamente als Wahlorgane. *Politische Vierteljahresschrift* 49 (2): 251–282.

Siefken, Sven T. 2013. Repräsentation vor Ort: Selbstverständnis und Verhalten von Bundestagsabgeordneten bei der Wahlkreisarbeit. *Zeitschrift für Parlamentsfragen* 44 (3): 486–508.

Signorelli, Salvatore. 2012. *The EU and public opinions: A love-hate relationship?* Paris: Notre Europe.

Snijders, Tom A. B., und Roel J. Bosker. 2012. *Multilevel Analysis. An Introduction to Basic and Advanced Multilevel Modeling.* London: Sage.

Solomon, Richard L. 1949. An extension of control group design. *Psychological Bulletin* 46 (2): 127–150.

Spieß, Martin. 2010. Der Umgang mit fehlenden Werten. In *Handbuch der sozialwissenschaftlichen Datenanalyse,* Hrsg. Christof Wolf und Henning Best, 117–142. Wiesbaden: VS Verlag für Sozialwissenschaften.

Stähli, Michèle Ernst, und Dominique Joye. 2016. Incentives as a Possible Measure to Increase Response Rates. In *The Sage Handbook of Survey Methodology,* Hrsg. Christof Wolf, Dominique Joye, Tom E. C. Smith und Yang-Chih Fu, 425–440. London: SAGE.

Stark, Toralf, und Christoph Mohamad-Klotzbach. 2016. Zentrale Datenquellen, Handbücher und Zeitschriften. In *Handbuch Vergleichende Politikwissenschaft,* Hrsg. Hans-Joachim Lauth, Marianne Kneuer und Gert Pickel, 947–969. Wiesbaden: Springer VS.

Statistisches Bundesamt. 2009. In fast jedem zehnten Haushalt ersetzen Handys das Festnetz. Pressemitteilung Nr. 184 vom 14. 05. 2009.

Statistisches Bundesamt. 2016. *Statistik und Wissenschaft. Demographische Standards. Ausgabe 2016.* Wiesbaden: Statistisches Bundesamt.

Stein, Petra. 2014. Forschungsdesigns für die quantitative Sozialforschung. In *Handbuch Methoden der empirischen Sozialforschung,* Hrsg. Nina Baur und Jörg Blasius, 135–151. Wiesbaden: Springer VS.

Steinbrecher, Markus, und Hans Rattinger. 2011. Die Wahlbeteiligung. In *Zwischen Langeweile und Extremen: Die Bundestagswahl 2009,* Hrsg. Hans Rattinger, Sigrid Roßteutscher, Rüdiger Schmitt-Beck und Bernhard Weßels, 77–90. Baden-Baden: Nomos.

Stevens, S. S. 1946. On the Theory of Scales of Measurement. *Science* 103 (2684): 677–680.

Stocké, Volker. 2014. Persönlich-mündliche Befragung. In *Handbuch Methoden der empirischen Sozialforschung,* Hrsg. Nina Baur und Jörg Blasius, 619–629. Wiesbaden: Springer VS.

Stoop, Ineke A. L. 2016. Unit Nonresponse. In *The Sage Handbook of Survey Methodology,* Hrsg. Christof Wolf, Dominique Joye, Tom E. C. Smith und Yang-Chih Fu, 409–424. London: SAGE.

Stouffer, Samuel A. 1949. *The American soldier.* Princeton: Princeton University Press.

Strohmeier, Gerd. 2005. Politik bei Benjamin Blümchen und Bibi Blocksberg. *Aus Politik und Zeitgeschichte* 55 (41): 7–15.

Sturm, Gabriele. 2006. Abduktion. In *Methoden der Politikwissenschaft. Neuere qualitative und quantitative Analyseverfahren,* Hrsg. Joachim Behnke, Thomas Gschwend, Delia Schindler und Kai-Uwe Schnapp, 27–35. Baden-Baden: Nomos.

Sturm, Roland. 2010. Der Vertrag von Lissabon. In *Vertrag von Lissabon,* Hrsg. Bundeszentrale für politische Bildung, 15–30. Bonn: Bundeszentrale für politische Bildung.

Stürmer, Stefan. 2011. Soziale Repräsentationen von Bürgerprotesten: Der Wutbürger – soziale Realität, Feuilleton-Chimäre oder politischer Kampfbegriff? *Politische Psychologie* 1 (1): 9–18.

Stykow, Petra, Christopher Daase, Janet MacKenzie, und Nikola Moosauer. 2010. *Politikwissenschaftliche Arbeitstechniken.* Stuttgart: UTB.

Tashakkori, Abbas, und Charles Teddlie, Hrsg. 2010. *SAGE Handbook of Mixed Methods in Social & Behavioral Research.* Thousand Oaks: Sage.

Tausendpfund, Markus. 2008. Demokratie Leben Lernen – Erste Ergebnisse der dritten Welle. Politische Orientierungen von Kindern im vierten Grundschuljahr. Mannheim. Mannheimer Zentrum für Europäische Sozialforschung (Arbeitspapier Nr. 116).

Tausendpfund, Markus. 2013a. Gemeinden als Fundament der Demokratie? Individuelle und kontextuelle Bestimmungsfaktoren der Zufriedenheit der Bürger mit der Demokratie. In *Die Versprechen der Demokratie. 25. wissenschaftlicher Kongress der Deutschen Vereinigung für Politische Wissenschaft,* Hrsg. Hubertus Buchstein, 269–292. Baden-Baden: Nomos.

Tausendpfund, Markus. 2013b. *Gemeinden als Rettungsanker der EU? Individuelle und kontextuelle Faktoren der Zustimmung der Bürgerinnen und Bürger zur Europäischen Union.* Baden-Baden: Nomos.

Tausendpfund, Markus. 2013c. Zustimmung von Kommunalpolitikern zur weiteren europäischen Integration. *Zeitschrift für Politikwissenschaft* 23 (3): 365–391.

Tausendpfund, Markus, und Daniela Braun. 2008. Die schwierige Suche nach Ergebnissen der Wahlen zum Europäischen Parlament: Ein neuer Datensatz für die Wahlen 1979 bis 2004. *Zeitschrift für Parlamentsfragen* 39 (1): 84–93.

Tausendpfund, Markus, und Angelika Vetter, Hrsg. 2017. *Politische Einstellungen von Kommunalpolitikern im Vergleich.* Wiesbaden: Springer VS.

Tenscher, Jens, und Philipp Scherer. 2012. *Jugend, Politik und Medien. Politische Orientierungen und Verhaltensweisen von Jugendlichen in Rheinland-Pfalz.* Münster: LIT.

Terwey, Michael. 2014. Zur Gewichtung von ALLBUS-Umfragedaten. In *Variable Report ALLBUS/Allgemeine Bevölkerungsumfrage der Sozialwissenschaften Kumulation 1980–2012. ZA-Nr. 4578. GESIS – Variable Reports 2014/07,* Hrsg. Michael Terwey und Stefan Baltzer, iii–xxiii. Köln: GESIS.

Theocharis, Yannis, und Jan W. van Deth. 2016. The continuous expansion of citizen participation: a new taxonomy. *European Political Science Review:* doi: 10.1017/S1755773916000230.

Theocharis, Yannis, und Jan W. van Deth. 2018. *Political Participation in a Changing World. Conceptual and Empirical Challenges in the Study of Citizen Engagement.* New York: Routledge.

Töller, Annette Elisabeth, Sylvia Pannowitsch, Céline Kuschek, und Christian Mennrich. 2011. Direkte Demokratie und Schulpolitik. Lehren aus einer politikfeldanalytischen Betrachtung des Scheiterns der Hamburger Schulreform. *Zeitschrift für Parlamentsfragen* 42 (3): 503–523.

Trüdinger, Eva-Maria, und Uwe Bollow. 2011. Andere Zeiten, andere Inhalte. Bedeutungsgehalt und Bedeutungswandel der politischen Richtungsbegriffe Links und Rechts im innerdeutschen Vergleich. *Zeitschrift für Parlamentsfragen* 42 (2): 398–418.

Urban, Dieter, und Jochen Mayerl. 2011. *Regressionsanalyse: Theorie, Technik und Anwendung.* Wiesbaden: VS Verlag für Sozialwissenschaften.

van Deth, Jan. 2014. A conceptual map of political participation. *Acta Politica* 49 (3): 349–367.

van Deth, Jan W. 2003. Vergleichende politische Partizipationsforschung. In *Vergleichende Politikwissenschaft,* Hrsg. Dirk Berg-Schlosser und Ferdinand Müller-Rommel, 167–187. Opladen: Leske+Budrich.

van Deth, Jan W., Hrsg. 2004. *Deutschland in Europa. Ergebnisse des European Social Survey 2002–2003.* Wiesbaden: VS Verlag für Sozialwissenschaften.

van Deth, Jan W. 2009. Politische Partizipation. In *Politische Soziologie. Ein Studienbuch,* Hrsg. Viktoria Kaina und Andrea Römmele, 141–161. Wiesbaden: VS Verlag für Sozialwissenschaften.

van Deth, Jan W. 2012. New modes of participation and norms of citizenship. In *New Participatory Dimensions in Civil Society. Professionalization and Individualized Collective Action,* Hrsg. Jan W. van Deth und William Maloney, 115–138. London: Routledge.

van Deth, Jan W., Simone Abendschön, Julia Rathke, und Meike Vollmar. 2007. *Kinder und Politik. Politische Einstellungen von jungen Kindern im ersten Grundschuljahr.* Wiesbaden: VS Verlag für Sozialwissenschaften.

van Deth, Jan W., und Elinor Scarbrough. 1995. The Concept of Values. In *The Impact of Values,* Hrsg. Jan W. van Deth und Elinor Scarbrough, 21–47. Oxford: Oxford University Press.

van Deth, Jan W., und Markus Tausendpfund, Hrsg. 2013. *Politik im Kontext: Ist alle Politik lokale Politik? Individuelle und kontextuelle Determinanten politischer Orientierungen.* Wiesbaden: Springer VS.

Verba, Sidney, und Norman H. Nie. 1972. *Participation in America. Political Democracy and Social Equality.* New York: Harper & Row.

Verba, Sidney, Kay Lehman Schlozman, und Henry E. Brady. 1995. *Voice and Equality. Civic Voluntarism in American Politics.* Cambridge: Harvard University Press.

Vetter, Angelika. 2008. Kommunale Wahlbeteiligung im Bundesländervergleich: Politische Institutionen und ihre Folgen. *Die Öffentliche Verwaltung* 61 (21): 885–894.

Völkl, Kerstin, Kai-Uwe Schnapp, Everhard Holtmann, und Oscar W. Gabriel. 2008. *Wähler und Landtagswahlen in der Bundesrepublik Deutschland.* Baden-Baden: Nomos.

von dem Berge, Benjamin, Thomas Poguntke, Peter Obert, und Diana Tipei. 2013. *Measuring Intra-Party Democracy. A Guide for the Content Analysis of Party Statutes with Examples from Hungary, Slovakia and Romania.* Heidelberg: Springer.

von der Heyde, Christian. 2014a. Allgemeine Theorie von Random-Stichproben. In *Stichproben-Verfahren in der Umfrageforschung. Eine Darstellung für die Praxis,* Hrsg. ADM Arbeitskreis Deutscher Markt- und Sozialforschungsinstitute e. V., 25–36. Wiesbaden: Springer VS.

von der Heyde, Christian. 2014b. Einwohnermeldeamts-Stichproben (EWA-Stichproben). In *Stichproben-Verfahren in der Umfrageforschung. Eine Darstellung für die Praxis,* Hrsg. ADM Arbeitskreis Deutscher Markt- und Sozialforschungsinstitute e. V., 191–195. Wiesbaden: Springer VS.

Voss, Kathrin, Hrsg. 2014. *Internet und Partizipation. Bottom-up oder Top-down? Politische Beteiligungsmöglichkeiten im Internet.* Wiesbaden: Springer VS.

Waal, Ton de, Jeroen Pannekoek, und Sander Scholtus. 2011. *Handbook of Statistical Data Editing and Imputation.* Hoboken, New Jersey: Wiley.

Wagner, Gert G., Jan Göbel, Peter Krause, Rainer Pischner, und Ingo Sieber. 2008. Das Sozio-oekonomische Panel (SOEP): Multidisziplinäres Haushaltspanel und Kohortenstudie für Deutschland – Eine Einführung (für neue Datennutzer) mit einem Ausblick (für erfahrene Anwender). *AStA Wirtschafts- und Sozialstatistisches Archiv* 2 (4): 301–328.

Wagner, Pia, und Linda Hering. 2014. Online-Befragung. In *Handbuch Methoden der empirischen Sozialforschung,* Hrsg. Nina Baur und Jörg Blasius, 661–673. Wiesbaden: Springer VS.

Wagschal, Uwe, und Sebastian Jäckle. 2011. Aggregate Data Analysis. In *International Encyclopedia of Political Science. Volume 1,* Hrsg. Bertrand Badie, Dirk Berg-Schlosser und Leonardo Morlino, 53–57. Thousand Oaks: Sage.

Walter-Rogg, Melanie. 2008. Direkte Demokratie. In *Die EU-Staaten im Vergleich. Strukturen, Prozesse, Politikinhalte,* Hrsg. Oscar W. Gabriel und Sabine Kropp, 236–267. Wiesbaden: VS Verlag für Sozialwissenschaften.

Walter, Franz, und Jöran Klatt. 2014. Robert Michels: Zur Soziologie des Parteienwesens in der modernen Demokratie. In *Klassiker der Sozialwissenschaften. 100 Schlüsselwerke im Portrait,* Hrsg. Samuel Salzborn, 68–71. Wiesbaden: Springer VS.

Warwick, Paul V. 2002. Toward a Common Dimensionality in West European Policy Spaces. *Party Politics* 8 (1): 101–122.

Wasmer, Martina, Michael Blohm, Jessica Walter, Regina Jutz, und Evi Scholz. 2017. *Konzeption und Durchführung der „Allgemeinen Bevölkerungsumfrage der Sozialwissenschaften" (ALLBUS) 2014. GESIS Papers 2017/20.* Mannheim: GESIS – Leibniz-Institut für Sozialwissenschaften.

Watteler, Oliver. 2017. Recherche nach sozialwissenschaftlichen Forschungsdaten. In *Grundlagen sozialwissenschaftlichen Arbeitens. Eine anwendungsorientierte Einführung,* Hrsg. Ina Berninger, Katrin Botzen, Christian Kolle, Dominikus Vogl und Oliver Watteler, 127–155. Stuttgart: UTB.

Weichbold, Martin. 2014. Pretest. In *Handbuch Methoden der empirischen Sozialforschung,* Hrsg. Nina Baur und Jörg Blasius, 299–304. Wiesbaden: Springer VS.

Weichbold, Martin, Johann Bacher, und Christof Wolf, Hrsg. 2009. *Umfrageforschung. Herausforderungen und Grenzen.* Wiesbaden: VS Verlag für Sozialwissenschaften.

Weischer, Christoph. 2015a. Befragung. In *Methoden-Lexikon für die Sozialwissenschaften,* Hrsg. Rainer Diaz-Bone und Christoph Weischer, 31–35. Wiesbaden: Springer VS.

Weischer, Christoph. 2015b. Beobachtung. In *Methoden-Lexikon für die Sozialwissenschaften,* Hrsg. Rainer Diaz-Bone und Christoph Weischer, 38. Wiesbaden: Springer VS.

Weixner, Bärbel Martina. 2006. Direkte Demokratie in den Bundesländern. *Aus Politik und Zeitgeschichte* 56 (10): 18–24.

Weller, Ingeborg. 1996. *Kontexteffekte in Eurobarometer-Umfragen. Theoretische Implikationen und praktische Bedeutung.* Münster: Waxmann.

Wenzelburger, Georg, Sebastian Jäckle, und Pascal König. 2014. *Weiterführende statistische Methoden für Politikwissenschaftler. Eine anwendungsbezogene Einführung mit Stata.* München: Oldenbourg.

Weßels, Bernhard. 2009. Spielarten des Euroskeptizismus. In *Die Verfassung Europas. Perspektiven des Integrationsprojekts,* Hrsg. Frank Decker und Marcus Höreth, 50–68. Wiesbaden: VS Verlag für Sozialwissenschaften.

Westle, Bettina. 1989. *Politische Legitimität – Theorien, Konzepte, empirische Befunde.* Baden-Baden: Nomos.

Westle, Bettina. 2009a. Auswahl einer Forschungsfrage und Konzeptspezifikation. In *Methoden der Politikwissenschaft,* Hrsg. Bettina Westle, 125–131. Baden-Baden: Nomos.

Westle, Bettina. 2009b. Auswahlverfahren. In *Methoden der Politikwissenschaft,* Hrsg. Bettina Westle, 157–176. Baden-Baden: Nomos.

Westle, Bettina. 2009c. Forschungsdesigns. In *Methoden der Politikwissenschaft,* Hrsg. Bettina Westle, 133–155. Baden-Baden: Nomos.

Westle, Bettina. 2009d. Grundgedanken und Grundelemente quantitativer Forschung. In *Methoden der Politikwissenschaft,* Hrsg. Bettina Westle, 75–96. Baden-Baden: Nomos.

Westle, Bettina. 2009e. Operationalisierung und Messung. In *Methoden der Politikwissenschaft,* Hrsg. Bettina Westle, 177–205. Baden-Baden: Nomos.

Westle, Bettina. 2012. Souveräne Teilhabe unter Unsicherheit und Halbwissen: Politisches Wissen und politische Partizipation. In *Die verstimmte Demokratie. Moderne Volksherrschaft zwischen Aufbruch und Frustration,* Hrsg. Stephan Braun und Alexander Geisler, 51–68. Wiesbaden: Springer VS.

Westle, Bettina, und Thomas Krumm. 2009. Beobachtung. In *Methoden der Politikwissenschaft,* Hrsg. Bettina Westle, 260–271. Baden-Baden: Nomos.

Westle, Bettina, Thomas Schübel, und Aribert Heyder. 2013. Düstere Zukunft: Es gibt wieder Wahlen, aber niemand geht hin? In *Wahlen und Wähler. Analysen aus Anlass der Bundestagswahl 2009,* Hrsg. Bernhard Weßels, Harald Schoen und Oscar W. Gabriel, 475–495. Wiesbaden: Springer VS.

Willis, Gordon B. 2005. *Cognitive Interviewing. A Tool for Improving Questionnaire Design.* Thousand Oaks: Sage.

Willis, Gordon B. 2016. Questionnaire Pretesting. In *The Sage Handbook of Survey Methodology,* Hrsg. Christof Wolf, Dominique Joye, Tom E. C. Smith und Yang-Chih Fu, 359–381. London: SAGE.

Wirth, Heike, und Walter Müller. 2006. Mikrodaten der amtlichen Statistik – Ihr Potenzial in der empirischen Sozialforschung. In *Methoden der Sozialforschung. Kölner Zeitschrift für Soziologie und Sozialpsychologie. Sonderheft 44/2004,* Hrsg. Andreas Diekmann, 93–127. Wiesbaden: VS Verlag für Sozialwissenschaften.

Wolf, Christof. 1995. Sozio-Ökonomischer Status und Berufliches Prestige. Ein kleines Kompendiurn sozialwissenschaftlicher Skalen auf Basis der beruflichen Stellung und Tatigkeit. *ZUMA-Nachrichten* 37: 102–136.

Wolf, Christof, und Henning Best. 2010a. *Handbuch der sozialwissenschaftlichen Datenanalyse.* Wiesbaden: VS Verlag für Sozialwissenschaften.

Wolf, Christof, und Henning Best. 2010b. Lineare Regressionsanalyse. In *Handbuch der sozialwissenschaftlichen Datenanalyse,* Hrsg. Christof Wolf und Henning Best, 607–638. Wiesbaden: VS Verlag für Sozialwissenschaften.

Wolf, Christof, Dominique Joye, Tom E. C. Smith, und Yang-Chih Fu, Hrsg. 2016. *The Sage Handbook of Survey Methodology.* London: SAGE.

Wonka, Arndt. 2007. Um was geht es? Konzeptspezifikation in der politikwissenschaftlichen Forschung. In *Forschungsdesign in der Politikwissenschaft. Probleme – Strategien – Anwendungen,* Hrsg. Thomas Gschwend und Frank Schimmelfennig, 63–89. Frankfurt: Campus.

Wüst, Andreas M., Hrsg. 2003. *Politbarometer.* Opladen: Leske+Budrich.

Wüst, Andreas M., und Markus Tausendpfund. 2009. 30 Jahre Europawahlen. *Aus Politik und Zeitgeschichte* 59 (23-24): 3–9.

Zeglovits, Eva, und Sylvia Kritzinger. 2014. New Attempts to Reduce Overreporting of Voter Turnout and Their Effects. *International Journal of Public Opinion Research* 26 (2): 224–234.

Züll, Cornelia, und Natalja Menold. 2014. Offene Fragen. In *Handbuch Methoden der empirischen Sozialforschung,* Hrsg. Nina Baur und Jörg Blasius, 713–719. Wiesbaden: Springer VS.

Züll, Cornelia, Evi Scholz, und Hermann Schmitt. 2010. *Kategorienschema für die Vercodung von Assoziationen mit den politischen Richtungsbegriffen ‚links' und ‚rechts'. GESIS-Technical Reports 2010/19.* Mannheim: GESIS – Leibniz-Institut für Sozialwissenschaften.