Zusammenfassung Forschungsmethoden

[**1. Was sind Forschungsmethoden der Psychologie?** 3](#_Toc3457646)

[**1.1 Definition** 3](#_Toc3457647)

[**1.2 Wofür braucht man das?** 3](#_Toc3457648)

[**2. Psychologie als empirische Wissenschaft** 3](#_Toc3457649)

[**2.1 Alltagspsychologie vs. wissenschaftliche Psychologie** 3](#_Toc3457650)

[2.1.1 Alltagspsychologie 4](#_Toc3457651)

[2.1.2 Wissenschaftliche Psychologie 4](#_Toc3457652)

[**2.2 Exkurs: Wissenschaftstheorie** 4](#_Toc3457653)

[**2.3 Historischer Hintergrund** 5](#_Toc3457654)

[**2.4 Qualitative vs. Quantitative Methoden** 6](#_Toc3457655)

[**2.5 Begriffserklärungen: Variablen, Operationalisierungen** 7](#_Toc3457656)

[**2.6 Basisziele der Psychologie** 7](#_Toc3457657)

[2.6.1 Beschreiben 7](#_Toc3457658)

[2.6.2 Erklären 7](#_Toc3457659)

[2.6.3 Vorhersagen 8](#_Toc3457660)

[2.6.4 Verändern 9](#_Toc3457661)

[**3. Quantitative Erhebungsmethoden** 10](#_Toc3457662)

[**3.1 Forschungsidee / Forschungsfrage finden** 10](#_Toc3457663)

[**3.2 Ethische Richtlinien** 10](#_Toc3457664)

[3.2.1 Ethische Richtlinien im Umgang mit VPN 10](#_Toc3457665)

[3.1.2 Ethik im Hinblick auf wissenschaftliche Kriterien 11](#_Toc3457666)

[3.1.3 Open Science 12](#_Toc3457667)

[**3.2 Hypothesen formulieren** 14](#_Toc3457668)

[**3.3 Definition und Messung der Variablen** 15](#_Toc3457669)

[3.3.1 Probleme des Selbstberichts 15](#_Toc3457670)

[3.3.2 Quantitative Gütekriterien 16](#_Toc3457671)

[3.3.2.1 Objektivität 16](#_Toc3457672)

[3.3.2.2 Reliabilität 17](#_Toc3457673)

[3.3.2.3. Validität 20](#_Toc3457674)

[3.3.3. Beobachten 21](#_Toc3457675)

[3.3.3.1. Arten der Beobachtung 22](#_Toc3457676)

[3.3.3.2. Formen der Beobachtung 22](#_Toc3457677)

[3.3.4 Messen 23](#_Toc3457678)

[3.3.5 Zählen 23](#_Toc3457679)

[3.3.6 Befragung 24](#_Toc3457680)

[3.3.6.1 Schriftliche vs. Mündliche Befragung 24](#_Toc3457681)

[3.3.6.2 Selbstberichtverfahren: Rating 25](#_Toc3457682)

[3.3.6.2.1 Ratingskalen: Anzahl der Skalenstufen 26](#_Toc3457683)

[3.3.7 Testen 27](#_Toc3457684)

[3.3.7.1 Testarten: Leistungs- und Persönlichkeitstests 27](#_Toc3457685)

[**3.4 Identifizierung und Auswahl der Studienteilnehmenden** 29](#_Toc3457686)

[3.4.1 Repräsentativität von Stichproben 29](#_Toc3457687)

[3.4.2 Arten von Stichproben 30](#_Toc3457688)

[**4. Quantitative Forschungsmethoden** 32](#_Toc3457689)

[**4.1 Forschungsstrategie / Forschungsdesign** 32](#_Toc3457690)

[4.1.1 Deskriptives Design 32](#_Toc3457691)

[4.1.2 Korrelatives Design 32](#_Toc3457692)

[4.1.3 Nichtexperimentelles Forschungsdesign 34](#_Toc3457693)

[4.1.4 Zusammenhänge / Korrelationen und Kausalität 34](#_Toc3457694)

[4.1.5 Experimente 35](#_Toc3457695)

[4.1.5.1 Begriffe 35](#_Toc3457696)

[4.1.5.2 Kontrollgruppen 36](#_Toc3457697)

[4.1.5.3 Versuchsplan 36](#_Toc3457698)

[4.1.5 Quasiexperimente 41](#_Toc3457699)

[4.1.6 Vergleich der Forschungsdesigns 42](#_Toc3457700)

[4.1.7 Meta-Analyse 43](#_Toc3457701)

[4.1.7.1 Vorgehen 43](#_Toc3457702)

[4.1.7.2 Fazit 44](#_Toc3457703)

[4.2 Interne und Externe Validität 45](#_Toc3457704)

[4.2.1 Interne Validität 45](#_Toc3457705)

[4.2.2 Externe Validität 45](#_Toc3457706)

[4.2.3 Feld- & Laboruntersuchungen 45](#_Toc3457707)

[4.2.4 Gefährdungen der internen Validität 46](#_Toc3457708)

[4.3 Statistische Regressionseffekte – Regression zur Mitte 48](#_Toc3457709)

# **1. Was sind Forschungsmethoden der Psychologie?**

## **1.1 Definition**

Der *Gegenstand der Psychologie ist das Erleben, Verhalten und Handeln* des Menschen. Unter *psychologischen Methoden* verstehen wir also *Vorgehensweisen, mit deren Hilfe wir Antworten auf Fragen aus dem Gegenstandsbereich* der Psychologie erhalten können.

Die Forschungsmethoden sind das Fundament der Psychologie, falsche Forschungsmethoden führen zu falschen Schlüssen.

## **1.2 Wofür braucht man das?**

* Um eigene Fragestellung untersuchen zu können.
* Um Sachverhalte und Merkmalsunterschiede genau beschreiben und erklären zu können
* Um Wissen zu erweitern, welches für effektives Intervenieren notwendig ist
* Um Wissen zu erweitern, womit man Fortschritte erlangen kann (Erziehungstipps, ...)
* Um im Alltag Informationen sammeln und beurteilen zu können
* Um die Arbeiten anderer verstehen und beurteilen zu können

# **2. Psychologie als empirische Wissenschaft**

Empirisch kommt aus dem Griechischen und bedeutet “auf Erfahrung begründet”.

**Definition**: Eine empirische Wissenschaft ist daran interessiert, Hypothesen und Theorien zu den Fragen zu entwerfen, mit denen man sich gerade beschäftigt. Diese werden nun mit der Realität konfrontiert. Man vergleicht also die gedankliche Antwort auf die Frage mit den in der Realität diesbezüglich vorfindbaren Sachverhalten.

## **2.1 Alltagspsychologie vs. wissenschaftliche Psychologie**

Die Grundfragen der alltäglichen und der wissenschaftlichen Psychologie unterscheiden sich nicht wesentlich, da sie den gleichen Interessengegenstand haben. Der Unterschied ist, wie diese Fragen nun beantwortet werden.

### 2.1.1 Alltagspsychologie

Die Wahrnehmung in der Alltagspsychologie ist selektiv, nicht wertfrei und hängt von Kontext und den Hypothesen ab. Sowohl die Wahrnehmung als auch die zugrundeliegenden Hypothesen werden im Alltag häufig nicht reflektiert, sondern intuitiv übernommen.

**Kritik**: Die Begriffe der Alltagspsychologie sind meist ungenau & inkonsistent. Die Erklärungen sind meist unvollständig und es findet keine systematische Überprüfung statt.

### 2.1.2 Wissenschaftliche Psychologie

Systematische Forschung führt zu Überwindung von Wahrnehmungsverzerrungen und präziser Realitätserfassung. Dafür werden wissenschaftliche Methoden benutzt.

2 Wege der Erkenntnis: induktives oder deduktives Vorgehen

**Induktion**: von wiederholten Einzelbeobachtungen oder sorgfältiger Einzelfallanalyse wird eine generelle Regel für die Allgemeinheit gemacht.

Problem: Induktive Schlüsse haben nur Wahrscheinlichkeitscharakter und lassen sich nicht richtig begründen und beweisen.

**Deduktion**: Von einer Regel für die Allgemeinheit wird auf den Einzelfall geschlossen.

## **2.2 Exkurs: Wissenschaftstheorie**

Bei der Wissenschaftstheorie handelt es sich um jene Disziplin, die sich mit dem Begriff und der Einteilung von Wissenschaften, ihren Erkenntnisprinzipien und Methoden, sowie ihrer Sprache beschäftigt. Ziel der Wissenschaftstheorie liegt «in der Klärung wissenschaftlicher Begriffe, Aussagen, Methoden und Theoriebildungen».

Die für die Psychologie wichtigsten wissenschaftstheoretischen Ansätze sind:

* **Logik** (Lehre des vernünftigen Schlussfolgerns)
* **Empirismus** (Wissenschaftliche Erkenntnisse sind auf Beobachtung und Experiment zurückzuführen)
* **Phänomenologie** (Wissenschaft, Lehre, die von der geistigen Anschauung des Wesens der Gegenstände oder Sachverhalte ausgeht, die die geistig-intuitive Wesensschau (anstelle rationaler Erkenntnis) vertritt.)
* **Kritischer Rationalismus** (Deduktion wird als einzig mögliche Vorgehensweise angesehen (Mit Ausnahme der Theoriebildung, wo auch Induktion erlaubt ist). Alles ist theoriegeleitet und aus der Theorie werden Hypothesen abgeleitet, die empirisch prüfbar sind. Der kritische Rationalismus besagt auch, dass nur Falsifikation möglich ist, aber nie Verifikation.)
* **Konstruktivismus** (Lehre, die den konstruktiven Aufbau besonders der Mathematik und der Logik vertritt)
* **Strukturalismus** (Sammelbegriff für interdisziplinäre Methoden und Forschungsprogramme, die Strukturen und Beziehungsgefüge in den weitgehend unbewusst funktionierenden Mechanismen kultureller Symbolsysteme untersuchen.)
* **Systemtheorie** (soziologische Theorie, durch die Gesellschaft als ein „umfassendes soziales System, das alle anderen sozialen Systeme in sich einschließt“ beschrieben und erklärt wird.)

## **2.3 Historischer Hintergrund**

Als Geburtsstunde der wissenschaftlichen Psychologie gilt die Gründung des ersten Instituts für experimentelle Psychologie im Jahr 1879 in Leipzig durch **Wilhelm Wundt**. Die experimentelle Psychologie nach Wundt hatte einen naturwissenschaftlichen Ansatz, welche das Experiment als das Paradigma der psychologischen Forschung ansah. Die Datenerhebung in diesen Experimenten kam aus Introspektion.

Aus diesen ersten methodischen Ansätzen entwickelten sich die quantitativen Methoden.

Die wesentlichen Merkmale experimentellen Vorgehens sind

* Objektive Beschreibbarkeit seelischer Vorgänge
* Willkürliche Herbeiführbarkeit und Veränderbarkeit seelischer Vorgänge
* Wiederholbarkeit experimenteller Studien

**Wilhelm Dilthey**: Begründer der Erkenntnistheorie der Geisteswissenschaften und als einer der Hauptvertreter der hermeneutischen Wissenschaften. Er wollte statt Wundts experimentellen Ansatz einen verstehenden, ganzheitlichen, hermeneutischen Ansatz sowie den Fragestellungen angemessene Methoden.

Aus seinen Ansätzen entstanden die qualitativen Methoden.

**Erkenntnistheorie**: Beschäftigt sich mit der Herkunft & Zustandekommen von Wissen

**Hermeneutik**: Theorie der Interpretation von Texten und des Verstehens

## **2.4 Qualitative vs. Quantitative Methoden**

|  |  |
| --- | --- |
| **Qualitative Methoden** | **Quantitative Methoden** |
| **Ziel**: Entwicklung von Konzepten, um das Erleben und Empfinden von Personen beschreiben zu können. | **Ziel**: Allgemeingültige Aussagen, die für die meisten Personen zutreffen |
| **Vorgehen**: Befragung / Beobachtung / Diskussion einzelner Personen / Gruppen | **Vorgehen**: Formulierungen von theoriegeleiteten Hypothesen und experimentelles Vorgehen / Befragung / Beobachtung / Messung vieler Personen unter möglichst gleichen Bedingungen |
| **Auswertung** dieser subjektiven Ergebnisse | **Auswertung**: über standardisierte (statistische) Auswertungsverfahren |
| **Interpretation** dieser Aussagen durch einzelne Forscherperson und evtl. Erstellung theoretischer Annahmen | **Interpretation**: Abgleich der Ergebnisse mit den Hypothesen |

## 

## **2.5 Begriffserklärungen: Variablen, Operationalisierungen**

**Variablen** sind veränderliche Beobachtungsgrössen. Jede Variable hat mindestens zwei Ausprägungen.

**Konstanten** sind Beobachtungsgrössen mit nur einer Ausprägung.

**Qualitative vs. Quantitative Variable**: Merkmalsausprägungen unterscheiden sich bei quantitativen Variablen nach dem Zahlenwert, bei qualitativen Variablen nach ihrer Beschaffenheit (Qualität).

**Konkrete (manifeste) vs. abstrakte (latente) Variablen**: Konkrete Variablen sind direkt messbar. (Bsp. Kopfumfang, Körpergrösse). Abstrakte Variablen sind hypothetische Konstrukte, die zwar nicht direkt messbar sind, aber trotzdem real sind (Bsp.: Intelligenz).

**Operationalisierung**: Messbarmachung der Variablen. Dazu weist man den Variablen konkret mess- bzw. beobachtbare Grössen zu. Es muss darauf geachtet werden, dass der Bedeutungskern mit der Operationalisierung getroffen wird.

## **2.6 Basisziele der Psychologie**

### 2.6.1 Beschreiben

**Definition**: Beim Beschreiben werden Angaben über die Erscheinungsformen und Merkmale von mindestens einem Sachverhalt gemacht. Dies geschieht durch Benennen, Ordnen und Klassifizieren, sowie Angaben zu Häufigkeit bzw. Ausprägungsgrad.

Durch das Beschreiben zweier Merkmale können *Zusammenhangsrelationen* gefunden werden. ACHTUNG: Zusammenhangsrelationen sind keine *Kausalrelationen*!

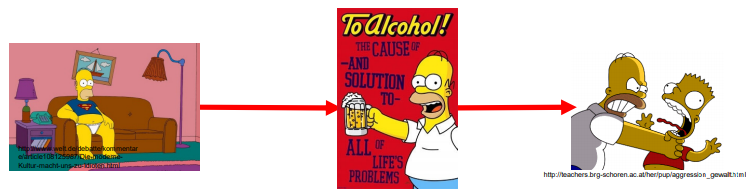
Beschreiben ist der erste Schritt, reicht aber noch nicht aus. Die Frage nach der Ursache wird erst durch das Erklären geklärt.

### 2.6.2 Erklären

**Definition**: Erklärungen sind Angaben über Bedingungsverhältnisse von Sachverhalten bzw. Angaben über Abhängigkeiten zwischen Sachverhalten. Erklärungen setzen die Beschreibung von mindestens zwei Sachverhalten voraus. Erklären ist also die Suche nach einem Ursache-Wirkungs-Zusammenhang / Kausalrelation. Kausalrelationen sind nur durch Experimente überprüfbar.

**Erklären 1. Ordnung**: Direkter Zusammenhang zwischen beobachteter Variable 1 und beobachteter Variable 2.

**Erklären 2. Ordnung**: Der Zusammenhang wird durch eine vermittelnde Variable (Mediator) erklärt.



### 2.6.3 Vorhersagen

Vorhersagen (=Prognosen) sind vorwärts gerichtete Erklärungen. Wenn nach Sachverhalt A Sachverhalt B folgt, kann bei Eintreten von Sachverhalt A Sachverhalt B vorhergesagt werden.

Prognosemodelle sollten theoriegeleitet sein und Kompromisse zwischen Wahrscheinlichkeiten und deterministischen Zusammenhängen finden.

**Determinismus**: Der Determinismus ist die Auffassung, dass alle – insbesondere auch zukünftige – Ereignisse durch Vorbedingungen eindeutig festgelegt sind.

**Bedingungen für die Prognosewahrscheinlichkeit**

* Präzision der Beschreibung der am Prognosemodell beteiligten Sachverhalte
* Adäquate Auswahl der theoriegeleiteten Prädiktoren
* potentielle Gewichtung der Prädiktoren
* Zeitraum der Prognose

### 

### 2.6.4 Verändern

**Verändern als Korrektur**: Ausgangszustand gilt als “problematisch”, die Intervention soll folglich in erster Linie etwas Negatives aufheben und nicht so sehr etwas Positives entwickeln.

**Verändern als Förderung**: Besserer Zustand wird angestrebt, ohne dass Ausgangszustand als problematisch gilt.

**Verändern als Prävention**: Intervention zur Verhinderung eines schlechten Zustands.

**3 Phasen des Veränderns**

**Diagnostik**: Die psychologische Diagnostik repräsentiert Vorgehensweisen, welche eine Erfassung von Charakteristika von Personen, Personengruppen, Institutionen, Situationen, etc. zur Folge haben. Die Erfassung und Gewinnung von Charakteristika erfolgt zielgerichtet und systematisch mit wissenschaftlich fundierten Methoden, wie Testverfahren, Fragebogen, Verhaltensbeobachtungen und Anamnesen (Erfragen von wichtigen Informationen durch Fachpersonal). Mit der Diagnostik wird das Ziel verfolgt, Erkenntnisse über die Merkmalsträger (Probanden, Klienten, Patienten) zu gewinnen und für eine Entscheidung über eine nachfolgende Massnahme, wie Beratung, Therapie, Training, etc., zu nutzen.

Bsp.: psychologische Tests, Fragebögen, diagnostische Interviews, Verhaltensbeobachtung.

**Intervention**: Unter einer Intervention versteht man in der Psychologie geplant und gezielt eingesetzte Massnahmen, um Störungen vorzubeugen ([primäre] Prävention), sie zu beheben ([sekundäre Prävention] Psychotherapie) oder deren negative Folgen einzudämmen ([tertiäre Prävention] Rehabilitation).

Intervention findet sich im klinisch-therapeutischen / präventiven Bereich, aber auch in Form von randomisierten, kontrollierten Interventionsstudien in der Forschung.

**Evaluation**: Unter Evaluation ist die systematische, empirische Analyse von Konzepten, Bedingungen, Prozessen und Wirkungen zielgerichteter Aktivitäten zum Zwecke ihrer Bewertung und Modifikation zu verstehen.

**Fünf Funktionen der Evaluationsforschung**

1. Erkenntnisfunktion (Erkenntnisgewinn)
2. Lern- und Dialogfunktion (z.B. Erkennen von Problemen bei Dialogen in Firmen, ...)
3. Optimierungsfunktion (Optimieren von Methoden)
4. Entscheidungsfunktion (Hilfe bei Entscheidungen über nächste Schritte)
5. Legitimationsfunktion (Rechtfertigung der zuvor getätigten Schritte)

# 

# **3. Quantitative Erhebungsmethoden**

## **3.1 Forschungsidee / Forschungsfrage finden**

Bei der Suche nach einer Forschungsidee finden sich zwei Herausforderungen:

1. Eine geeignete Forschungsidee zu finden
2. Bewertung der Forschungsidee bezüglich wissenschaftlicher und praktischer Relevanz, sowie ihrer empirischen Untersuchbarkeit

**Forschungsidee finden**

* Kennenlernen des bisherigen Wissenstands durch lesen von Primär- und Sekundärliteratur
* Bibliotheken, Fachdatenbanken durchsuchen
* Idee mit bestehender Forschung abgleichen

**Bewertung der Forschungsidee**

*Wissenschaftliche Relevanz*: Welchen Nutzen hätte eine Forschung für die Wissenschaft? Bestehen in diesem Bereich bereits Studien?

*Praktische Relevanz*: Kann das Experiment überhaupt durchgeführt werden? Werden zu grosse Stichproben benötigt?

*Empirische Untersuchbarkeit*: Kann das Experiment überhaupt gemessen werden? Sind die zu messenden Variablen überhaupt operationalisierbar und hält das Experiment die ethischen Richtlinien ein?

## **3.2 Ethische Richtlinien**

### 3.2.1 Ethische Richtlinien im Umgang mit VPN

**Milgram-Experiment**

Cover-Story: Studie zur Lernleistung unter Bestrafung

Eigentliche Studie: Messen, wie weit sich Personen dazu drängen lassen, einer anderen Person Schaden zuzufügen

Drei Beteiligte Personen: Versuchsleiter, Schüler (Schauspieler, der in das Experiment eingebunden ist), Lehrer (VPN)

Schüler soll Assoziationspaare lernen. Bei einem Fehler sollte der Lehrer ihn mit einem Stromschlag bestrafen. Der Schüler zeigt offensichtlich Schmerzen und weist auf Kreislaufprobleme hin, wodurch der Lehrer das Experiment beenden. Der Versuchsleiter versucht nun, ihn zum Weitermachen zu zwingen.

**Heute gängige ethische Grundsätze**

1. Die Gewährleistung der psychischen und physischen Unversehrtheit und Integrität der VPN
2. Die Transparenz der Untersuchungssituation
3. Vermeidung von Täuschungen
4. Freiwillige Untersuchungsteilnahme
5. Vertraulichkeit der Untersuchungsergebnisse (Datenschutz)
6. Information über die Untersuchung nach Abschluss (Debriefing)
7. Einverständniserklärung
8. VPN kann das Experiment jederzeit ohne einen Grund angeben zu müssen abbrechen.
9. Die VPN kann jederzeit die Löschung ihrer Daten verlangen

**Cover-Stories**

Cover-Stories können dazu führen, dass die VPN an einem Experiment teilnimmt, an dem sie bei Bekanntheit des Untersuchungsziel nicht teilgenommen hätte. Deshalb sind Cover-Stories nur erlaubt, wenn sie absolut notwendig sind. Am Ende des Experiments muss die VPN unbedingt über die Täuschung aufgeklärt werden (Debriefing). Sie muss erneut darauf aufmerksam gemacht werden, dass sie jederzeit das Recht hat, grundlos die Löschung ihrer Daten zu verlangen.

Cover-Stories sind nur erlaubt, wenn…

1. Die Untersuchung des interessierenden Phänomens ohne Verschleierung nicht möglich ist
2. Die Studie einen bedeutsamen Beitrag zum wissenschaftlichen Erkenntnisgewinn leistet
3. Die Täuschung nicht deutlichen Schaden oder emotionalen Stress für die Teilnehmenden mit sich bringt
4. Wenn die Cover-Story nicht dazu missbraucht wird, im VPN für ein unangenehmes Experiment zu gewinnen, dass sie sonst abgelehnt hätten

### 3.1.2 Ethik im Hinblick auf wissenschaftliche Kriterien

**Umgang mit Daten**

-Fälschen von Daten absolutes No-Go!

-Falsche Anreizstrukturen: Für die wissenschaftliche Karriere ist die Anzahl der Publikationen ausschlaggebend. Deshalb tendiert man dazu, nur spannende Ergebnisse überhaupt zu veröffentlichen.

**Questionable Research Practices (QRP)**

1. **Outcome switching**: Erhebung mehrerer AVs und selektiven Berichten derjenigen, die signifikant wurden.
2. **p-hacking**: Rechnen mehrerer Analysen und selektives Berichten derjenigen, die signifikant wurden
3. **HARKing**: Hypothese erstellen, nachdem die Resultate bekannt wurden. Darstellung von explorativen Ergebnissen als konfirmatorisch
4. **Optional stopping**: Analyse der Daten während der Erhebung und Stoppen der Erhebung bei signifikantem Ergebnis

QRPs erhöhen die Wahrscheinlichkeit eines Fehlers 1. Art dramatisch!

**Publication Bias**: Studien mit neuen, spannenden Befunden haben bessere Chancen auf Veröffentlichung (als z.B. Replikationsstudien).

**File Drawer**: Da i.d.R. nur signifikante Ergebnisse publiziert werden gibt es eine Dunkelziffer an durchgeführten Studien, die aufgrund nicht-signifikanter Ergebnisse nie veröffentlicht wurden. Dies führt bei Meta-Analysen dazu, dass keine zuverlässige Schätzung der wahren Effektstärken möglich ist. Es gibt zwar Methoden zur Korrektur des Publication Bias, diese sind aber nicht einfach anzuwenden.

**Geringe Power von Psychologischen Studien**: Psychologische Test haben eine sehr niedrige Power (=Teststärke), also Wahrscheinlichkeit einen Effekt zu entdecken, wenn er tatsächlich in der Population existiert. Da die Power von Effektstärke, Stichprobengrösse und Alpha-Level abhängig ist, kann die Power nur mit mehr VPN verbessert werden. Da aber durchschnittlich nur etwa 40 Personen gefunden werden können, beträgt die Power bloss 34%. Dadurch ergeben sich viele nicht informative Studien.

**Umgang mit Quellen: Richtig zitieren**

* Wenn etwas als Tatsache formuliert wird, immer die Quelle (Referenz) dazu angeben
* Wenn eine Idee aus einer anderen Quelle in eigene Worte gefasst wird (paraphrasieren) 🡪 Quelle angeben
* Zitate immer als solche Kennzeichnen
* Nicht aus anderen Quellen abschreiben / wörtlich übersetzen (ausser für ein wörtliches Zitat), sondern umformulieren und Quelle angeben

**Plagiat**: Ganze oder teilweise Übernahme eines fremden Werks ohne Angabe der Quelle des Urhebers bzw. der Urheberin zu verstehen

### 3.1.3 Open Science

**Definition:** Der Begriff Open Science bündelt Strategien und Verfahren, die allesamt darauf abzielen, die Chance der Digitalisierung konsequent zu nutzen, um alle Bestandteile des wissenschaftlichen Prozesses über das Internet offen zugänglich und nachnutzbar zu machen. Damit sollen Wissenschaft, Gesellschaft und Wirtschaft neue Möglichkeiten im Umgang mit wissenschaftlichen Erkenntnissen eröffnet werden.

**Funktionen von Open Science**

* Nachvollziehbarkeit und Reproduzierbarkeit (Open Methodology)
* Bewertung von Qualität und Reputation (Open Peer Review, Open Source)
* Nachnutzbarkeit
* Öffnung (Open Access)
* Vernetzung und Integration (Open Data)
* Verbreitung (Open Access, Open Educational Ressources)
* Finanzierung

**Präregistrierung**

Verbesserung der Forschung, durch

* Klare Unterscheidung von konfirmatorischer vs. Explorativer Forschung
* Entgegenwirkung von Publication Bias
* Verhinderung der meisten Formen von questionable research practices / p-hacking

**Open Data**: Open Data sind Daten, die von jedermann frei verwendet, nachgenutzt und verbreitet werden können – maximal eingeschränkt durch Pflichten zur Quellennennung. Davon ausgenommen sind personenbezogene Daten. Die wichtigsten Eigenschaften sind Verfügbarkeit und Zugang als Ganzes, Wiederverwendung und Nachnutzung und universelle Beteiligung.

**Open Access**: Die Literatur sollte kostenfrei und öffentlich im Internet zugänglich sein, sodass alle, die daran interessiert sind, sie nutzen können. Die einzige Einschränkung sollte darin bestehen, dass die Arbeiten den zugehörigen Autoren anerkannt werden und sicher zu stellen, dass sie korrekt zitiert werden.

**Argumente für Open Access**

* Freier, schneller, kostenloser Zugang zu öffentlich finanzierten Forschungsergebnissen
* Erhöhte Sichtbarkeit und Zitierhäufigkeit, gute Auffindbarkeit über Suchmaschinen und Nachweisdienste
* Vollständige Ausschöpfung der Vorteile digitaler Dokumente
* Förderung internationaler und interdisziplinärer Zusammenarbeit
* Förderung der Forschungseffizienz
* Langfristige Verfügbarkeit der Dokumente
* Prioritätssicherung

**Vorbehalte gegen Open Access**

* Keine angemessene Anerkennung von Open Access-Beiträgen führt zu Nachteilen für die wissenschaftliche Laufbahn
* Garantie der langfristigen und guten Auffindbarkeit fragwürdig
* Rechtliche Vorbehalte: Kontrolle über Nutzung und urheberrechtliche Verwertungsrechte
* Kosten? Finanzierbarkeit des Author-Pays-Modells
* Publikationsmittelverteilung und Interessenkonflikte
* Zeitaufwand

## **3.2 Hypothesen formulieren**

**Definition**: Die Forschungshypothese wird aus etablierten Theorien und/ oder gut gesicherten empirischen Befunden abgeleitet und postuliert die Existenz, Richtung und Stärke eines bestimmten Effekts. Anhand von Hypothesenprüfungen werden bestehende Theorien getestet und weiterentwickelt.

**Merkmale von wissenschaftlichen Hypothesen**

* Präzise und widerspruchsfreie Formulierung (impliziter Konditionalsatz)
* Prinzipielle Widerlegbarkeit (Falsifizierbarkeit)
* Operationalisierbarkeit
* Begründbarkeit
* Positive Formulierung über Existenz\*

\*Hypothesen, dass etwas nicht existiert, sind nicht testbar! Solche Hypothesen sollten also nie aufgestellt werden.

**Falsifikation, Verifikation**

Falsifikation bedeutet nicht falsch, Verifikation bedeutet nicht richtig!

Falsifikation bedeutet Zurückweisung der Hypothese, Verifikation Beibehaltung der Hypothese. Zur annähernden Bestätigung einer Hypothese sind viele (Replikations-)Studien notwendig. Eine Hypothese ist nie komplett bestätigt.

**Arten von Hypothesengültigkeit**

* **Universelle Hypothese**: hat generellen Gültigkeitsanspruch. Ein einziger gegenteiliger Fall reicht aus, um sie zu falsifizieren. Sie ist nie endgültig bestätigbar.
* **Beschränkt universelle Hypothese**: Einschränkung für die Gültigkeit vorhanden (personen-, situations-, zeitbezogen)
* **Quasiuniverselle Hypothese**: Wahrscheinlichkeitsaussagen, Regelhaftigkeiten. Diese Hypothesenart hat keinen Anspruch auf uneingeschränkte Geltung und ist in der Psychologie am häufigsten.

**Zeitpunkt der Hypothesenformulierung**

* **Hypothesenprüfende Untersuchung (konfirmatorisch)**: Hypothesen werden vorab formuliert. Konfirmatorische Untersuchungen liefern eine geprüfte Aussage und eine Entscheidung über Annahme oder Zurückweisung der Hypothese ist möglich.
* **Hypothesengenerierende Untersuchung (explorativ)**: Hypothesen werden im Nachhinein generiert. Explorative Untersuchungen liefern ungeprüfte Hypothesen und eine Entscheidung über Annahme oder Zurückweisung der Hypothese ist nicht möglich. Dies ist erst nach einer Überprüfung der Hypothese durch eine konfirmatorische Hypothese möglich.

**Operationalisierungen und Hypothesen**

Durch das Operationalisieren der Hypothese ergeben sich 2 Hypothesenebenen:

* Die **theoretisch-inhaltliche Hypothese (TIH**): nicht operationalisierte Form der Hypothese. Bsp.: Fluchen reduziert die Empfindung von Schmerz.
* Die **empirisch-inhaltliche Hypothese (EIH)**: operationalisierte Form der Hypothese Bsp.: VPN sollten eine höhere Schmerztoleranz haben, wenn sie fluchen, wie wenn sie nicht fluchen. & VPN sollten weniger stark auf Schmerz reagieren, wenn sie fluchen, wie wenn sie nicht fluchen.

Die empirisch-inhaltliche Hypothese ist direkt messbar während die theoretisch-inhaltliche Hypothese nur teilweise direkt messbar ist.

**Hypothesenarten bezüglich Unterschieden**

1. **Zusammenhanghypothese**: Es wird ein Zusammenhang zwischen zwei Variablen vermutet.
2. **Unterschiedshypothese**: Es wird ein Unterschied vermutet.
3. **Veränderungshypothese**: Es wird eine Veränderung einer abhängigen Variable im Laufe der Zeit vermutet.

**Gerichtete vs. ungerichtete Hypothesen**

Gerichtet: besser/schlechter, höher/niedriger, …

Ungerichtet: Es gibt einen Unterschied, aber die Richtung (besser/ schlechter, höher/ niedriger) ist unklar.

## **3.3 Definition und Messung der Variablen**

**Vorteile von quantifizierten Informationen**

1. Präzision
2. Vergleichbarkeit
3. Verknüpfbarkeit mit einfachen Operationen
4. Übersichtlichkeit und Sparsamkeit der Zusammenfassung

**Datenerhebungsverfahren**

* Beobachten, Zählen, Messen
* Selbstbericht
* Psychologische Tests
* (biopsychologische, neurowissenschaftliche Messungen)

### 3.3.1 Probleme des Selbstberichts

* Zugänglichkeit
* Verzerrungen
* Reaktivität

**Definition Reaktivität**: Veränderung bzw. Verzerrung der erhobenen Daten alleine aufgrund der Kenntnis der untersuchten Personen darüber, dass sie Gegenstand einer Untersuchung sind.

**Hawthorne-Effekt**

Einfluss, den die bloße Teilnahme an einem Experiment auf die Versuchsperson und damit auf das Experimentsergebnis auszuüben vermag.

**Massnahmen zur Reduzierung von Verzerrungen**

* Cover-Story
* Konkrete Anweisung auf Nichtverzerrung zu achten
* Auf Anonymisierung hinweisen
* Untersuchte in Unkenntnis darüber lassen, dass sie untersucht werden
* Masse einsetzen, die die VPN nicht kontrollieren können

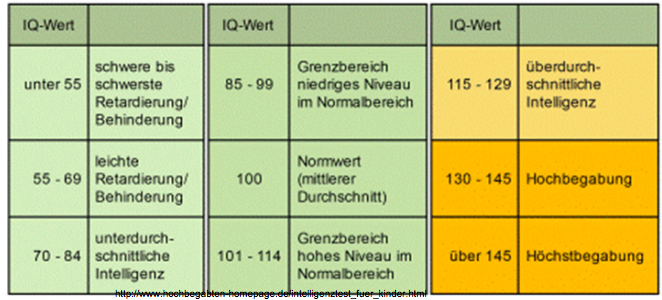
### 3.3.2 Quantitative Gütekriterien

#### 3.3.2.1 Objektivität

Die Objektivität (=Anwenderunabhängigkeit) einer Untersuchung / eines Tests / eines Fragebogens gibt an, in welchem Ausmass die Testergebnisse von den Testanwendern unabhängig sind.

**3 Arten von Objektivität**

* **Durchführungsobjektivität**: Die Durchführung des Experiments findet bei allen Testpersonen genau gleich statt, es wird keine zusätzliche Hilfestellung gegeben, wenn eine Person nicht weiterkommt, wenn nicht jede andere Person genau die gleiche Hilfe ebenfalls bekommt.
* **Auswertungsobjektivität**: Die Bewertung von Richtig und Falsch findet bei allen VPN genau gleich statt. Eine allfällige Punktewertung ist bei allen genau gleich.
* **Interpretationsobjektivität**: Interpretation findet anhand von Normwerttabellen statt (Bsp.: IQ-Stufen)



Objektivität wird erhöht durch Standardisierung von Durchführung, Auswertung und Interpretation der Untersuchung / des Tests / des Fragebogens.

🡪 Instruktionen im Testhandbuch / Manual / Handelsanweisung

🡪 einfach zu erreichendes Gütekriterium

Die Objektivität ist die Voraussetzung für die weiteren Gütekriterien

#### 3.3.2.2 Reliabilität

Die Reliabilität gibt an, wie gering oder stark ein Test durch Messfehler verzerrt ist. Sie ist ein Synonym zu Zuverlässigkeit, Präzision und Messgenauigkeit an.

Die Reliabilität basiert auf Annahmen der Klassischen Testtheorie:

- Testwert X = wahrer Wert T + Messfehler E

- Reliabilität umso höher, je kleiner der zu X gehörende Messfehler E.

- Für die perfekte Reliabilität beträgt X=T.

- Die **Fehlervarianz** ist die unsystematischen Abweichungen von wahren Werten

Reliabilität = Anteil wahre Varianz / beobachtete Varianz

Reliabilität > 0.7 akzeptabel

**Reliabilitätsarten**

1. Test-Retest-Reliabilität
2. Paralleltest-Reliabilität
3. Testhalbierungs-Reliabilität
4. Interne Konsistenz
5. Interrater-Reliabilität

**Test-Retest-Reliabilität**

=Ausmass der Übereinstimmung bei einer wiederholten Anwendung des Tests / Fragebogens / der Untersuchung bei der gleichen Stichprobe. Höhere Korrelation der Testwerte des Tests A und der des Tests B bedeutet hohe Test-Retest-Reliabilität.



Bei stabilen Merkmalen führt eine reliable Testung bei Wiederholung unter gleichen Bedingungen zu gleichen / sehr ähnlichen Ergebnissen.

Probleme: Erinnerungseffekte, aufwendig, nicht geeignet bei instabilen Merkmalen

Häufige Quellen für Messfehler: Beobachtungsfehler (Unkonzentriertheit, Flüchtigkeitsfehler, …), Umweltveränderungen, Veränderungen der Teilnehmenden.

**Paralleltest-Reliabilität**

Übereinstimmung zweier Versionen (Äquivalenz) des gleichen Tests innerhalb einer Stichprobe. Hohe Korrelation der Testwerte von Version A und Version B führt zu hoher Paralleltest-Reliabilität. Paralleltests finden Anwendung bei Gruppentests im Leistungsbereich oder bei wiederholten Testungen gleicher Personen.

Probleme: Sehr aufwendig in der Entwicklung.

**Testhalbierungs-Reliabilität**

Die Testhalbierungsreliabilität bezieht sich auf die Korrelation zwischen zwei Hälften des gleichen Tests. Die Items müssen dafür aber homogen sein, also alle das gleiche Merkmal messen!

Ein durchgeführter IQ-Test kann etwa in zwei Hälften unterteilt werden, zu denen sich für die untersuchten Personen jeweils ein Kennwert bestimmen lässt. Wenn die einzelnen Personen in den beiden Testhälften ähnlich hohe Kennwerte erzielen, dann besitzt der Gesamttest eine hohe Testhalbierungsreliabilität.

Vorteilhaft an der Reliabilitätsbestimmung über die Testhalbierung im Vergleich zur Retest- oder Paralleltestmethode ist der geringere Untersuchungsaufwand.

Die Reliabilität wird durch die Testhalbierung unterschätzt, da nur noch halb so viele Items pro gemessene Reliabilität vorhanden sind. Dies kann mit der Spearman-Brown-Prophecy-Formula korrigiert werden.

**Spearman-Brown-Prophecy-Formula**



where *n* is the number of "tests" combined (see below) and *p* is the reliability of the current "test". The formula predicts the reliability of a new test composed by replicating the current test *n* times (or, equivalently, creating a test with *n* parallel forms of the current exam). Thus *n* = 2 implies doubling the exam length by adding items with the same properties as those in the current exam. Values of *n* less than one may be used to predict the effect of shortening a test.

**Interne Konsistenz**

Erweiterung der Testhalbierung. Einteilung des Tests in kleinste Einheiten (Items). Mit jedem Item wird nun ein Paralleltest zu den anderen Items durchgeführt. Korrelierende Ergebnisse führen zu hoher Reliabilität. Das gebräuchlichste Mass für die interne Konsistenz ist **Cronbach’s Alpha.** Die interne Konsistenz ist ein Indikator für die Homogenität eines Tests.

**Vor- und Nachteile der verschiedenen Reliabilitätsarten**

|  |  |  |  |
| --- | --- | --- | --- |
| **Reliabilität** | **Vorteile** | **Nachteile** | **Besonderheiten** |
| Test-Retest | * Gleiche Testitems * Keine Extraarbeit bei der Entwicklung | * Erinnerungseffekte * Nicht geeignet für instabile Merkmale * Erfordert zwei Messzeitpunkte * Abhängig vom gewählten Zeitinterval | * Mass für die Stabilität eines Merkmals |
| Paralleltest | * Minimiert Wiederholungseffekte * In Gruppensettings anwendbar * Auch geeignet für Prä-Posttest-Designs | * Verwendung verschiedener Items verringert Reliabilität * Aufwendig in der Entwicklung | * Anwendung in Leistungstests in Gruppensettings |
| Test-halbierung | * Minimiert Wiederholungs-effekte * Unaufwendig | * Verwendung verschiedener Items verringert Reliabilität * Benötigt längeren Test * Geringere Anzahl Items verringert Reliabilität * Reliabilität auch abhängig von Art der Halbierung |  |
| Interne Konsistenz |  | * Bei mehrdimensionalen Tests kommt es zu Unterschätzungen | * Mass für Homogenität des Tests |

**Spezialfall: Interrater-Reliabilität**

Die Interrater-reliabilität ist die Höhe der Übereinstimmungen der Einschätzungsergebnisse unterschiedlicher Beobachter / Testanwender (Rater). Sie ist hoch, wenn verschiedene Rater bei den gleichen Testpersonen zu gleichen oder ähnlichen Einschätzungen kommen.

Berechnet wird sie als Prozent der Übereinstimmung, Cohen’s Kappa (Bei 2 Ratern) oder Krippendorff’s Alpha (Bei mehr als 2 Ratern).

**Cohen’s Kappa**



p0 ist der gemessene Übereinstimmungswert.

pC ist die zufällig erwartete Übereinstimmung.

1= 100% Übereinstimmung gemessen.

0=Übereinstimmung entspricht dem Ausmass des Zufalls.

#### 3.3.2.3. Validität

**Definition**: Der Testwertmisst tatsächlich das Merkmal, das er laut Testbezeichnung bzw. Testbeschreibung zu messen beansprucht und primär kein anderes.

Validität ist das wichtigste Gütekriterium. Allerdings setzt sie Objektivität und Reliabilität voraus und ihre Überprüfung ist aufwendig.

Validitätsarten

1. Inhaltsvalidität
2. Kriteriumsvalidität
   1. Übereinstimmungsvalidität
   2. Prognostische Validität / Vorhersagevalidität
3. Konstruktvalidität
   1. Konvergente Validität
   2. Diskriminante Validität
4. Interne / externe Validität von Untersuchungen

**Inhaltsvalidität**

Test erfasst das Zielmerkmal offensichtlich in den wichtigsten Aspekten. Die Repräsentativität der Items für das zu messende Merkmal sind gegeben.

Die Inhaltsvalidität kann nicht numerisch bestimmt werden. Sie wird durch subjektive Einschätzung oder ExpertInnenratings bestimmt. Diese Art der Validität ist vor allem anwendbar, wenn das Testverhalten das interessierende Merkmal direkt repräsentiert.

**Kriteriumsvalidität**

Kriteriumsvalidität bezieht sich auf den Zusammenhang zwischen den Ergebnissen des Messinstruments und einem empirischen Kriterium.

Ein Problem der Kriteriumsvalidität ist die Wahl eines geeigneten Aussenkriteriums zum Vergleich. Sollte auch hier kein eindeutiges Aussenkriterium vorhanden sein, so können mehrere benutzt werden. Die Kriteriumsvalidität kann populationsabhängig sein.

Übereinstimmungsvalidität

Das Außenkriterium, das bereits valide sein muss (z. B. ein anderer Test), wird zeitgleich mit dem zu validierenden Messinstrument den gleichen Versuchspersonen dargeboten. Die Ergebnisse der beiden Messinstrumente werden korreliert. Die Höhe der Korrelation ist das Maß für die Übereinstimmungsvalidität.

Prognostische Validität

Die Messdaten werden zu einem Zeitpunkt erhoben, der vor der Erhebung des Außenkriteriums liegt. Im Unterschied zur Übereinstimmungsvalidität liegt bei der Bestimmung der Vorhersagevalidität das Prognoseintervall zwischen den beiden Messungen. So kann der Grad bestimmt werden, in dem die Messdaten das Kriterium vorhersagen. Ein Test erfüllt die Vorhersagevalidität, wenn seine Vorhersagen hoch mit dem tatsächlich später eingetretenen Ergebnis korrelieren.

**Konstruktvalidität**

Unter Konstruktvalidität versteht man empirische Belege dafür, dass ein Test das Konstrukt erfasst, welches er erfassen soll – und nicht ein anderes. Konstruktvalidität ist ein Netz von Hypothesen, dass durch Konstrukte und Relationen zu anderen manifesten und latenten Variablen besteht.

Für eine hohe Konstruktvalidität benötigt man mehrere Indikatoren durch eine Reihe von Studien, die am besten mit unterschiedlichen Forschungsdesigns durchgeführt wurden. Die anderen Variablen, durch die man die Konstruktvalidität belegt, müssen mit anderen validen Tests gemessen werden. Zudem muss die abgeleitete Hypothese gültig sein.

Eine besondere Methode zur Konstruktvalidierung ist die Multitrait-Multimethode-Matrix. Dabei wird das Konstrukt durch konvergente und diskriminante Validität bestätigt.

**Konvergente Validität**: zwei verschiedenen Operationalisierungen des gleichen Konstrukts sollen hoch miteinander korrelieren.

**Diskriminante Validität**: niedrige Korrelation zwischen Zielkonstrukt und davon unterschiedlichem Konstrukt.

### 3.3.3. Beobachten

Wissenschaftliche Beobachtung ist die systematische und regelgeleitete Registrierung des Auftretens bzw. der Ausprägung von ausgewählten, psychologisch relevanten Merkmalen oder Ereignissen.

Eine wissenschaftliche Beobachtung ist nichtkommunikativ, selektiv, zielgerichtet und methodisch kontrolliert. Sie ist zudem standardisiert, dokumentiert und intersubjektiv überprüfbar.

Wissenschaftliche Beobachtungen folgen einem zuvor festgelegten Beobachtungsplan zu Beobachtungsorten, -zeiten, -objekten und -einheiten.

**Zeitstichprobe**: Die Beobachtungen sind in festen Intervallen aufgezeichnet. Es handelt sich um eine breite und umfassende Beschreibung des Gegenstands. Sie wird gewählt, wenn der Fokus auf der umfassenden Beschreibung des Verhaltens liegt.

**Ereignisstichprobe**: Der zeitliche Ablauf ist irrelevant. Es wird das Auftreten, die Auftretensdauer und die Auftretenshäufigkeit von definierten Ereignissen betrachtet. Mit der Ereignisstichprobe sind Ereigniskombinationen und seltene Ereignisse untersuchbar. Sie wird verwendet, wenn der Fokus auf der Beobachtung eines speziellen Verhaltens liegt.

#### 3.3.3.1. Arten der Beobachtung

* **Freie Beobachtung (=offene / unstrukturierte / qualitative Beobachtung)**
  + Keine Richtlinien
  + Erkundungscharakter
* **Teilstrukturierte Beobachtung**
  + Vorgaben durch offene Kategorien
  + Angebracht, wenn die Umstände / Ursachen für Verhalten unklar sind
* **Strukturierte Beobachtung**
  + Präzise Festlegung was, wie, wie lange beobachtet wird und wie interpretiert wird
  + Der zu beobachtende Gegenstand ist im Prinzip bekannt und lässt sich in einzelne Elemente zerlegen, die beobachtet werden sollen.

#### 3.3.3.2. Formen der Beobachtung

**Teilnehmende Beobachtung**: Beobachter ist Teil der Aktivität, ist in das Geschehen involviert

**Nicht-teilnehmende Beobachtung**: Beobachter ist nicht Teil der Aktivität, ist nicht in das Geschehen involviert

**Offene oder verdeckte Beobachtungen**: Es ist bekannt, dass der Beobachter beobachtet oder nicht.

|  |  |  |
| --- | --- | --- |
|  | Vorteile | Nachteile |
| Teilnehmend | -Besondere Einsicht in Untersuchungsgegenstand  -Geeignet für Erkundungsstudien im natürlichen Umfeld | -Zeitintensiv  -Reaktivität  -durch Teilnahme ggf. geändertes Verhalten  -fehleranfällig  -Häufig geringer Grad an Systematisierung |
| Nichtteilnehmend | -Beobachtung im natürlichen Kontext  -Volle Konzentration auf Beobachten und Protokollieren | -Zeitintensiv  -besondere Sorgfalt für Nichtbeeinflussung der Situation notwendig |
| Verdeckt | -Natürliches Verhalten wird nicht gestört | -Ethische Richtlinien werden nicht eingehalten |
| Offen | -Ethische Richtlinien werden eingehalten | -Das natürliche Verhalten kann gestört werden |

### 3.3.4 Messen

**Definition**: Messen ist die Zuordnung von Zahlen zu Objekten oder Ereignissen bezüglich der Ausprägung eines Merkmals / einer Eigenschaft so dass bestimmte Relationen zwischen den Zahlen vorhandene Relationen zwischen Objekten oder Ereignissen homomorph abbilden.

Eine homomorphe Abbildung bildet Relationen zwischen Objekten bzw. Ereignissen (dem empirischen Relativ) durch zugeordnete Zahlen (dem numerischen Relativ) so ab, dass die Objekte bzw. Ereignisse und die Zahlen im korrekten Verhältnis zu einander stehen.

Es werden vier Skalenniveaus unterschieden, um festlegen zu können, ob eine Messung eine homomorphe Abbildung leistet:

1. **Nominalskala** (Relation der Verschiedenheit; wenn die Ausprägung eines Merkmals zwar verschiedene Zustände hat, aber keine Rangordnung gemacht werden kann)

Bsp.: Blutgruppe, Geschlecht, Religionszugehörigkeit, ...

1. **Ordinalskala** (Relation der Rangordnung; wenn die Ausprägung eines Merkmals in eine Rangordnung gebracht werden kann)

Bsp.: Zufriedenheit mit einem Produkt (sehr gut>gut>mittel>schlecht>sehr schlecht), Dienstrang beim Militär (General > Major > Leutnant > Feldwebel > Unteroffizier > Gefreiter), ...

1. **Intervallskala** (Relation der Differenz; Abstände zwischen den Ausprägungen einer Eigenschaft können genau bestimmt werden, es wird mit Zahlen gearbeitet)

Bsp.: Temperatur über einen Tag (6:00 5°C, 9:00 10°C, 12:00 14°C, 15:00 16°C, 18:00 10° C, ...)

1. **Verhältnisskala** (Relation zwischen Merkmalsausprägungen; Es handelt sich im Grunde um eine Intervallskala, allerdings gibt es einen Nullpunkt)

Bsp.: Temperatur in Kelvin (Nullpunkt: 0K), Zeitdauer (Nullpunkt: 0s), Preis (Nullpunkt: kostenlos), Geschwindigkeit (Nullpunkt: Stillstand)

Diese Skalenniveaus sind wichtig für die Entscheidung, welche Art von Aussagen über empirische Verhältnisse gemacht werden dürfen und welche mathematischen / statistischen Verfahren anwendbar sind.

### 3.3.5 Zählen

Zählen: Häufigkeit beobachteter Ereignisse

Messen: feinere Erfassung der Ausprägung eines Merkmals

Wann nimmt man was? → Hängt von der Art des untersuchten Merkmals und der Fragestellung ab

### 3.3.6 Befragung

Die am häufigsten eingesetzte Erhebungsmethode ist der Selbstbericht. Dabei gibt die Person selbst Auskunft über psychologische relevante Sachverhalte.

Die verschiedenen Befragungsverfahren lassen sich nach den folgenden Kriterien unterteilen:

-mündlich / schriftlich

-standardisiert / unstandardisiert

-strukturiert / unstrukturiert (Reihenfolge der Fragen wird jedes Mal verändert)

-Einzel-/Gruppenbefragungen / Umfragen

#### 3.3.6.1 Schriftliche vs. Mündliche Befragung

|  |  |
| --- | --- |
| **Schriftliche Befragung** | **Mündliche Befragung** |
| Typischer in der quantitativen Forschung | Typischer in der qualitativen Forschung |
| Höherer Grad an Standardisierung | Häufig kein typischer vorgegebener Ablauf vorhanden |
| Meist geschlossene Fragen mit vorgegebenen Antworten | Häufig aufwändiger und kostenintensiver |
| Vorwissen zur Formulierung der Fragen nötig | Befragte sind oft häufiger bereit sich zu äussern |
| Weniger mit dem Problem der Reaktivität behaftet | Stärker mit dem Problem der Reaktivität behaftet |

**Wie gelangen Befragte zu Selbstauskünften?**

Es werden die folgenden Prozesse angenommen:

-Die Frage wird unterschiedlich interpretiert

-Abruf und die Konstruktion eines eigenen Urteils

-Übersetzung des Urteils in eine kommunizierte Auskunft

**Wie werden Selbstberichte kommuniziert?**

Selbstauskünfte in Befragungen sind intentionale Akte der Kommunikation. Daher muss neben der Informationsvermittlung auch die Motivation berücksichtigt werden.

**Zu beachtende Punkte bei Selbstberichterhebungen:**

-Wurde die Frage/ Themenstellung durch die Befragten korrekt interpretiert / verstanden, bzw. so wie es die Befragenden wollten?

-Wurden die Antworten auf die Fragen möglicherweise durch die Art der Befragung beeinflusst?

-Ist gewährleistet, dass die Befragten ihre Antworten überhaupt beantworten können?

-Es müssen die Motive der Befragten, die den Antworten zugrunde liegen, abgeschätzt werden.

**Tipps für Itemformulierungen**

-Einfache Formulierung und gute Verständlichkeit

-Keine zu hohen Anforderungen an die mentale oder kognitive Leistungsfähigkeit der Befragten

-Adressatenorientierte Formulierung

-Keine Verneinung in den Fragen

-Keine überfrachteten, uneindeutigen Fragen

-Keine “Forced Choice” bei unabhängig beantwortbaren Aspekten (Zwei Dinge, die nicht voneinander abhängig sind oder nichts miteinander zu tun haben (Bsp.: Was haben sie lieber: Game of Thrones oder Lakritze?))

-Keine Fragen, die alle Befragten sehr ähnlich beantworten

-Einsatz mehrerer Items zur Beantwortung einer Frage

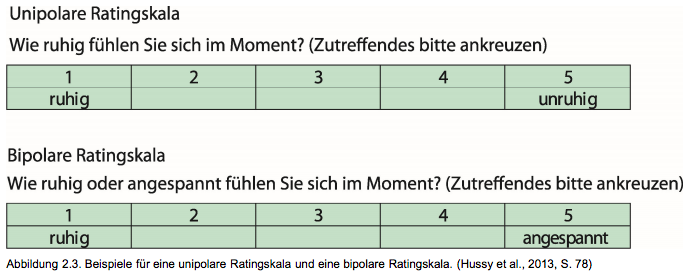
-Beachtung der Ausgewogenheit in der Reihenfolge der Fragen

-Eine klare und Informative Instruktion

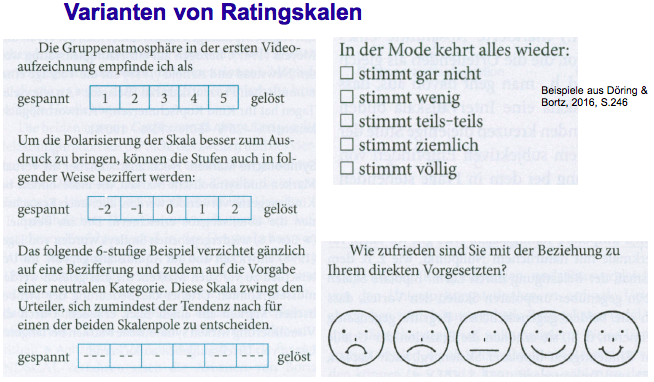
#### 3.3.6.2 Selbstberichtverfahren: Rating

-Frage oder Aussage

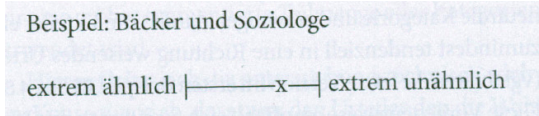
“Wie ruhig fühlen sie sich im Moment?” vs. “Ich fühle mich im Moment sehr ruhig”. Wie sehr trifft das auf sie zu?



*unipolar oder bipolar*



*Numerische, verbale, symbolische Marken*



*Grafisches Rating (visuelle Analogskala)*

##### 3.3.6.2.1 Ratingskalen: Anzahl der Skalenstufen

Geradzahlige oder ungeradzahlige Skalenstufe?

🡪 Ungerade: Es gibt eine neutrale Mittelkategorie

🡪 Gerade: Es wird eine Tendenz in eine Richtung erzwungen

**Ambivalenz-Indifferenz-Problem**: das Problem, dass bei Wahl einer neutralen Kategorie nicht klar ist, ob der Urteiler der Meinung ist, dass es keine Tendenz gibt, oder ob er sich nicht positionieren kann oder will.

Sollen viele oder wenige Abstufungen benutzt werden?

🡪 Viele Stufen führen zu einer besseren Differenzierbarkeit der Skala

🡪 Viele Stufen können die Befragten aber überfordern

🡪 Präferenz bei Befragten: 5-Stufige Skala

🡪 Stufenzahl zwischen 5 und 7 führt zu den validesten und reliabelsten Resultaten

Eine spezielle Form von Ratingverfahren ist das semantische Differenzial. Dabei werden charakteristische Polaritätsprofile durch bipolare Items, die eine schnelle Orientierung über zentrale Merkmale bzw. Unterschiede zwischen Merkmalsträgern erlauben.

**Probleme und Urteilsfehler beim Einsatz von Ratingskalen**

-Antworttendenzen

-Gedankenlose Reproduktion

-Primacy-Effekt

-Halo-Effekt (von gewissen bekannten Eigenschaften wird auf andere unbekannte Eigenschaften geschlossen)

### 3.3.7 Testen

Test(s) kann/können...

-Verfahren zur möglichst genauen quantitativen Erfassung einer Merkmalsausprägung bei Individuen

-Vorgang der Durchführung einer Untersuchung

-Gesamtheit der zur Durchführung notwendigen Requisiten

-Jede Untersuchung, sofern sie Stichprobencharakter hat

-Mathematisch-statistische Prüfverfahren

...sein

**Definition**: „Ein Test ist ein wissenschaftliches Routineverfahren zu Untersuchung eines oder mehrerer empirisch unterscheidbarer (Persönlichkeits-)Merkmale mit dem Ziel einer möglichst genauen quantitativen Aussage über den relativen Grad der individuellen Merkmalsausprägung.

Ein Test besteht in der Regel aus mehreren Aufgaben oder Fragen (Items), die von verschiedenen Menschen mit unterschiedlichen Fähigkeiten oder Eigenschaften unterschiedlich gelöst bzw. beantwortet werden.

In einem abstrakten methodischen Sinn wird ein Test auch als eine standardisierte Verhaltensstichprobe definiert, die aus Antworten auf eine Mehrzahl von Items besteht. Aus den Antworten wird der Testwert der untersuchten Person aggregiert.“

#### 3.3.7.1 Testarten: Leistungs- und Persönlichkeitstests

**Leistungstests**

-Aufgaben richtig oder falsch lösbar

-Speedtests

-Powertests1

**Arten von Leistungstests**

-Intelligenztests

-Entwicklungstests

-Schultests

-Allgemeine Leistungs- und spezielle Funktions- und Eignungstests

**Persönlichkeitstests**

-Kein richtig oder falsch

-erfassen die Ausprägung von Eigenschaften, Motiven, Interessen, Einstellungen, etc.

-Objektive und subjektive Persönlichkeitstests

**Projektive Tests**

Rorschach-Test: Tintenkleckse lassen Interpretation in verschiedene Objekte zu. Je nach Objekt soll das etwas über die Persönlichkeit aussagen.

Familie in Tieren: Sehr beliebter Test obwohl er stark unzulässige Gütekriterien hat. Angaben zu Reliabilität und Objektivität fehlen völlig.

**Testnormierung**

Wie gut lässt sich das Testergebnis mit den Ergebnissen anderer Menschen vergleichen?

Normierung = Bezugssystem, um individuelle Testwerte im Vergleich zu einer repräsentativen Stichprobe einordnen zu können

Normen sind wichtig

Normierung ist aber nicht unbedingt nötig je nach Forschung

**Testen: Verfälschungen und Gegenmassnahmen**

|  |  |  |  |
| --- | --- | --- | --- |
| Testart | Verfälschung | Gegenmassnahmen | Auch bei Befragung |
| Leistung | Raten | -Einsatz von falschen Antwortmöglichkeiten (Distraktoren)  -Ratekorrektur | Nein |
| Persönlichkeit | Psoitive Selbstdarstellung | -Aufforderung zu korrektem Testverhalten  -Randomized-Response-Technik | Ja |
| Persönlichkeit | Soziale Erwünschtheit | -Kontrollskalen | Ja |
| Persönlichkeit | Schematische Antworttendenzen | -Ausbalancierte Antwortvorgaben | Ja |

**Randomized Response Technik**

2 Fragen: Ein mit offensichtlicher Antwort, andere mit tatsächlicher zu untersuchender Frage. Sammeln der Antworten, danach Anteil der korrekten Antworten.

## **3.4 Identifizierung und Auswahl der Studienteilnehmenden**

**Grundgesamtheit / Population**: Menge aller potenziellen Untersuchungsobjekte für eine gegebene Fragestellung.

**Stichprobe**: Teilmenge einer Grundgesamtheit.

**Vollerhebung**: Untersuchung aller Objekte der Grundgesamtheit.

## 3.4.1 Repräsentativität von Stichproben

Die Repräsentativität gibt an, wie gut bzw. unverzerrt die Merkmalszusammensetzung in der Population wiederspiegelt.

**Merkmalsspezifisch-repräsentativ**: Eigenschaft trifft auf Personen eines bestimmten Merkmals zu.

**Global-repräsentativ**: Eigenschaft trifft auf alle Personen zu.

**Zielpopulation**: Gruppe aller Entitäten, die untersucht werden sollen.

**Zugängliche Population**: Gruppe der Entitäten, welche theoretisch für die Untersuchung zur Verfügung stehen könnte.

**Stichprobe**: Gruppe der Entitäten, welche theoretisch für die Untersuchung zur Verfügung steht

Die Stichprobengrösse alleine ist nicht ausreichend, um die Repräsentativität zu gewährleisten.

**Probleme: Abdeckungsfehler**

* **Über- / Unterabdeckung**: Gruppe mit gewissen Merkmalen ist zu stark / schwach in der Stichprobe vertreten.
* **Stichprobenausfälle / Non-Response**: wird über die Ausschöpfungsrate / Rücklaufquote beschrieben
* **Abdeckungsfehler**: Rückschluss kann nur über die Gesamtheit aller Objekte, aus der die Stichprobe tatsächlich stammt (Inferenzpopulation) gemacht werden.

**Muss eine Stichprobe immer repräsentativ sein?**

**Ja**, wenn eine Populationsbeschreibung gemacht wird.

**Nein**, wenn eine Hypothesenprüfung stattfindet, welche bloss einen kausalen Zusammenhang finden möchte.

### 3.4.2 Arten von Stichproben

1. **Zufallsstichproben (gut für Repräsentativität)**
   1. **einfache Zufallsstichprobe**

Stichprobe wird durch Zufall aus der Grundgesamtheit gezogen, wobei jedes Individuum die genau gleiche Auswahlwahrscheinlichkeit hat. Für diese Stichprobenart muss allerdings jedes Untersuchungsobjekt erfasst sein und ausgewählt werden können.

*Vorteil*: faire, unverzerrte Selektion

*Nachteil*: Repräsentativität ist nicht garantiert

* 1. **geschichtete Zufallsstichprobe**

Population wird auf Basis von gewissen Merkmalen in Untergruppen unterteilt. Von jeder Schicht wird per Zufall eine gleich grosse Auswahl an Personen gezogen.

*Vorteil*: Jede Schicht ist gleich gut repräsentiert; genauere Parameterschätzungen

*Nachteil*: keine Repräsentativität für die Zielpopulation

**proportional geschichtete Stichprobe**: Population wird wieder in Subgruppen unterteilt, aber es wird von jeder Subgruppe eine vorher festgelegte Anzahl Personen gezogen, welche proportional zur Grundgesamtheit sind.

*Vorteil*: Proportionen der Stichprobe repräsentiert die Grundgesamtheit

*Nachteil*: aufwendig, da Proportionen bekannt sein müssen; Problem mit sehr kleinen Schichten

* 1. **Klumpenstichprobe**

Es werden schon natürliche, schon bestehende Gruppen (Cluster / Klumpen) benutzt. Von allen in der Population bestehenden Clustern werden einige der Cluster gezogen und dann alle Entitäten in diesen Clustern untersucht.

*Vorteil*: einfache Methode um eine grosse, relativ repräsentative Stichprobe zu ziehen

*Nachteil*: alle in der Population enthaltenen Cluster müssen bekannt sein, Daten

1. **Stichproben ohne Zufallsauswahl**
   1. **Gelegenheitsstichprobe**

Nicht-zufälliges Auswahlverfahren = aktuell leicht verfügbare Teilnehmende.

Häufigste Methode der Stichprobenziehung in der psychologischen Forschung.

*Vorteil*: Einfache Methode zur Stichprobenziehung

*Nachteil*: Stichprobe höchstwahrscheinlich verzerrt, nicht repräsentativ für Zielpopulation

**Snowballl-sampling**: Man findet eine Person, die an der Studie teilnehmen möchte und bittet sie, doch auch noch Personen aus ihrem Umfeld ebenfalls für die Studie zu gewinnen.

* 1. **Quotenstichprobe**

Gelegenheitsstichprobe mit vorgegebenen Quoten hinsichtlich bestimmter Merkmale

*Vorteil*: Forschende können Zusammensetzung der Stichprobe kontrollieren

*Nachteil*: Stichprobe enthält wahrscheinlich immer noch Verzerrung

# 

# **4. Quantitative Forschungsmethoden**

## **4.1 Forschungsstrategie / Forschungsdesign**

Es wird unter folgenden Forschungsdesigns-Arten unterschieden:

* **Deskriptiv**: reine Beschreibung einzelner Merkmale
* **Korrelativ**: Zusammenhänge zwischen zwei Variablen, keine Erklärung
* **Nicht-experimentell**: Zusammenhänge zwischen zwei Variablen (i.d.R. Gruppenunterschiede), keine Erklärung
* **Quasi-experimentell**: Versuch einer Annäherung an Ursache-Wirkungs-Zusammenhänge (Versuch der Erklärung); Problem der natürlichen Gruppen und Konfundierung von Alternativerklärungen mit dem Design
* **Experimentell**: Ursache-Wirkungs-Zusammenhänge (Erklärung) zwischen Variablen

Es findet zudem eine Unterscheidung zwischen Quer- & Längsschnittdesigns statt.

### 4.1.1 Deskriptives Design

**Querschnittsstudie**

Ziel: Beschreibung der Ausprägung eines Merkmals anhand einer (oder mehrerer) möglichst repräsentativen Stichproben zu einem Messzeitpunkt.

→ Umfrageforschung

**Längsschnittstudie**

Ziel: Beschreibung der Ausprägung eines Merkmals anhand einer oder mehrerer verschiedener möglichst repräsentativer Stichproben zu verschiedenen Messzeitpunkten.

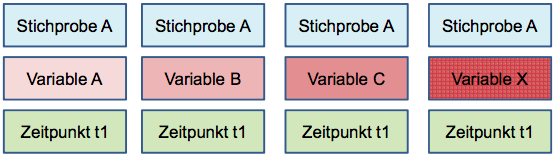
### 4.1.2 Korrelatives Design

Ziel: Art und Stärke der Zusammenhänge zwischen Variablen prüfen.

🡪Es sind keine Kausalaussagen möglich (dazu ist ein Experiment nötig)

🡪Es sind keine Aussagen über Gruppenunterschiede möglich (dazu ist ein nichtexperimentelles Forschungsdesign nötig)

Bei der **korrelativen Querschnittsmethode** werden Zusammenhänge zwischen verschiedenen Variablen zum gleichen Messzeitpunkt in der gleichen Stichprobe gesucht.



Vorteile

-Zusammenhänge zwischen Variablen sind prüfbar

-Es gibt keine Probleme mit Stichprobenausfall, Übungseffekten, ...

-schnell und einfach umsetzbar

-nützlicher Vergleich zwischen Gruppen

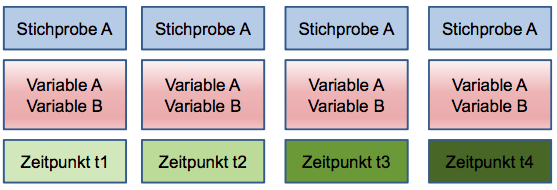
Nachteile

-Alle Vorteile der Längsschnittstudie sind nicht möglich

-Übertragbarkeit auf andere Erhebungszeitpunkte fraglich

-Stichproben möglicherweise nicht vergleichbar

Bei der **korrelativen Längsschnittmethode** werden Zusammenhänge zwischen Variablen zu verschiedenen Messzeitpunkten in der gleichen Stichprobe gesucht.



Vorteile

-Korrelationen innerhalb einer Variablen über die Zeit sind prüfbar

-Stabilität / Retest-Reliabilität kann überprüft werden

-Zusammenhänge zwischen zwei Variablen über die Zeit sind prüfbar

Nachteile

-Mögliches Auftreten von Testungseffekten

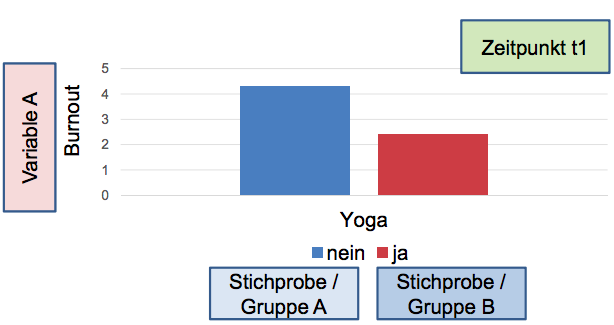
-Selektive Stichprobenausfälle möglich

-Hoher Zeit- und Personalaufwand

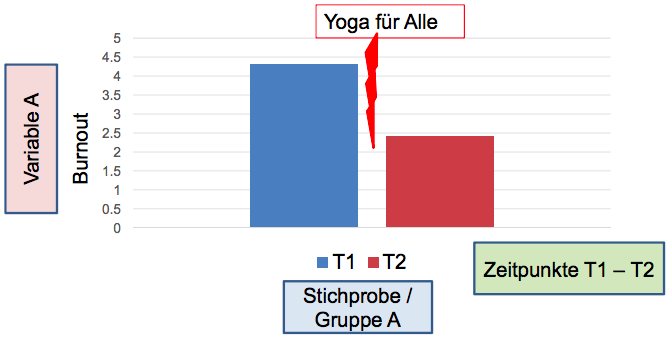
### 4.1.3 Nichtexperimentelles Forschungsdesign

Ziel: Veränderungen des Verhaltens durch Experimentalumgebung beseitigen durch Beobachtung in der natürlichen Umwelt. So sind allerdings viele Störvariablen vorhanden.

Bei der **nichtexperimentellen Querschnittsstudie** werden Mittelwertsunterschiede zwischen verschiedenen Stichproben / natürlichen Gruppen zum gleichen Messzeitpunkt verglichen.



Bei der **nichtexperimentellen Längsschnittstudie** werden Mittelwertsunterschiede innerhalb einer Stichprobe / natürlichen Gruppe über zwei Messzeitpunkte mit zwischengeschaltetem Treatment verglichen.



### 4.1.4 Zusammenhänge / Korrelationen und Kausalität

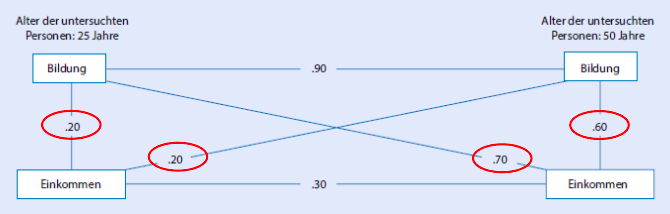
**Korrelationen sind nicht geeignet, die Gültigkeit eines Kausalmodells nachzuweisen. Allerdings ist es möglich, durch Nullkorrelationen Kausalmodelle zu falsifizieren, da Kausalrelationen Korrelationen implizieren.**

*Annäherung an Kausalitätsprüfung durch “cross-lagged-panel design”*

-Zeitversetzte Überprüfung konkurrierender Hypothesen zur Richtung des Zusammenhangs

-Annäherung an Kausalitätsprüfung

-**Kausalitätsprüfung nicht abschliessend möglich**



**Problem von korrelativen Forschungsdesigns**

*-Funktionen von korrel. Forsch.Designs*: Zusammenhänge zwischen Variablen erkennen & Kausalhypothesen ausschliessen

-Korrelative Forschungsdesigns haben geringe interne Validität, vor allem bei Querschnittdesigns.

-“Cross-lagged panel designs” haben die höchste interne Validität bei den korrelativen Forschungsdesigns.

-Drittvariablenproblem führt dazu, dass nie abschliessend kausale Aussagen möglich sind.

### 4.1.5 Experimente

**Definition:** *Unter einem Experiment versteht man die systematische Beobachtung einer abhängigen Variablen unter verschiedenen Bedingungen einer unabhängigen Variablen bei gleichzeitiger Kontrolle der Störvariablen, wobei die zufällige Zuordnung (Randomisierung) von Probanden und experimentellen Bedingungen gewährleistet sein muss. Das Experiment ist die einzige Möglichkeit, um Kausalhypothesen zu prüfen. Das Experiment ist aber aus ethischen, ökonomischen oder praktischen Gründen häufig nicht durchführbar.*

#### 4.1.5.1 Begriffe

**Unabhängige Variable (UV):** von den Forschenden manipulierte Variable

**Abhängige Variable (AV):** wird im Hinblick auf ihr Auftreten oder ihre Ausprägung beobachtet, um Effekte der UV zu messen.

**Störvariablen:** Einflussgrössen, die systematisch mit der UV variieren und auf die AV einwirken. → Konfundierung

**Konfundierung**: Drittvariable, die sowohl mit der UV als auch mit der AV in Zusammenhang steht und zu einer Verzerrung des Zusammenhangs zwischen UV und AV führt.

**Randomisierung**: zufällige Einteilung der Versuchspersonen in die experimentellen Gruppen. Durch Randomisierung werden bei ausreichend grosser Gruppengrösse personenbezogene Störvariablen neutralisiert. Beim Quasi-Experiment werden natürliche Gruppen zur Unterteilung verwendet.

**Wichtigster Unterschied zwischen Experimenten und Quasiexperimenten ist die Randomisierung.**

Unterschied Zufallsstichprobe und zufällige (randomisierte) Zuteilung zu Experimentalgruppen: Bei der Zufallsstichprobe wird bestimmt, wer überhaupt als VPN ausgewählt wird. Bei der Einteilung werden die ausgewählten VPN dann zufällig in die Untergruppen aufgeteilt.

#### 4.1.5.2 Kontrollgruppen

**Definition**: Kontrollgruppe heisst die Gruppe, die nicht in der Behandlungs- / Interventions- / experimentellen Manipulationsbedingung ist. Das Pendant dazu ist die Experimentalgruppe, die der Manipulationsbedingung ausgesetzt ist.

Generell wird im Experiment ein Vergleich der abhängigen Variable in verschiedenen Bedingungen gemacht.

2 Arten von Kontrollgruppen (Aktive / Inaktiv):

1. No-treatment Kontrollgruppen, die der Manipulation nicht ausgesetzt werden.
2. Placebo-Kontrollgruppen / aktive Kontrollgruppen

Aktive Kontrollgruppen:

Inaktive Kontrollgruppen:

#### 4.1.5.3 Versuchsplan

**Definition**: *logischer Aufbau einer empirischen Untersuchung im Hinblick auf Hypothesenprüfung*

*4 Entscheidungen*

1. vollständige oder unvollständige Pläne
2. Bestimmung der Anzahl der Beobachtungen pro experimenteller Bedingung
3. interindividuelle oder intraindividuelle Bedingungsvariation
4. randomisierte oder nichtrandomisierte Zuordnung der VPN zu den experimentellen Bedingungen.

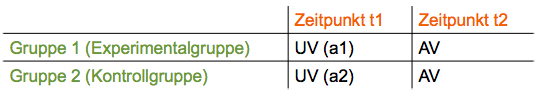
**Zweigruppenplan**

1 UV mit zwei Abstufungen

🡪2 Gruppen (pro Abstufung der UV eine)

🡪Zwei Zeitpunkte t1 und t2

Dies ist der einfachste experimentelle Plan. Meistens gibt es 2 Gruppen, die Experimentalgruppe und die Kontrollgruppe.



**Manipulation Check**

Der Manipulation Check ist eine Methode, um zu überprüfen, inwieweit die Teilnehmenden die experimentelle Manipulation wahrgenommen und interpretiert haben und/oder um die direkte Wirkung der experimentellen Manipulation zu erfassen.

Zwei Arten des MC:

1. Direkte Messung der unabhängigen Variablen
2. Im Nachbefragungs-Fragebogen enthaltene Fragen zum Experiment

MCs sollten angewendet werden, wenn…

1. die Manipulation an den Teilnehmenden angesetzt hat
2. die Manipulation subtil war
3. eine Simulation Teil des Experiments war
4. es eine Placebo-Kontrollgruppe gab

**Zweigruppenplan Prä-Post**

Variante: Messung der AV vor dem Experiment (Prätest)

*Vorteile der Vorhermessung:*

-Prüfung von möglichen Unterschieden im Ausgangsniveau

-Veränderung feststellbar

*Nachteile der Vorhermessung:*

-Kosten

-Nicht immer möglich

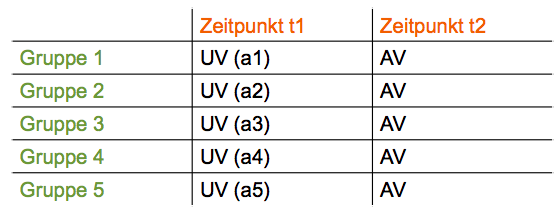
-Auswirkungen der Vorhermessung auf die Nachhermessung



**Experiment: Mehrgruppenplan / Einfaktorieller Versuchsplan**

UV mit mehr als 2 Stufen

i.d.R. Prüfung von Unterschiedshypothesen

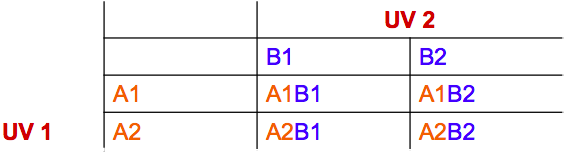


**Mehrfaktorielle Versuchspläne**

Mehrere UVs mit x Stufen

UV = Faktor

Bei zwei UVs: zweifaktorieller Plan, bei drei UVs: dreifaktorieller Plan, ...



Problem: Die Anzahl UVs wird begrenzt durch:

-die notwendige Anzahl Teilnehmende

-statistische Schwierigkeit

-Interaktionen / Wechselwirkungen zwischen UVs schwierig mit zu vielen Variablen

**Mehrfaktorielle Designs: Haupteffekte und Interaktionen**

Hypothesen über die UVs: **Haupteffekthypothesen**

→ Haupteffekt = Wirkung einer UV auf die AV

Hypothesen über die Wechselwirkung zwischen UVs: **Interaktionshypothesen**

→ Interaktion / Wechselwirkung = Wirkung einer UV1 auf AV ist abhängig von der anderen UV2

→ UV2 wird Moderator genannt

→ Ein Moderator verändert die Stärke und/ oder die Richtung der Wirkung einer UV auf die AV

**→ Interaktionen sollten zur besseren Veranschaulichung immer grafisch dargestellt oder alle Mittelwerte berichtet werden!**

**→ Interaktionen sind nicht in einfaktoriellen Designs prüfbar**

Es sind auch Interaktionen zwischen mehr als 2 Variablen möglich.

→ Interaktion zwischen drei Variablen = Dreifachinteraktion, Interaktion 2. Ordnung

→ Ab mehr als 3 UVs → schwierig

**Techniken zur Kontrolle potenzieller Störvariablen:**

-Kontrollgruppe

-Randomisieren

-Blindversuche

-Konstanthalten

-Parallelisieren / Matching

-systematische Variation / statistische Kontrolle

-zufällige Variation

-Elimination

-Registrieren

**Within/Between-Subjects Designs**

In einem within-subjects design werden zwei oder mehrere experimentelle Treatments innerhalb der gleichen Individuen verglichen.

In einem between-subjects design werden zwei oder mehrere experimentelle Treatments, die mit unterschiedlichen Personen durchgeführt wurden, verglichen.

**Vor- und Nachteile von within- und between-subject designs**

|  |  |  |
| --- | --- | --- |
|  | Within-Subjects | Between-Subjects |
| Vorteile | -kein Problem mit personenspezifischer Konfundierung von Störvariablen  -Weniger Teilnehmende notwendig  -Effekte der UV werden eher entdeckt als in between-subjects designs (grössere “Power”) | -Teilnahme an einer experimentellen Bedingung hat keinen Einfluss auf die Teilnahme an einer anderen experimentellen Bedingung  -Innerhalb einer Bedingung können mehr Daten gesammelt werden, als wenn jemand an mehreren Bedingungen teilnimmt. |
| Nachteile | -Dropout durch mehrere Messzeitpunkte  -Gefährdung der internen Validität durch zeitliche Veränderung  -Positionseffekte  -Übertragungseffekte | -Immer potentielle Konfundierung von Personvariablen mit experimenteller Bedingung |

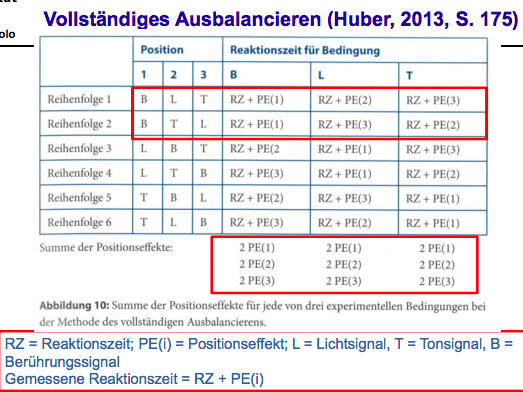
**Spezielle Störeffekte bei wiederholter Messung**

*Messwiederholung*: mehrere Messungen pro Person zu unterschiedlichen Zeitpunkten. Diese Definition ist bezogen auf within-subjects Experimente: Personen nehmen mehrfach innerhalb des gleichen Experiments in verschiedenen Bedingungen teil.

**Zeitliche Veränderung** ausserhalb der Untersuchungssituation

**Positionseffekt:** Mögliche Auswirkungen der Anordnung oder Reihung der verschiedenen Messungen/Items/… des Experiments. Position einer experimentellen Bedingung in einer Folge von Bedingungen hat einen Effekt auf die Resultate. Positionseffekte können mit Ausbalancieren kontrolliert werden.

**Vollständiges Ausbalancieren**: Jede mögliche Kombination an Reihenfolge wird durchgeführt.



Anzahl an Möglichkeiten=n!

Vorteil: Jede Kombination abgedeckt

Nachteile: Sehr aufwendig, viele VPN benötigt. Praktisch kaum umsetzbar

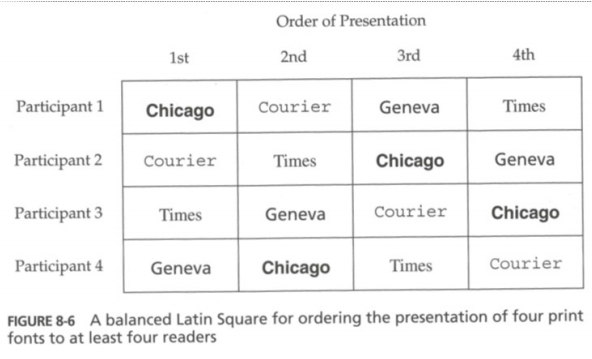
**Unvollständiges Ausbalancieren**: Nur Teile der Möglichkeiten werden durchgeführt. Als unvollständige Methoden gelten Zufallsauswahl, Spiegelbildmethode und das lateinische Quadrat.

**Zufallsauswahl**: Es gibt eine randomisierte Auswahl aus allen möglichen Reihenfolgen. Jede VPN erhält eine andere Reihenfolge. Dies ist bei grossen Stichproben möglich.

**Spiegelbildmethode**: Es wird eine Reihenfolge gewählt, die dann gespiegelt wird (Bsp: AB-BA, ABC-CBA). Alle VPN bekommen die gleiche Sequenz. Mit dieser Methode kann für jede Person eine Kontrolle des Positionseffektes durchgeführt werden.

Für diese Methode muss aber jede experimentelle Bedingung zweimal pro Person durchführbar sein und der Positionseffekt muss linear und vorher bekannt sein.

**Lateinisches Quadrat**: Gleich viele Reihenfolgen wie es experimentelle Bedingungen hat. Die VPN werden auf die Reihenfolgen verteilt. Jede Reihenfolge hat eine Bedingung zu einem unterschiedlichen Zeitpunkt.



Vorteil: Das lateinische Quadrat kann in einen faktoriellen Versuchsplan integriert werden:

-Die Position der einzelnen experimentellen Bedingung wird als UV eingeführt

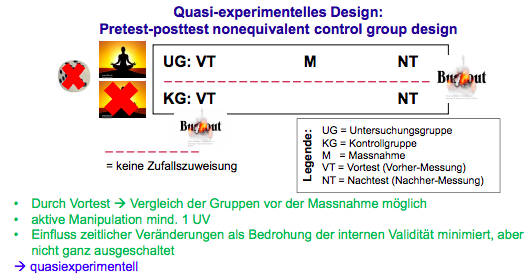
-VPN werden in Gruppen zugeteilt

→ Überprüfung (nicht nur Kontrolle) möglich

**Übertragungseffekte**: Effekte einer Messung auf eine andere Messung. Es ist eine Störvariable, die davon herrührt, dass eine frühere experimentelle Bedingung eine spätere inhaltlich beeinflusst. Übertragungseffekte können durch Beseitigung der Ursache, Between-subjects designs und “Wash-out periods” beseitigt werden.

**Wash-out periods**: Zeit zwischen zwei Bedingungen um Übertragungseffekte zu beseitigen.

### 4.1.5 Quasiexperimente



Achtung: Beim quasi-experimentellen Vorgehen findet keine randomisierte Zuteilung zu Gruppen statt.

### 4.1.6 Vergleich der Forschungsdesigns

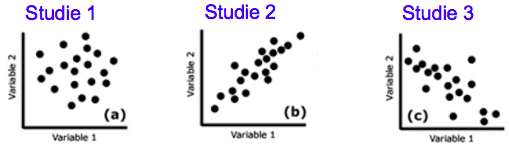
|  |  |  |
| --- | --- | --- |
| **Was** | **Vorteile/ Was wird gemessen?** | **Nachteile/ Was wird nicht gemessen?** |
| Deskriptiv | -reine Beschreibung einzelner Merkmale (z.B. Häufigkeit) | -keine Zusammenhänge  -keine Gruppenunterschiede  -keine experimentelle Manipulation  -keine Kausalitätsprüfung |
| Korrelativ | -Zusammenhang / Zusammenhänge zwischen Variablen | -keine Gruppenunterschiede  -keine experimentelle Manipulation  -keine Kausalitätsprüfung |
| Nicht- experimentell | -Unterschiede natürlicher Gruppen auf AV  -im within-subjects-Kontext: Unterschiede einer Gruppe über die Zeit | -keine Randomisierung (nur natürliche Gruppen)  -keinerlei Kontrolle von Störvariablen |
| Quasi- experimentell | Versuch einer Annäherung an Kausalitätsprüfung durch experimentelle Manipulation von min. 1 UV und Versuch der Kontrolle von Störvariablen → mehr als reine Beschreibung von Mittelwertsunterschieden | -keine Randomisierung (sondern natürliche Gruppen)  -keine vollständige Kontrolle von Störvariablen  -Kausalaussagen nie abschliessend möglich |
| Experimentell | -Kausalitätsprüfung zwischen UVs und AVs durch randomisierte Zuteilung zu den Bedingungen (between oder within) und aktive Manipulation von min. 1 UV | -Auch hier gewisse Gefährdungen der internen Validität durch entsprechende Störvariablen möglich |

### 4.1.7 Meta-Analyse

**Definition**: Zusammenfassung verschiedener Studien mithilfe statistischer Analysen. Bei der Metaanalyse werden die Daten aus verschiedenen Studien zusammengefasst und erneut statistisch ausgewertet. So können Durchschnittswerte über die verschiedenen Studien gemacht werden. Metaanalysen gewähren einen Überblick über das “grosse Ganze”.

**Ausgangslage**: Eine einzelne Studie kann Hinweise für die Beantwortung der Forschungsfrage geben, aber nichts abschliessend “beweisen”. → Replikationen sind notwendig

**Problem**: Ergebnisse können uneinheitlich sein.



**Ziel einer Metaanalyse**: Überblick über aktuellen Forschungsstand verschaffen

**Alternativen zur Metaanalyse**

-**Narratives Review**: Subjektive Zusammenfassung von Studien zur gleichen Forschungsfrage

-**Systematisches Review**: Systematische Zusammenfassung der Studienergebnisse nach umfassender Literaturrecherche; häufig Kodierung der Studien nach Qualität; häufig in der Gesundheitsforschung anzufinden. Das systematische Review ist aber keine statistische Zusammenfassung der Studien.

*Bemerkung*: Jede Metaanalyse ist auch ein systematic review!

#### 4.1.7.1 Vorgehen

1. **Bestimmung der empirisch-inhaltlichen Hypothese**

Themensuche und kreieren einer Hypothese, welche dann mit einer Metaanalyse überprüft werden soll.

1. **Umfassende Literatursuche**

Gefahr des “publication bias” / “file-drawer-problem”, wenn nur signifikante Ergebnisse veröffentlicht werden.

Publication bias: Irrelevante werden weniger oft publiziert wie relevante.

File-drawer-problem: Positive Resultate werden öfter publiziert wie negative oder unschlüssige Resultate.

Eine Metaanalyse bietet die Möglichkeit der statistischen Korrektur bzw. Berechnung

1. **Bewertung und Codierung der Studien**

Bewertung nach interner Validität und Validität der Erfassung der Variablen.

3 Probleme:

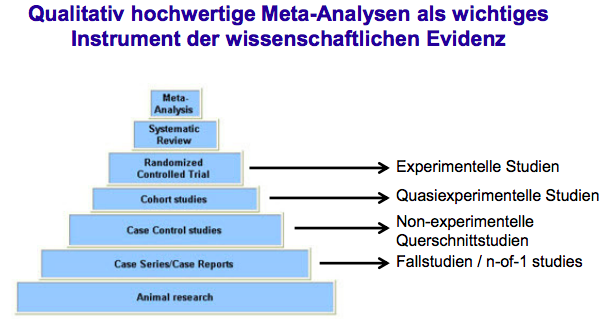
* *Müll-rein-Müll-raus-Problem*: Es müssen Mindeststandards für Primärstudien vorhanden sein. Die Studien sollten nach Qualität gewichtet werden und Studien, die den Mindeststandard nicht einhalten ausgeschlossen werden.
* *Äpfel-und-Birnen-Problem*: Die Studien müssen vergleichbar sein
* *Abhängigkeitsproblem*: Teilergebnisse einer Studie, die mehrfach in die Metaanalyse einfliessen führen zu Verzerrung. Man sollte sich also auf eine Effektgrösse pro Studie beschränken.

Übereinstimmungen in der Codierung müssen erwähnt werden.

1. **Statistische Analyse**
   1. Bestimmung der Gesamteffektgrösse
   2. Bestimmung der Homogenität der Varianzen
2. **Interpretation**

#### 4.1.7.2 Fazit

Metaanalyse kann für experimentelle, quasiexperimentelle, nichtexperimentelle, korrelative Primärstudien durchgeführt werden. Sie ist ein wichtiges Instrument zur “objektiven”, statistischen Erfassung des Forschungsgegenstands auf einem Gebiet. Die Metaanalyse muss hohen Qualitätsstandards genügen, sonst können falsche Schlussfolgerungen vorkommen.



## 4.2 Interne und Externe Validität

**Validität**: Gültigkeit einer Untersuchung, Frage ob die Ergebnisse aussagekräftig sind.

### 4.2.1 Interne Validität

Eine Untersuchung ist intern valide, wenn die Ergebnisse eindeutig kausal interpretierbar sind, die beobachteten Effekte bei der abhängigen Variable also zweifelsfrei auf die Veränderung der unabhängigen Variable zurückzuführen sind. Die interne Validität sinkt mit der Anzahl plausibler Alternativerklärungen für den beobachteten Effekt.

### 4.2.2 Externe Validität

Eine Untersuchung ist extern valide, wenn eine Generalisierung der Ergebnisse zulässig ist. Dazu müssen die Ergebnisse über die Untersuchungssituation und die untersuchten Personen hinaus generalisierbar sein. Die externe Validität sinkt mit der Zunahme der Unnatürlichkeit der Untersuchungssituation.

### 4.2.3 Feld- & Laboruntersuchungen

**Labor**

-jede Umgebung, die offensichtlich für eine wissenschaftliche Untersuchung eingerichtet wurde

-wird von den Teilnehmenden als künstlich wahrgenommen

-Kontrolle bzw. Ausschaltung untersuchungsbedingter Störvariablen

→ gut für interne Validität

-Ergebnisse nicht unbedingt übertragbar auf Alltag

→ Schlecht für externe Validität

**Feld**

-Umgebung wird von den Teilnehmenden als natürlich wahrgenommen

-kaum durch Versuchsleitende verändert

-Gute Übertragbarkeit auf den Alltag

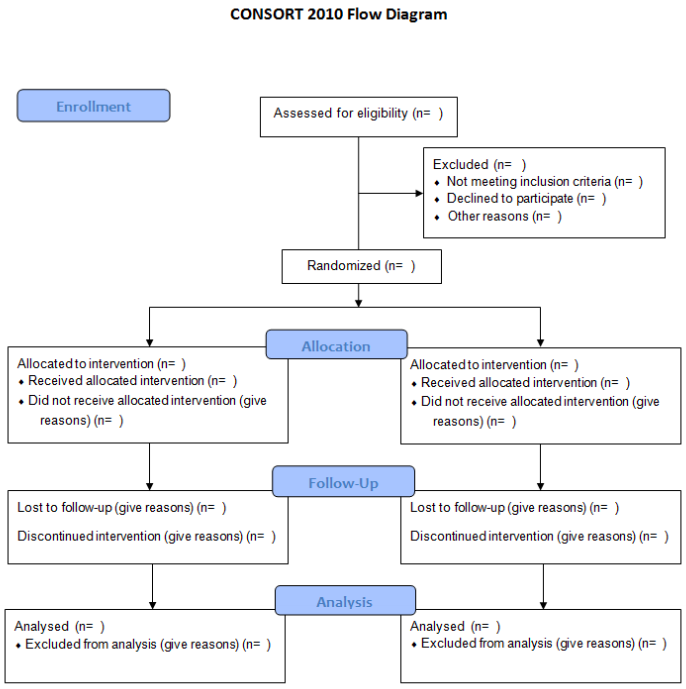
→ gut für externe Validität

-Verminderte Kontrolle von Störvariablen

→ schlecht für die interne Validität

### 4.2.4 Gefährdungen der internen Validität

|  |  |
| --- | --- |
| **Gefährdung** | **Umgang / Lösungsmöglichkeiten** |
| Unklare zeitliche Abfolge (unklar ob A oder B zuerst war) | Keine Kausalschlüsse |
| Selektionseffekte (In der Stichprobe mit 2 verschiedenen Gruppen anhand von Merkmalen befinden sich Gründe, warum Person A in Gruppe 1 und Person B in Gruppe 2 ist) | Randomisierung, Parallelisierung, Konstanthalten |
| Externe zeitliche Einflüsse (Veränderungen entstanden mit der Zeit aufgrund von anderen Ursachen und nicht aufgrund der UV) | Kontrollgruppe |
| Reifungsprozesse | Kontrollgruppe |
| Statistische Regressionseffekte | Vermeidung von Extremgruppen |
| Experimentelle Mortalität | Genaue Dokumentation, statistischer Umgang mit fehlenden Werte |
| Testübung (mehrfache Befragung mit gleichen Fragen) | Verschiedene Testversionen |
| Mangelnde instrumentelle Reliabilität | Verwendung standardisierter Messinstrumente mit hoher Reliabilität |
| Kombination der genannten Bedrohungen | Kombination von Umgang mit einzelnen Gefährdungen |



*Experimentelle Mortalität*

Woran kann ein Experiment scheitern?

-Enrollment: VPN erfüllen die Bedingungen nicht, die für die Teilnahme nötig sind.

-Allocation: Wer konnte keiner Gruppe zugewiesen werden?

-Follow-up: Follow-up zu späterem Zeitpunkt: Wer möchte dann nicht mehr mitmachen?

-Analysis: Gewisse Resultate invalide, können nicht gebraucht werden.

## 4.3 Statistische Regressionseffekte – Regression zur Mitte

Extremwerte haben die Tendenz, sich bei wiederholter Messung zur Mitte der Verteilung hin zu verändern. Dies ist immer vorhanden bei verbundenen Messungen (Messung eines Merkmals, dass nicht durch einen einzigen Messwert bestimmt werden kann).

Bei Veränderungsmessungen ist die Regression zur Mitte stärker ausgeprägt, wenn die Retestreliabilität (Stabilität) des Tests kleiner ist.

