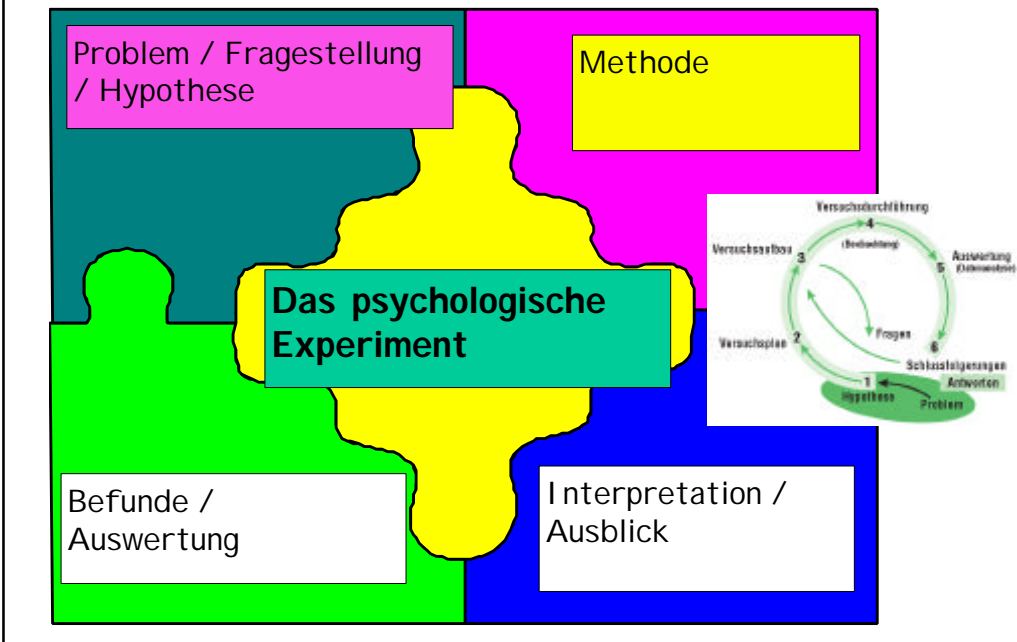


Das Experiment



Das Experiment

• E = {Erkundungs-, Entscheidungs-, Labor-, Feldexperiment}

1. Problem / Hypothesen

- Motive & Hintergrundwissen zur Fragestellung über Sachverhalte der Natur (Problem)
- Mögliche Antwort(en) auf Frage; Wette mit der Natur
- Kriterien & Klassifikation
- Modellvorhersagen

2. Methode

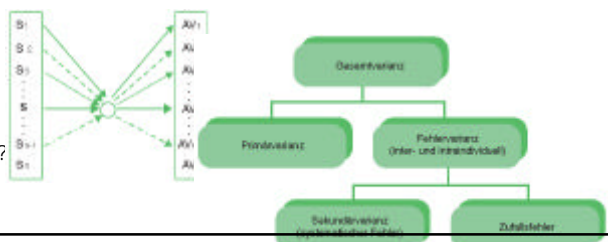
- Operationalisierung der Hypothesen (Konstrukte)
- Variablen, UV, AV, KV; Was messen & variieren? Wie? (MaxKonMin Prinzip), Validitätsfragen
- Versuchsplanung

3. Befunde / Auswertung

- Einzelfälle, Mittelwerte
- Deskriptiv / inferentiell

4. Diskussion / Ausblick

- Hypothese bestätigt?
- Modellvergleich - modifikation?



Das Experiment 2: Problem & Hypothesen

1. Problem

- Frage über Zusammenhänge zwischen Sachverhalten / Phänomenen
- EXKURS: Ethische Richtlinien? (APA, DGPs)

2. Hypothese

- Newtons „causa ficta“: Annahme, die geeignet ist, Phänomene zu erklären, aber noch nicht als die EINZIG MÖGLICHE und GÜLTIGE Erklärung des Phänomens erwiesen ist
- Deduktive Herleitung: wichtig für Theorieprüfung, aber nicht kreativ!
- Induktive Herleitung: aus „interessanter“ Einzelbeobachtung heraus
- Abduktive Herleitung: neue Fakten, die mit keiner existierenden Theorie erklärt werden können, führen zu neuer Hypothese, die dann selbst wieder Basis für Deduktionen wird
 - kein Schlußfolgern, sondern Problemlösen
 - kreativer Prozeß (schafft neues Wissen)

Das Experiment 3: Problem & Hypothesen 2

3. Kriterien einer erfahrungswissenschaftlichen Hypothese

- Hinausgehen über bereits festgestellte Sachverhalte, z.B. in Form einer empirischen GENERALISATION (logische Konstruktion aus Daten)
- Keine wilden Spekulationen / willkürliche Behauptungen / Tautologien (möglichst aus theoretischem Hintergrundwissen begründet)
- Empirische Überprüfbarkeit: konfrontierbar mit relevanten Daten

4. Klassifikation erfahrungswissenschaftlicher Hypothesen

- Singuläre Aussagen: Person P ist hochgradig leistungsmotiviert
- Existentielle Aussagen: Es gibt eine Psychologin mit IQ = 300 (nicht falsifizierbar)
- Universelle Aussagen: Die Kapazität des Kurzzeitgedächtnisses liegt unter 8 Einheiten (nicht verifizierbar; s. Popper, 1934)
- Quasi-universelle Aussagen: lassen bei Allhypothese Ausnahmen zu!
- Statistische Hypothesen (Wahrscheinlichkeitsverteilungen von Ereignissen; Intelligenz und Leistungsmotivation sind korreliert)
 - Nullhypothese vs. Alternativhypothesen: $m_2 - m_1 = 0$

Sarris/Reiss: Kurzer Leitfaden der Experimentalpsychologie

Tabelle 3.1: Typische Fehler bei der Generierung wissenschaftlicher Hypothesen (1).

- Es wird eine Hypothese formuliert, die zu allgemein gehalten ist, als dass sie in sinnvoller Weise untersucht werden könnte.
- Es werden die Begriffe in der Hypothese nicht operational definiert, so dass deren experimentelle Realisation als abhängige und unabhängige Variablen nicht möglich ist.
- Es wird versäumt, rechtzeitig an das Designing des Experiments, an dessen Durchführung sowie an die Auswertung der experimentellen Daten zu denken.
- Es wird versäumt, die einschlägige Fachliteratur zu studieren.

Sarris/Reiss: Kurzer Leitfaden der Experimentalpsychologie

Tabelle 3.1: Typische Fehler bei der Generierung wissenschaftlicher Hypothesen (2).

- Es wird zu sehr auf die Sekundärliteratur geachtet.
- Es wird beim Studium der Fachliteratur zu sehr auf die empirischen Befunde und zu wenig auf die dabei benutzten Methoden geachtet.
- Bei der Lektüre der Fachliteratur wird das wissenschaftliche Thema nicht genügend eingegrenzt (der Leser „verliert sich“ in der Literatur und wird dadurch entmutigt).
- Bei der Durchsicht der Fachliteratur werden zu viele Exzerpte angefertigt und dabei nicht klar genug zwischen Wichtigem und Unwichtigem unterschieden.

Hauptmerkmale guter Hypothesenbildung

- Testbarkeit und Falsifizierbarkeit
- Präzisierung
- Theorienrelevanz
- Morgan's Canon

(Dieses Gütekriterium besagt, dass bei sonst gleichem Voraussagewert einfachere Grundlagen solchen vorzuziehen sind, die eine kompliziertere und damit oft auch wohl spekulativere Basis der Hypothesenbildung zur Voraussetzung haben.)

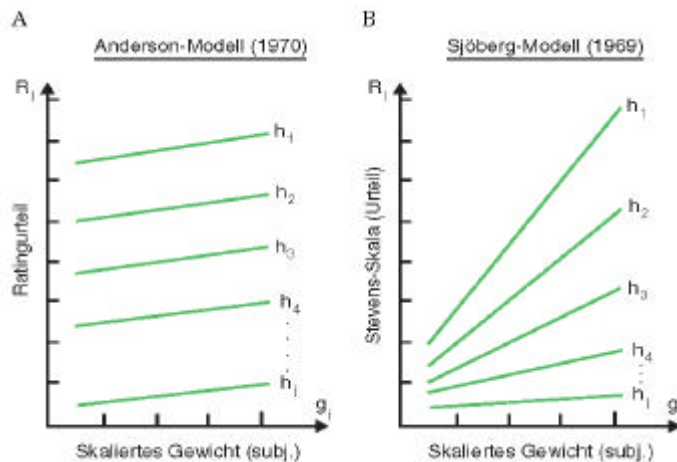


Abbildung 3.6: Schematische Repräsentation zweier mathematischer Modelle, die jeweils unterschiedliche Vorhersagen von Wahrnehmungstäuschungstrends für die Größengewichtstäuschung machen, wobei g_j das subjektive Gewicht und h_j die subjektive Höhe für verschieden schwere und große physikalische Objekte bedeuten. A: Andersons(1970) Paralleltrend-Vorhersage; B: Sjöbergs (1969) Wechselwirkungstrend-Vorhersage (nach Sarris 1995).

Das Experiment 4: Methode

Konstruktoperationalisierung

- **N**ominaldefinition (eines theoretischen Konstrukts / nicht beobachtbarer Sachverhalt):
genus proximum + differentia specifica (Auge = lichtempfindliches Sinnesorgan)
- **R**ealdefinition: Begriff wird durch reale Gegebenheiten, die er umfaßt, definiert
(Emotion = Freude, Wut, Trauer....)
- **O**perationalisierung (keine Definition im engeren Sinne): Bedeutung des Konstrukts wird durch Test- bzw. Meßoperation bestimmt

Beispiele

- N: A = Handlung, die Verletzung oder Kränkung eines Organismus intendiert bzw. bewirkt
(Dollard et al., 1930)
- R: A ist charakterisiert durch Veränderungen im Hormonhaushalt (Testosteron+,
Noradrenalin+) sowie Aktivierung von Hypothalamus und limbischem System
(insbesondere Mandelkern; Hofstätter, 1972)
- O: UV = Frustration (?), AV = Anzahl der Schläge auf Puppe

Das Experiment 5: Methode 2

Versuchsplanung / exp. Design

- **S<G> x T**; 2 UV Möglichkeiten: G = Group; T = Treatment
- S<G> (1 faktorieller, unabhängiger Gruppenplan)
 - $S = s_1 - s_n$
 - $G = g_1 - g_n$
 - Bsp.: **S₁₀<G₂>** = 10 VPn bekommen Alkohol (g_1), 10 bekommen Placebo (g_0), [AV = RZ(ms), F(%)]
- S x T (abhängige Messungen / Meßwiederholungsdesign)
 - $T = t_1 - t_n$
 - Bsp.: **S₂₀ x T₂** = (alle) 20 VPn bekommen zuerst geringe Menge Alkohol (t_1) dann große (t_2), [AV = RZ(ms), F(%)]
- S<G> x T (2 faktorieller gemischter Plan)
 - Bsp.: **S₁₀<G₂> x T₂** = 10 VPn ohne (g_0) & 10 mit Alkohol (g_1) hören sehen jeweils schwache (t_1) und starke (t_2) Lichtreize
 - $N = 10 \times 2 \times 2 = 40$ Meßwerte / Datenpunkte
- S<G<H>> x T x U x V

Sarris/Reiss: Kurzer Leitfaden der Experimentalpsychologie



Gütemerkmale eines Versuchsplans

Gütemerkmal 1: Kausaltheoretische Hypothese – vor Versuchsbeginn – ist vorhanden.

Gütemerkmal II: Experimentelle Variable ist manipulierbar

Gütemerkmal III: Alle übrigen Versuchsbedingungen sind kontrollierbar

	1 Strenges Experiment	2 Quasi- Experiment	3 Ex post facto Untersuchung	Korre- lative Untersuchung	4 Vorexperi- mentelle Untersuchung
I Kausaltheoretische Hypothese vor Versuchs- beginn vorhanden und hinreichend begründet	+	+	(+)	-	-
II Experimentelle Variablen manipulierbar bzw. manipuliert	+	+	-	-	-
III Alle übrigen Versuchs- bedingungen kontrollierbar bzw. kontrolliert	+	-	-	-	-

Abbildung 4.2: Allgemeines Gliederungsschema für eine Systematik der Versuchspläne nach den vier allgemeinen Designtypen und deren Bewertung nach verschiedenen Güteigenschaften (I, II, III). (Modifiziert nach Boesch & Eckensberger, 1969)

Weitere experimentelle Designs

Quasi-Experimente

Der quasi-experimentelle Versuchsansatz erreicht zwar nicht die kausaltheoretische Bedeutung des strengen Experiments

allerdings stellen die quasi-experimentellen Designs eine wichtige Ergänzung des klassischen, für sich allein zu einseitigen strengen Experiments dar.

Ex post facto- und korrelative Versuchsanordnungen

Bei der *ex post facto*-Versuchsanordnung wird nur Y systematisch beobachtet

Dann folgt rückblickend eine Suche nach X

Diese Anordnungen wie auch die *korrelativen Designs* besitzen gegenüber den *Quasi*-Experimenten - wenn überhaupt - nur einen sehr geringen bzw. gar keinen kausaltheoretischen Wert

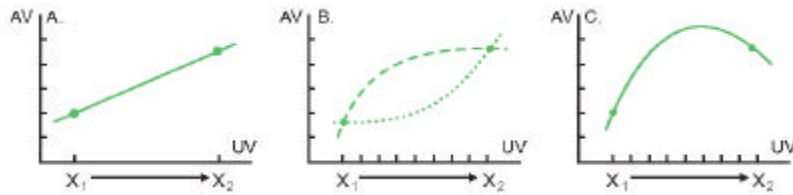


Abbildung 4.5: Zur Wahl eines trendanalytischen Mehrstichprobendesigns: Zwei gemäß einem Zweistichprobenversuchsplan – unter den Bedingungen X_1 und X_2 – erhaltene Durchschnittswerte werden in der grafischen Darstellung üblicherweise durch eine Gerade miteinander verbunden; dabei wird die Annahme einer linearen bzw. monotonen Beziehung zwischen der UV und AV, meist stillschweigend, gemacht (A.). Demgegenüber kann die tatsächliche Beziehung z.B. eher positiv oder eher negativ beschleunigt sein (B.) oder gar non-monoton, z.B. biton (C.) sein. – Die gesamte UV-AV-Beziehung lässt sich grundsätzlich durch die Verwendung eines Mehrstichprobenversuchsplans geeignet untersuchen, im Sinne einer trendanalytischen Designverwendung. (Sarris, 1992)

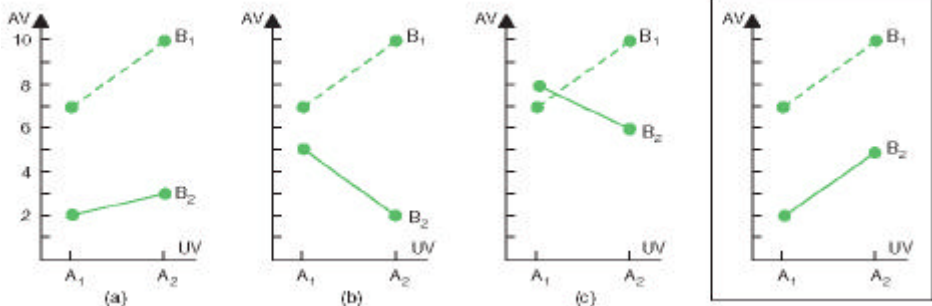


Abbildung 4.8: Typische Beispiele für Wechselwirkungen (Interaktionen) von zwei Faktoren, vereinfacht dargestellt mit jeweils nur zwei Merkmalsabstufungen. a: Ordinale; b und c: disordinale Wechselwirkung der Faktoren A und B im Vergleich zur Noninteraktion (keine Wechselwirkung: s. Paralleltrends in der Teilabb. rechts). (Nach Bredenkamp, 1980)

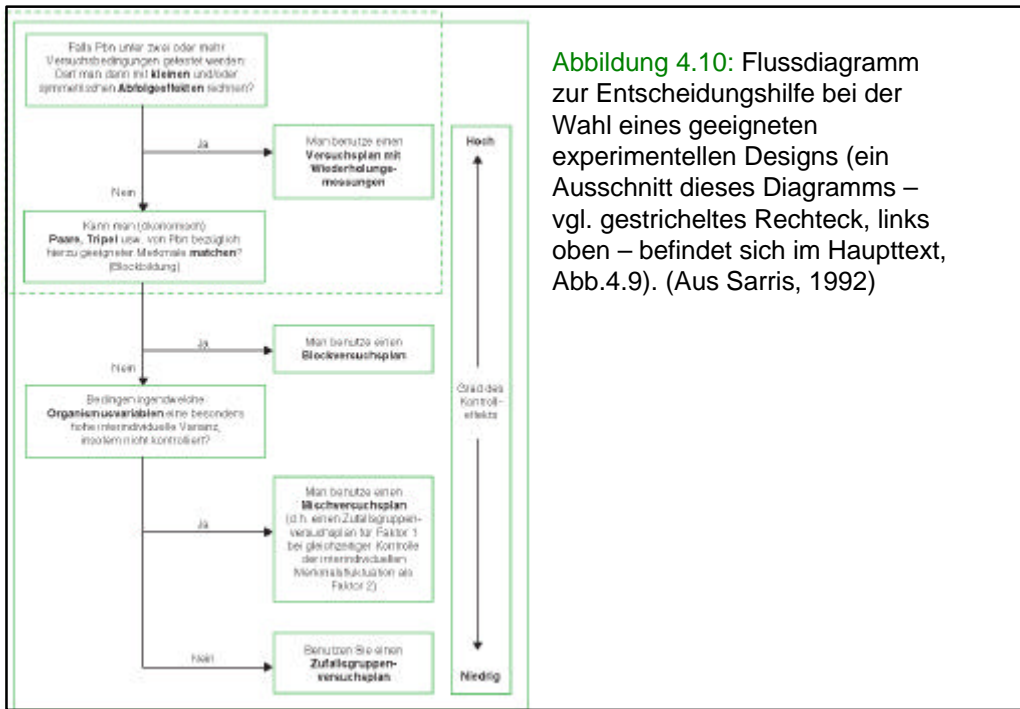
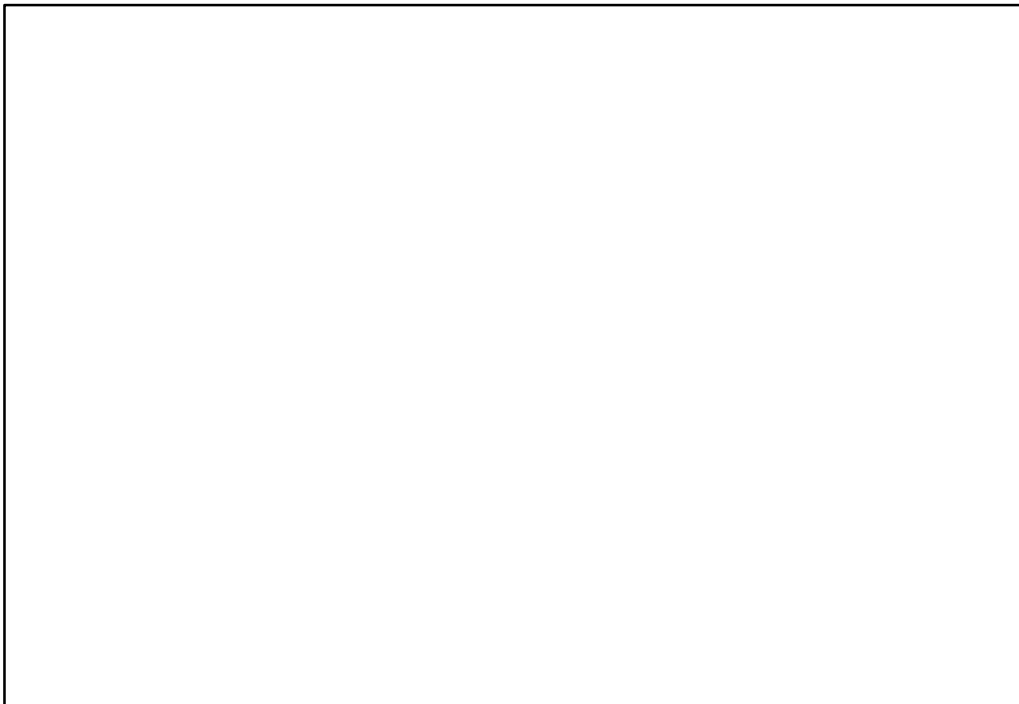


Abbildung 4.10: Flussdiagramm zur Entscheidungshilfe bei der Wahl eines geeigneten experimentellen Designs (ein Ausschnitt dieses Diagramms – vgl. gestricheltes Rechteck, links oben – befindet sich im Haupttext, Abb.4.9). (Aus Sarris, 1992)



Box 4.4: Validitätsbedrohung
(quasi-experimentelles Design)

Validitätsbeeinträchtiger Störfaktor	Kurze Charakterisierung des Artefakts
1. Zeitgeschehen (history)	Beobachtete Effekte gehen nicht allein auf experimentelle Bedingungskonstellation, sondern (zusätzlich) auf unkontrollierte <i>zwischenzeitliche</i> Ereignisse zurück.
2. Reifung (maturation)	Wenn sich der zu untersuchende Sachverhalt bezüglich biologischer (oder/und psychosozialer) Reifungsmerkmale verändert, ist mit reifungsbedingten Effekten zu rechnen, die den eigentlichen experimentellen Befund überlagern.

Box 4.4: Validitätsbedrohung
(quasi-experimentelles Designing)

Validitätsbeeinträchtiger Störfaktor	Kurze Charakterisierung des Artefakts
3. Mehrfache Testung (test sophistication)	Besonders bei mehrfacher Erhebung derselben Messdaten an ein und demselben Individuum können die während des zweiten (dritten,...) Messzeitpunkts erhobenen Daten aufgrund vorangegangener Testung beeinflusst sein (z. B. sensibilisierende Erfahrung im Umgang mit einem Test) .
4. Instrumentierung (instrumentation)	Die gemessenen Werte gehen z. T. auf die (zwischenzeitlich erfolgte) Veränderung der Messinstrumente zurück (z.B. aufgrund mangelnder Objektivität und Reliabilität eines Tests).

Box 4.4: Validitätsbedrohung
(quasi-experimentelles Designing)

Validitätsbeeinträchtigung
Störfaktor

- | | |
|---|--|
| 5. Statistische Regression (regression) | Werden mehr oder weniger extrem verschiedene Leistungsgruppen z.B. mithilfe eines Vortests gebildet, dann kann die mangelnde Reliabilität (Testzuverlässigkeit) des Messinstruments zu einer statistischen „Regression“ zur Mitte bei der zweiten (experimentellen) Testung führen . |
| 6. Auswahlverzerrung (selection) | Bei nicht-zufälliger Bildung von Versuchsgruppen können die damit von Anfang an bestehenden systematischen Ausgangsdifferenzen zwischen den Gruppen den eigentlichen experimentellen Effekt überlagern. |

Box 4.4: Validitätsbedrohung
(quasi-experimentelles Designing)

Validitätsbeeinträchtigung
Störfaktor

- | | |
|---|---|
| 7. Ausfalleffekte (experimental mortality) | Fallen im Untersuchungsverlauf Probanden von verschiedenen Versuchsgruppen aus, so kann das die eigentlichen experimentellen Effekte beeinflussen, insofern die Ausfallquote für die Gruppen systematisch verschieden ist |
| 8. Versuchsleiter-effekte (experimenter-bias effects, sog. Rosenthal-Effekte) | Bleiben die Eigenschaften, Verhaltensweisen oder/und Versuchserwartungen des Untersuchers unkontrolliert, kann das eine systematische Beeinträchtigung der eigentlichen experimentellen Befunde nach sich ziehen. |

Box 4.4: Validitätsbedrohung (quasi-experimentelles Designing)

Validitätsbeeinträchtiger Störfaktor

Kurze Charakterisierung des Artefakts

9. Interaktive Effekte
(interactive effects;
carry-over effects)

Wird ein Individuum unter verschiedenen experimentellen Bedingungen untersucht und bleiben dabei Übertragungseffekte unkontrolliert, können dadurch die experimentellen Befunde verfälscht werden.

Das Experiment 6: Methode 3

Variablen & Messen

- Veränderliche Größen (X, Y); kategorial (diskret, dichotom, 3-wertig) vs. quantitativ (kontinuierlich)
- Skalenniveaus (sensu Stevens, 1957)
 - Nominal
 - $A = B$; Zahlwörter = Namen (Blutgruppe, Geschlecht, Staatsangehörigkeit)
 - Ordinal
 - $A > B$, aber +, -, *, : bedeuten nichts! (Extraversionsscore)
 - Intervall
 - $Y = aX + b$; $a-b = b-c$, aber beliebiger Nullpunkt (Celsius, IQ?), daher nur Verhältnisaussagen über Differenzen, nicht über Variables selbst ($10^{\circ}\text{C} - 3^{\circ}\text{C} = 17^{\circ}\text{C} - 10^{\circ}\text{C}$, aber: $3^{\circ}\text{C} * 3^{\circ}\text{C} \neq 3^{\circ}\text{F} * 3^{\circ}\text{F}$)
 - Ratio
 - $Y = aX$; $a/b = b/c$ (Masse, Körpergröße, Gewicht, Tonhöhe-MEL?, Lautheit/SON?)

Das Experiment 6: Methode 3

Variablen & Messen 2

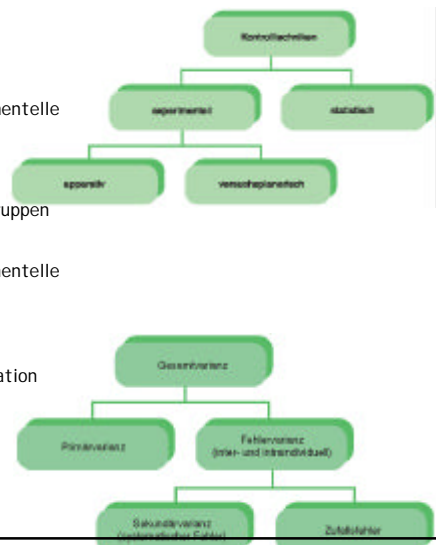
- Abhängige & Unabhängige Variablen (AV, UV)
 - $R = f(S)$; $S = UV = \text{unabhängig von der Versuchsperson (VP)}$; $R = AV = \text{abhängig von VP}$
 - $Y = f(X)$; AV wird gemessen, UV wird systematisch, willkürlich variiert
 - $Y = \{\text{Reaktionszeit, Fehlerrate, EEG Amplitude, BOLD Reaktion}\}$
 - $X = \{\text{Reizintensität, Reizvertrautheit}\}$
- Kontrollierte Variablen (KV), MaxKonMin Prinzip
 - Maximierung des Effektes der UV (Extrembedingungen, Mehrfaktorielle Designs, Faktoren mit mehreren Stufen)
 - Minimierung des Effektes von unsystematischen Störvariablen (Rauschen) durch Wiederholungsmessung, Blockbildung
 - Kontrolle systematischer Störvariablen (Abschirmung / Schallkammer, Eliminierung / Dunkelraum, Konstanthaltung / IQ Messung mit konstanter Leistungsmotivation)
 - Randomisierung (zufällige Abfolge der $S_1 - S_n$)
 - Parallelisierung (matching; Gruppen von VPn mit gleichem Alter, IQ, ...)

Sarris/Reiss: Kurzer Leitfaden der Experimentalpsychologie



Tabelle 2.1: Übersicht über die typischen Maßnahmen zur Gewährleistung der internen Validität gemäß dem Max-Kon-Min-Prinzip der Versuchsplanung

- 1. *Maximiere* Primärvarianz („Signal“)
 - Wahl von Extremgruppen
 - Wahl von so genannten optimalen Stufen
 - Umwandlung eines Störfaktors in eine weitere experimentelle UV („Kontrollvariable“)
- 2. Kontrolliere Sekundärvarianz (systematische Fehler)
 - Eliminierung eines Störfaktors
 - Konstanthaltung eines Störfaktors für alle Versuchsgruppen und experimentellen Bedingungen
 - Randomisierung der Probanden und der Bedingungen
 - Umwandlung eines Störfaktors in eine weitere experimentelle UV („Kontrollvariable“)
 - Nachträgliche statistische Kontrolle: Kovarianzanalyse
- 3. *Minimiere* Fehlervarianz („Rauschen“)
 - Wahl eines Wiederholungs- oder Block-Versuchsplans
 - Anheben der Standardisierung der Untersuchungssituation
 - Erhöhung der Zuverlässigkeit (und Gültigkeit) des Messinstruments



Das Experiment 7: Methode 4

Validitätsfragen

- Konstruktvalidität des Experiments (es mißt, was es zu messen vorgibt) ist gegeben wenn:
 - UV in psychologisch inhaltlich eindeutiger Weise auf AV wirkt und somit keine Konfundierung mit psychologisch irrelevanten Variablen vorliegt
 - UV & AV konstruktadäquat operationalisiert sind
 - Bsp.: Konstrukt „Angst“ als UV wird nur indirekt (über angstinduzierendes Verhalten des Versuchsleiters) operationalisiert
- Interne Validität
 - AV muß artefaktfrei auf UV zurückführbar sein
 - Typische Artefakte
 - Ermüdung, Übung, Stichprobenselektion,
 - Objektivität, Sensibilität & Reliabilität des Meßinstruments
- Externe Validität
 - wenn gesamte Versuchsanordnung (incl. UV, AV, ...) auch für natürliche Situationen Gültigkeit hat
- Inferenzstatistische Validität
 - Wenn von Stichprobe auf Population geschlossen werden kann (repräsentativ)

Gütekriterien instrumenteller Messungen

- Ein Verfahren ist messtechnisch *valide*, wenn es das zu Messende auch tatsächlich inhaltlich misst
- Als Voraussetzung für die Validität des Registrierungsverfahrens gelten die beiden Testkonstruktionskriterien der *Objektivität* und *Reliabilität*.

- *Objektivität*

- Messergebnisse dürfen nicht vom Untersucher selbst abhängig sein
- d.h. Verfahren muss so geartet sein, dass es intersubjektiv übereinstimmende Auswertung ermöglicht

- *Reliabilität*

- *Reliabilität* meint die (messtechnische) Zuverlässigkeit des Verfahrens
- Ein vollkommen reliables Verfahren führt zu verschiedenen Zeitpunkten im Prinzip zu identischen Messwerten für ein und denselben Sachverhalt
- In der Regel haben die in der Psychologie angewendeten Instrumente typischerweise einen „*Messfehler*“, der ein Maß der Variation der Messwerte bei wiederholten Messungen ist
- Bei der Datenerhebung ist naturgemäß darauf zu achten, dass nur solche Verfahren benutzt werden, deren Messfehler möglichst klein sind.

Die Instruktion *informiert* den Versuchsteilnehmer über

- den *Untersuchungsgegenstand* (allgemeiner Zweck des Versuches)
- die *Versuchsanordnung* (z.B. Funktion verschiedener für den Versuchsteilnehmer sichtbarer Apparate)
- die konkrete *Aufgabe*

Die Instruktion *motiviert* den Versuchsteilnehmer

- etwas Bestimmtes im Sinne der psychologischen Fragestellung zu leisten
- eine bestimmte Haltung während des Versuches einzunehmen (Entspannung, Aufmerksamkeit, Kooperation usw.)

Das Experiment 8: Befunde / Auswertung

- Visuelle Dateninspektion (JMP)
 - Korrelation über alle Datenpunkte, $Y = f(G, T)$
 - Effektdiagramme (bei $S \times T$ Design): $m_2 - m_1$ über alle VPn
- Deskriptive Statistik
 - Kennwerte der zentralen Tendenz
 - Kennwerte der Variabilität
 - Häufigkeitsverteilungen (Histogramme)
- Inferenzstatistik
 - Regression
 - Korrelation
 - Varianzanalyse
- Mathematische oder Computermodellierung (Kurvenanpassung)