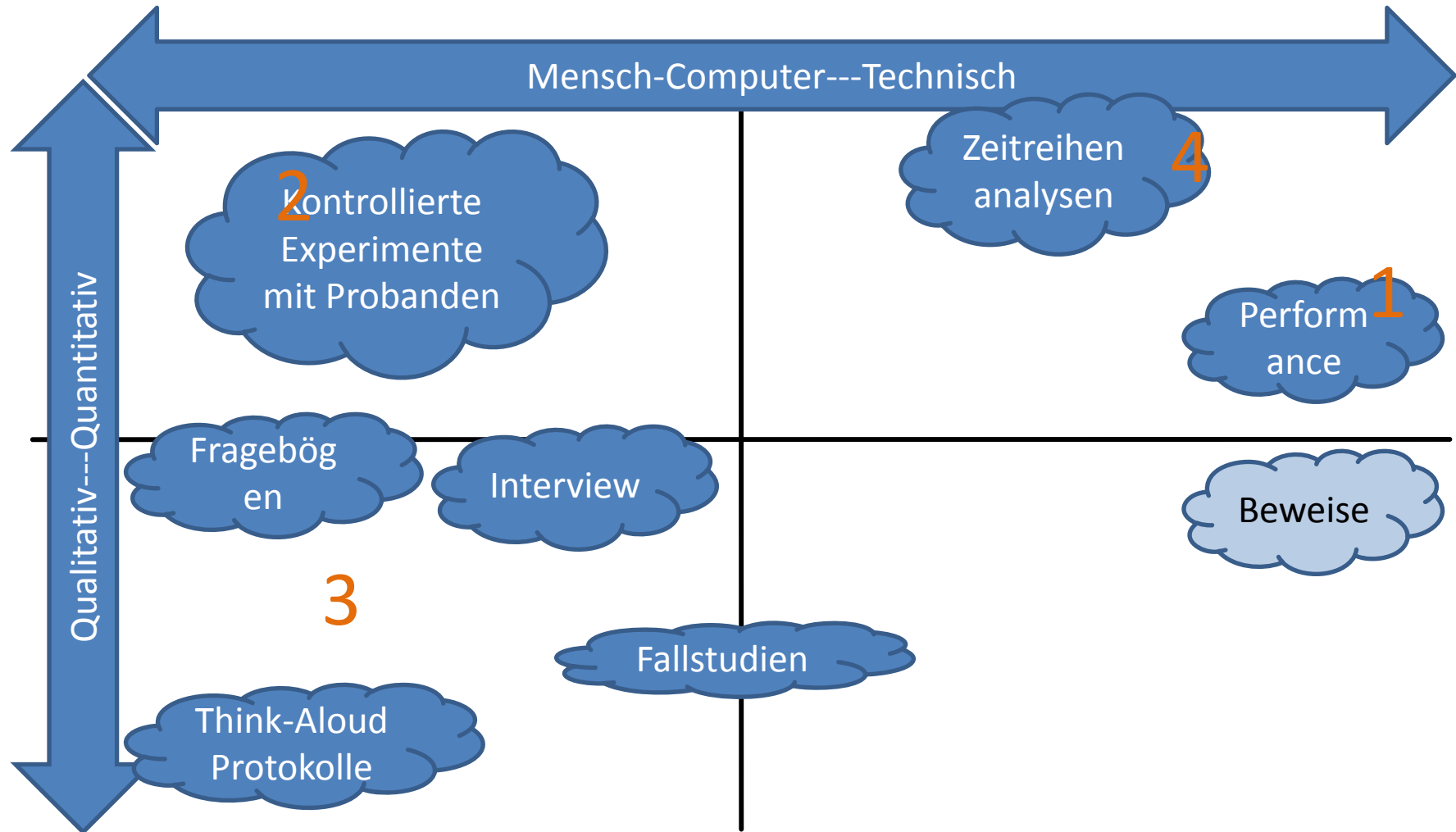


Kontrollierte Experimente

Einordnung



Lernziele

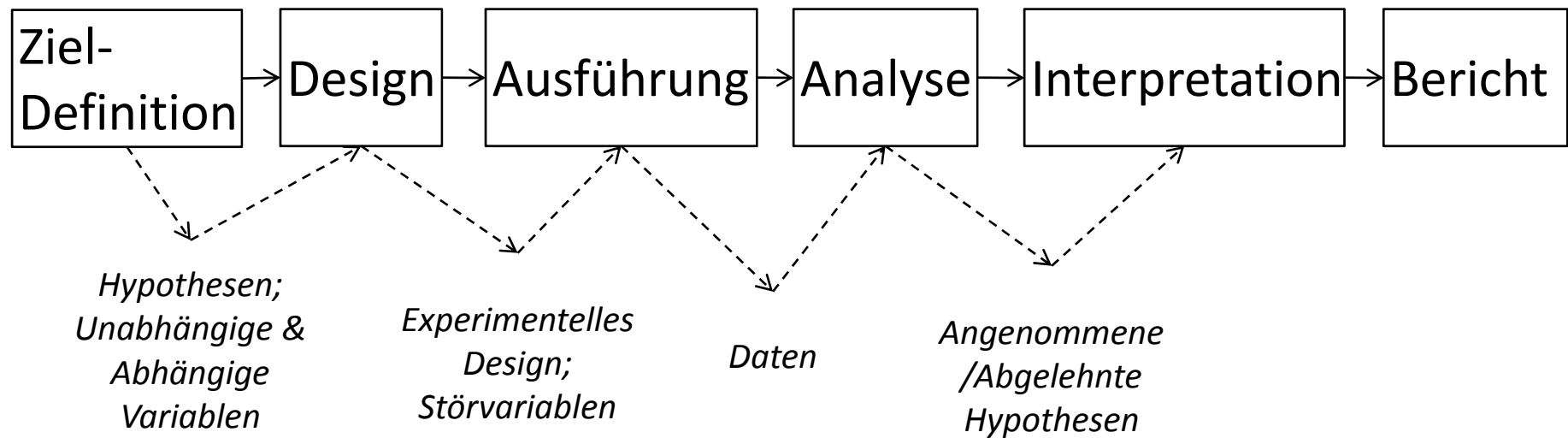
- Gute Hypothesen aufstellen können
- Experiment mit hoher interner oder hoher externer Validität entwerfen können

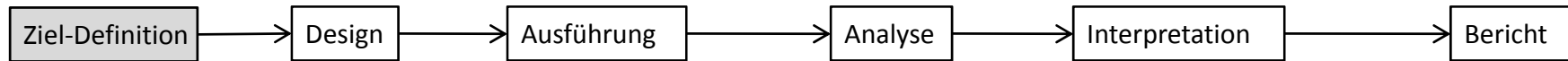


Definition

- Systematische Studie
- Ein oder mehrere Faktoren werden variiert
- Alles andere konstant halten
- Ergebnis der systematischen Variation wird beobachtet

Experimentelle Phasen





Variablen

Unabhängige Variablen

- Absichtlich, systematisch variiert durch Versuchsleiter
- Faktor, Prädiktor (-variable)
- Alternativen, Level, Stufen, Treatment
- Beispiele:
 - Programmierparadigma
 - Sprache
 - Workload

Abhängige Variable

- Ergebnis eines Experiments
- Hängen ab von Variation der unabhängigen Variablen
- Beobachtet
- Beispiele:
 - Performance
 - Programmverständnis
 - Produktivität von Entwicklern

Latente Variablen

- Konstrukt
- Nicht direkt beobachtbar
- Beispiele
 - Programmverständnis
 - Intelligenz
 - Performance

Operationalisierung

- Operationen definieren, mit denen man Variablen messen kann
- Darf gesundem Menschenverstand nicht widersprechen

Aufgabe

- Operationale Definitionen für folgende Variablen
 - Performance
 - Programmverständnis
 - Intelligenz
 - Wartbarkeit

Hypothesen

- Erwartungen über Ergebnisse
- Erwartungen müssen begründet sein in Theorie oder Praxis
- Hypothesen müssen einfach und klar formuliert sein
- Hypothesen müssen überprüfbar sein
- Falsifizierbarkeit

Hypothese-Negativbeispiel

- Schlechte Kommentare sind schlecht für Programmverständnis
- Gute Kommentare sind gut für Programmverständnis

Besser

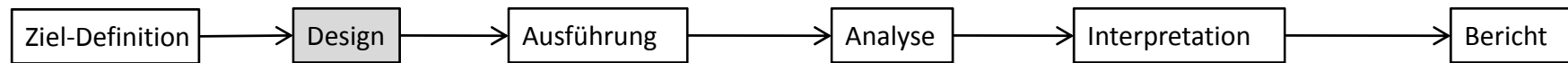
- Kommentare, die jedes Statement von Quelltext beschreiben, haben keinen Einfluss auf Antwortzeit beim Verstehen von Quelltext
- Kommentare, die falsche Informationen über Quelltext enthalten, verlangsamen Programmverständnis
- Kommentare, die den Zweck von Statements beschreiben, beschleunigen Programmverständnis

Wozu Hypothesen?

- Steuerungsfunktion
- Verhindert *Fishing for Results*
- Verbindung zwischen Theorie und Empirie
 - Abgeleitet aus Theorie
 - Überprüft mit Empirie

Aufgabe

- Stellen Sie je eine Hypothese zu folgenden Forschungsfragen auf:
 - Erhöht Objektorientierung die Produktivität von Entwicklern?
 - Ist Java besser als C++?
 - Welchen Einfluss hat Programmiererfahrung auf Entwicklungszeit?
- Die Hypothese muss überprüfbar sein, begründet sein; die Variablen müssen operationalisiert sein
= Beispiel für Prüfungsfrage



Design

Validität

- Wird das gemessen was gemessen werden soll?

Interne Validität

- Maß, in dem Wert der abhängigen Variablen auf Variation der unabhängigen Variablen zurückgeführt werden kann

Externe Validität

- Maß, in dem Ergebnisse aus einem Experiment auf andere Umstände (Probanden, Material,...) übertragen werden kann
- = Verallgemeinerbarkeit

Hausaufgabe

- Recherchieren Sie andere Validitätsarten

Gefahren/Bedrohungen

- Störvariablen:
 - Beeinflussen abhängige Variable zusätzlich zu unabhängiger Variablen
 - Lerneffekte
 - Hawthorne-Effekt
 - Messinstrumente
 - Selektion
 - ...

Aufgabe

- Messen von Programmverständnis: Welche Störvariablen gibt es?
- Wie könnte man diese Störvariablen kontrollieren?

Störvariablen

- Es gibt viele Störvariablen
- Sorgfältig identifizieren und kontrollieren
 - Randomisierung
 - Matching/Parallelisierung/Balancing
 - Störvariable als unabhängige Variable definieren
 - Störvariable konstant halten
 - Nachträgliche Analyse

Randomisierung

- Zufallszahlengenerator
 - Münze werfen
 - Würfeln
 - ...
-
- Probleme:
 - Gruppen müssen groß genug sein
 - 5 pro Gruppe zu wenig, 10 scheint akzeptabel

Matching/Parallelisierung/Balancing

Proband	Wert
P5	65
P9	56
P3	42
P4	34
P10	24
P6	23
P7	21
P8	16
P2	12
P1	5

Gruppe A	Gruppe B
65	56
34	42
24	23
16	21
12	6

odd-even-even-odd/
ABBA

Matching/Parallelisierung/Balancing

- Nachteil gegenüber Randomisierung:
 - Störvariable muss gemessen werden
 - Programmiererfahrung?
 - Intelligenz?

Parameter als unabhängige Variable definieren

- Wird systematisch vom Versuchsleiter variiert
- Störvariable wird operationalisiert
- Programmiererfahrung:
 - Viel/wenig Erfahrung
 - Viel/wenig Motivation
 - Java/C

Rechenbeispiel

- 23 Störvariablen, jede mit 2 Stufen
= 8 388 608 mögliche Kombinationen
- Wie viele Probanden sind nötig, um jede Kombination abzudecken?
 - min. 10 Probanden pro Gruppe
 - 83 886 080 (ganz Deutschland)

Konstant halten

- Nur ein Level einer Störvariable
- Programmierfahrung
 - Nur Bachelor-Studenten
 - Nur Programmierexperten
- Intelligenz
 - Nur Studenten mit bestimmter Note

Nachträgliche Analyse

- Variable wird während des Experiments gemessen
- Einfluss einer Variablen wird nach dem Experiment analysiert
- Probleme:
 - Kann zeigen, dass Ergebnisse unbrauchbar sind

Hausaufgabe

- Entwerfen Sie ein Experiment mit hoher interner Validität
- Entwerfen Sie ein Experiment mit hoher externer Validität

Verhältnis von Interner und Externer Validität

- Beide verlangen verschiedene Dinge
 - Intern: alles kontrollieren
 - Extern: allgemeines Setting
- Und jetzt?
 - Erst interne Validität maximieren
 - Dann schrittweise externe Validität erhöhen

Reliabilität

- Genauigkeit der Messinstrumente

Objektivität

- Durchführung eines Experiments darf nicht von Person der Versuchsleiter abhängen
- Dasselbe Experiment, durchgeführt von anderen Versuchsleitern, soll dasselbe Ergebnis liefern