

Mensch-Maschine-Interaktion



Kapitel 13 - Evaluation

- Arten der Evaluation

- Formativ vs. Summativ
- Quantitativ vs. Qualitativ
- Analytisch vs. Empirisch

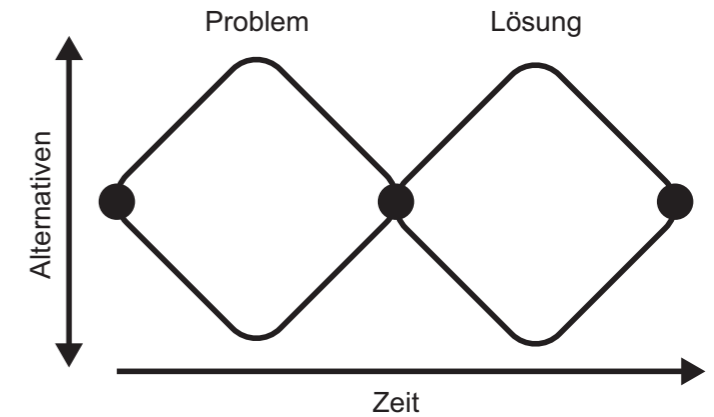
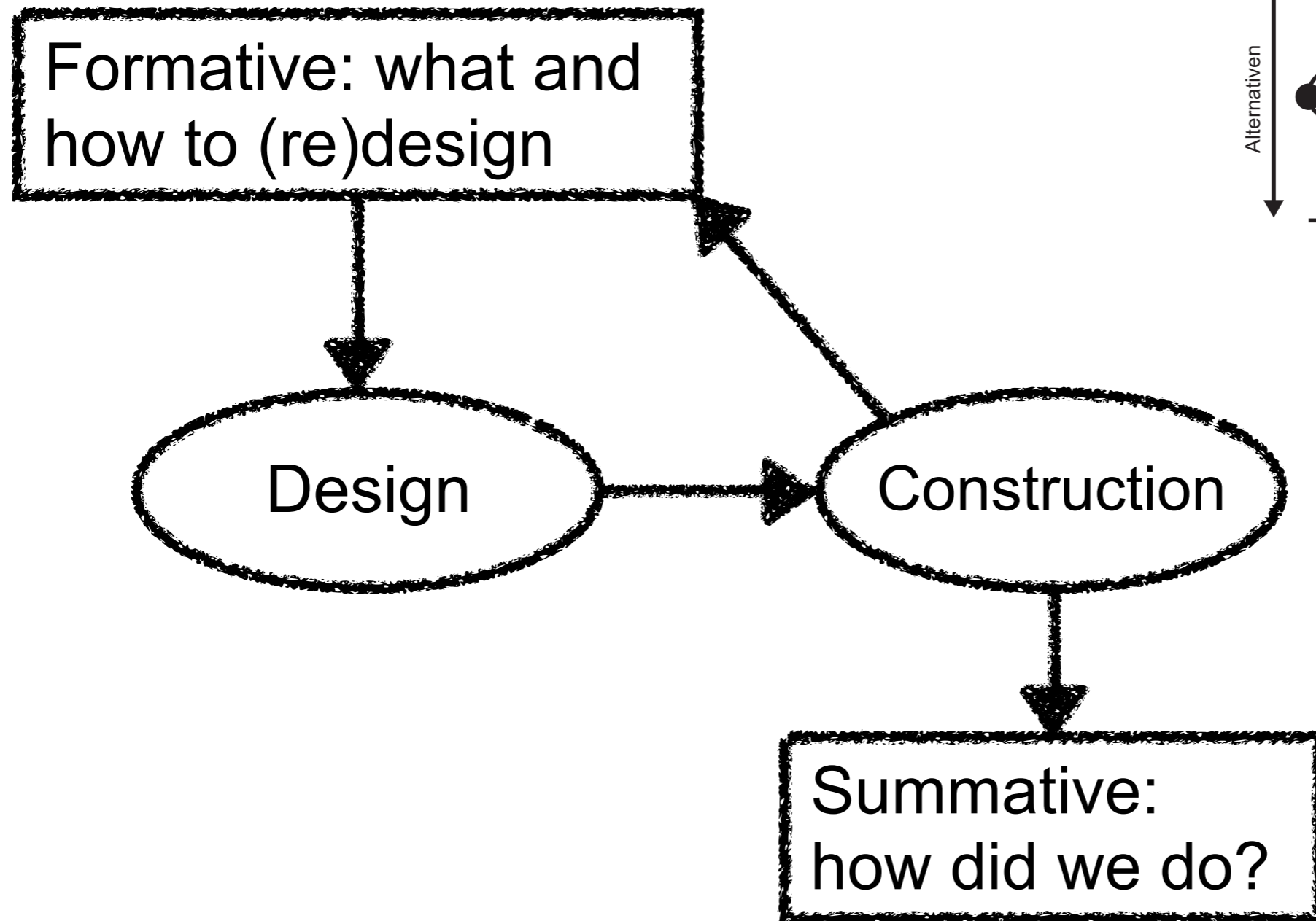
- Analytische Methoden

- Cognitive Walkthrough
- Heuristische Evaluation
- GOMS und KLM

- Empirische Methoden

- Variablen und Werte
- Probanden
- Beobachtungsstudien
- Kontrollierte Experimente
- Darstellung der Ergebnisse
- Statistische Auswertung
- Feldstudien und Laborstudien
- Langzeit- und Tagebuch-Studien

Formative vs. Summative Evaluation



- M. Scriven: The methodology of evaluation, 1967

Qualitative vs. Quantitative Evaluation



<http://www.scope-mr.ch/de/dienstleistungen/methoden/>



<http://www.scope-mr.ch/de/dienstleistungen/methoden/>



<https://studi-lektor.de/diplomarbeit/image/tipps/quantitativ-qualitativ.png>

Analytische vs. Empirische Evaluation

Scriven, 1967: “Will man ein Werkzeug, wie z.B. eine Axt evaluieren, dann kann man dies entweder analytisch tun und untersuchen, welcher Stahl für die Haue und welches Holz für den Griff verwendet wurde, wie die Balance der Axt oder die Schärfe der Klinge ist, oder man wählt den empirischen Weg und misst, wie gut die Axt in der Hand eines guten Holzfällers ihren Zweck erfüllt, also wieviel Holz er in gegebener Zeit damit fällt.”



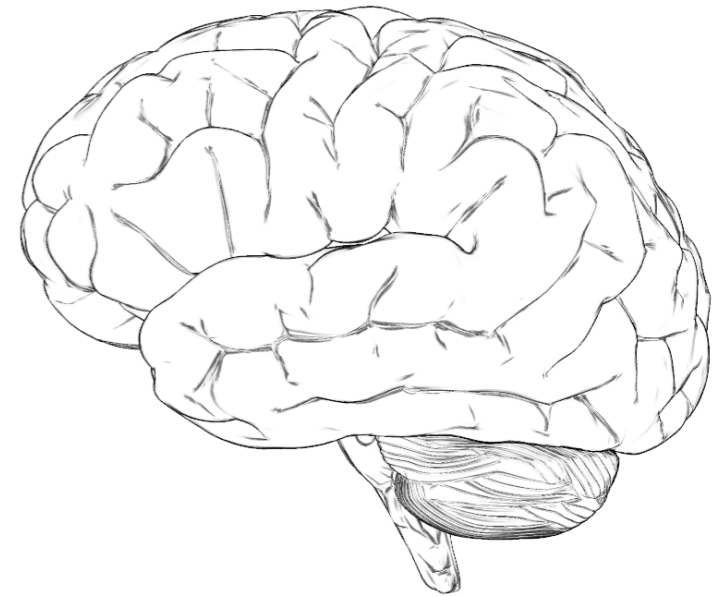
Kapitel 13 - Evaluation

- Arten der Evaluation
 - Formativ vs. Summativ
 - Quantitativ vs. Qualitativ
 - Analytisch vs. Empirisch
- Analytische Methoden
 - Cognitive Walkthrough
 - Heuristische Evaluation
 - GOMS und KLM
- Empirische Methoden
 - Variablen und Werte
 - Probanden
 - Beobachtungsstudien
 - Kontrollierte Experimente
 - Darstellung der Ergebnisse
 - Statistische Auswertung
 - Feldstudien und Laborstudien
 - Langzeit- und Tagebuch-Studien

Cognitive Walkthrough

...Schritt für Schritt...

...entlang definierter Aufgaben...



1. Ist die **korrekte Aktion** zur Ausführung einer Handlung ausreichend **klar**? Weiß der Benutzer überhaupt, was er tun soll?
2. Ist die korrekte Aktion als solche **erkennbar**? Findet sie der Benutzer?
3. Erhält der Benutzer eine ausreichende **Rückmeldung** nach Ausführung der Aktion, so dass er erkennen kann, dass die Handlung erfolgreich durchgeführt ist?

10 Usability Heuristics

- Visibility of system status
- Match between system and the real world
- User control and freedom
- Consistency and standards
- Error prevention
- Recognition rather than recall
- Flexibility and efficiency of use
- Aesthetic and minimalist design
- Help users recognize, diagnose, and recover from errors
- Help and documentation



Jakob Nielsen

Beispiel für eine detaillierte Checkliste

Usability Techniques Heuristic Evaluation - A System Checklist

By Denise Pierotti, Xerox Corporation

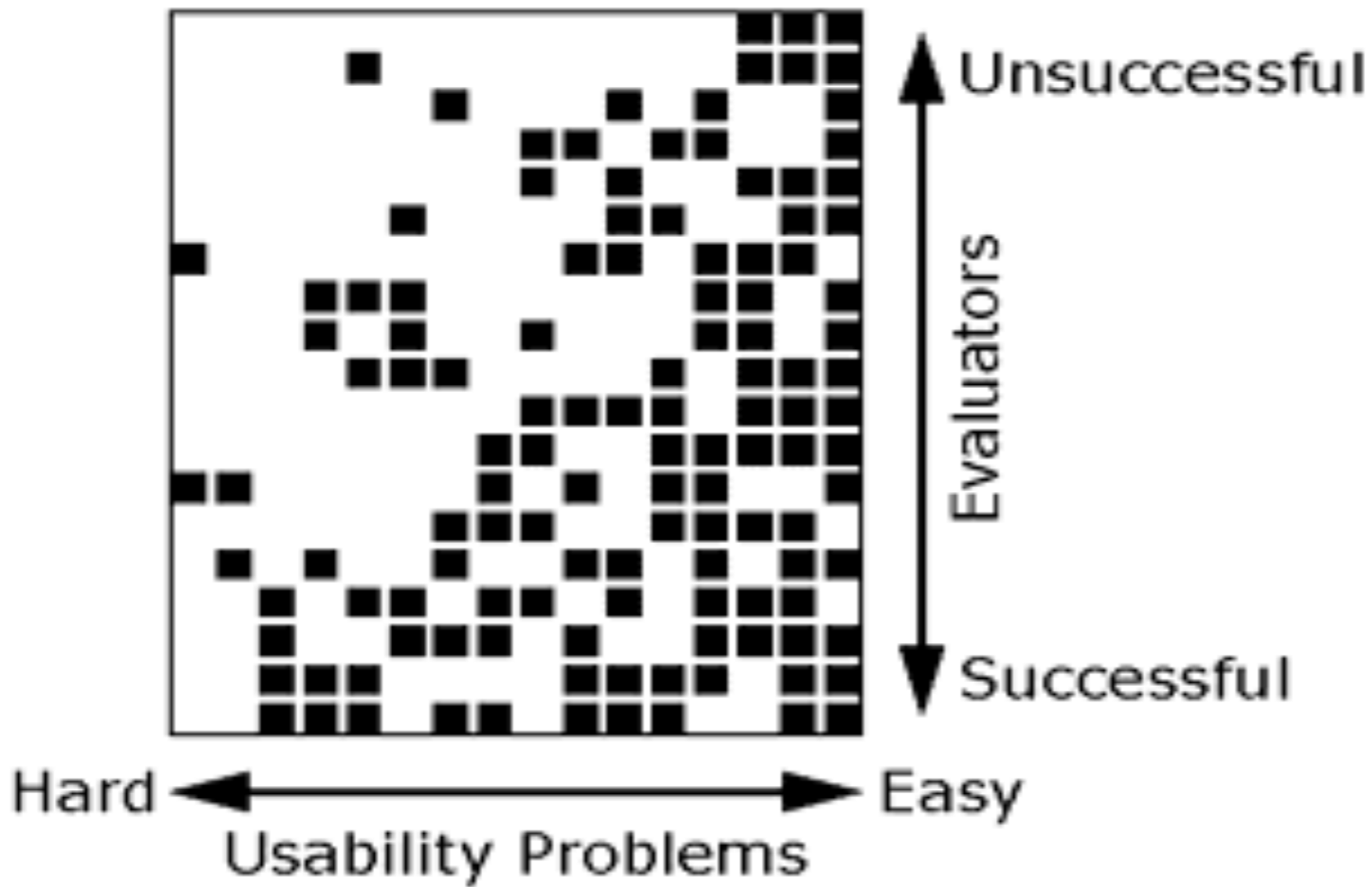
Heuristic Evaluation - A System Checklist

<http://www.stcsig.org/usability/topics/articles/he-checklist.html>

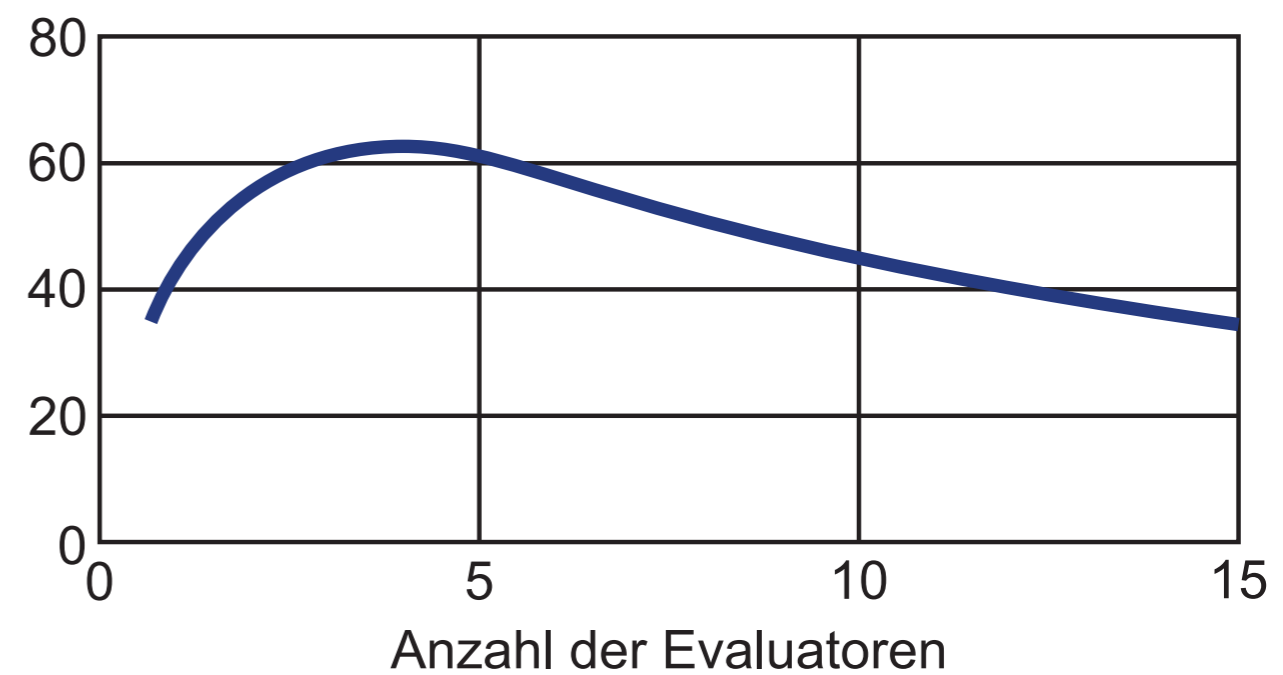
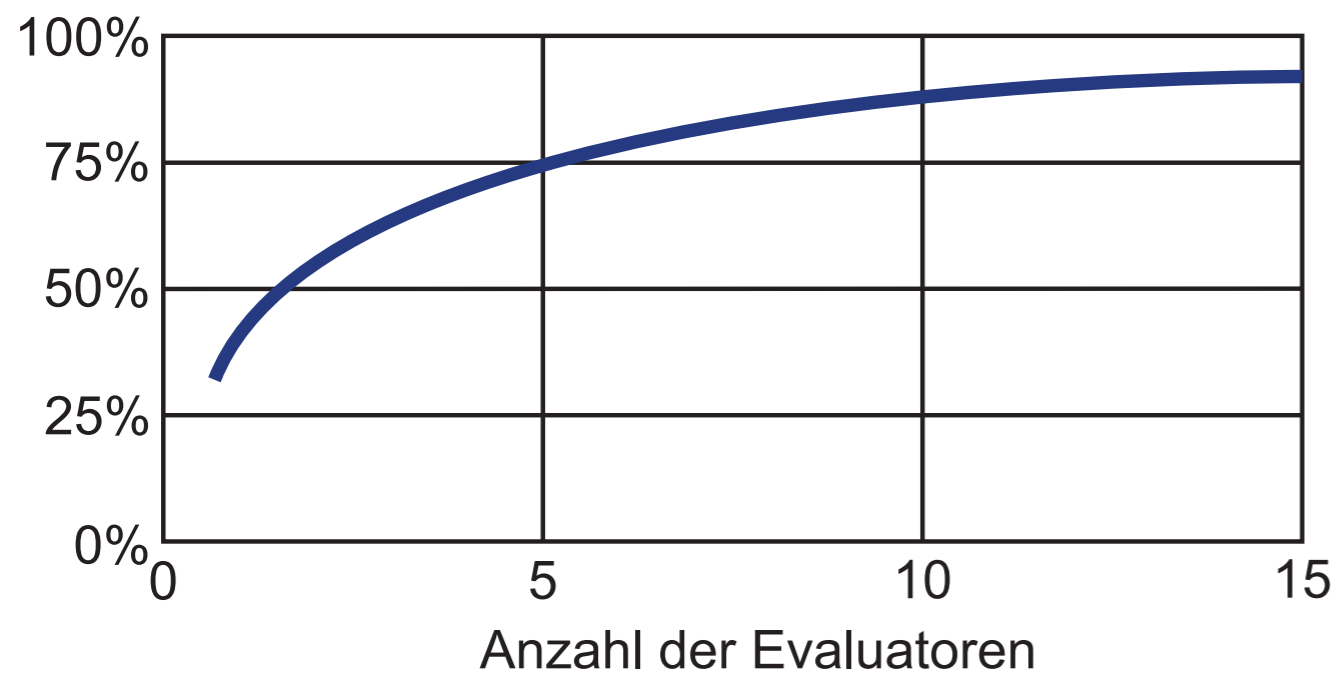
1. Visibility of System Status

The system should always keep user informed about what is going on, through appropriate feedback within reasonable time.

#	Review Checklist	Yes No N/A	Comments
1.1	Does every display begin with a title or header that describes screen contents?	<input type="radio"/> <input type="radio"/> <input type="radio"/>	
1.2	Is there a consistent icon design scheme and stylistic treatment across the system?	<input type="radio"/> <input type="radio"/> <input type="radio"/>	
1.3	Is a single, selected icon clearly visible when surrounded by unselected icons?	<input type="radio"/> <input type="radio"/> <input type="radio"/>	
1.4	Do menu instructions, prompts, and error messages appear in the same place(s) on each menu?	<input type="radio"/> <input type="radio"/> <input type="radio"/>	
1.5	In multipage data entry screens, is each page labeled to show its relation to others?	<input type="radio"/> <input type="radio"/> <input type="radio"/>	
1.6	If overwrite and insert mode are both available, is there a visible indication of which one the user is in?	<input type="radio"/> <input type="radio"/> <input type="radio"/>	
1.7	If pop-up windows are used to display error messages, do they allow the user to see the field in error?	<input type="radio"/> <input type="radio"/> <input type="radio"/>	
1.8	Is there some form of system feedback for every operator action?	<input type="radio"/> <input type="radio"/> <input type="radio"/>	
1.9	After the user completes an action (or group of actions), does the feedback indicate that the next group of actions can be started?	<input type="radio"/> <input type="radio"/> <input type="radio"/>	
1.10	Is there visual feedback in menus or dialog boxes about which choices are selectable?	<input type="radio"/> <input type="radio"/> <input type="radio"/>	
1.11	Is there visual feedback in menus or dialog boxes about which choice the cursor is on now?	<input type="radio"/> <input type="radio"/> <input type="radio"/>	



Jakob Nielsen



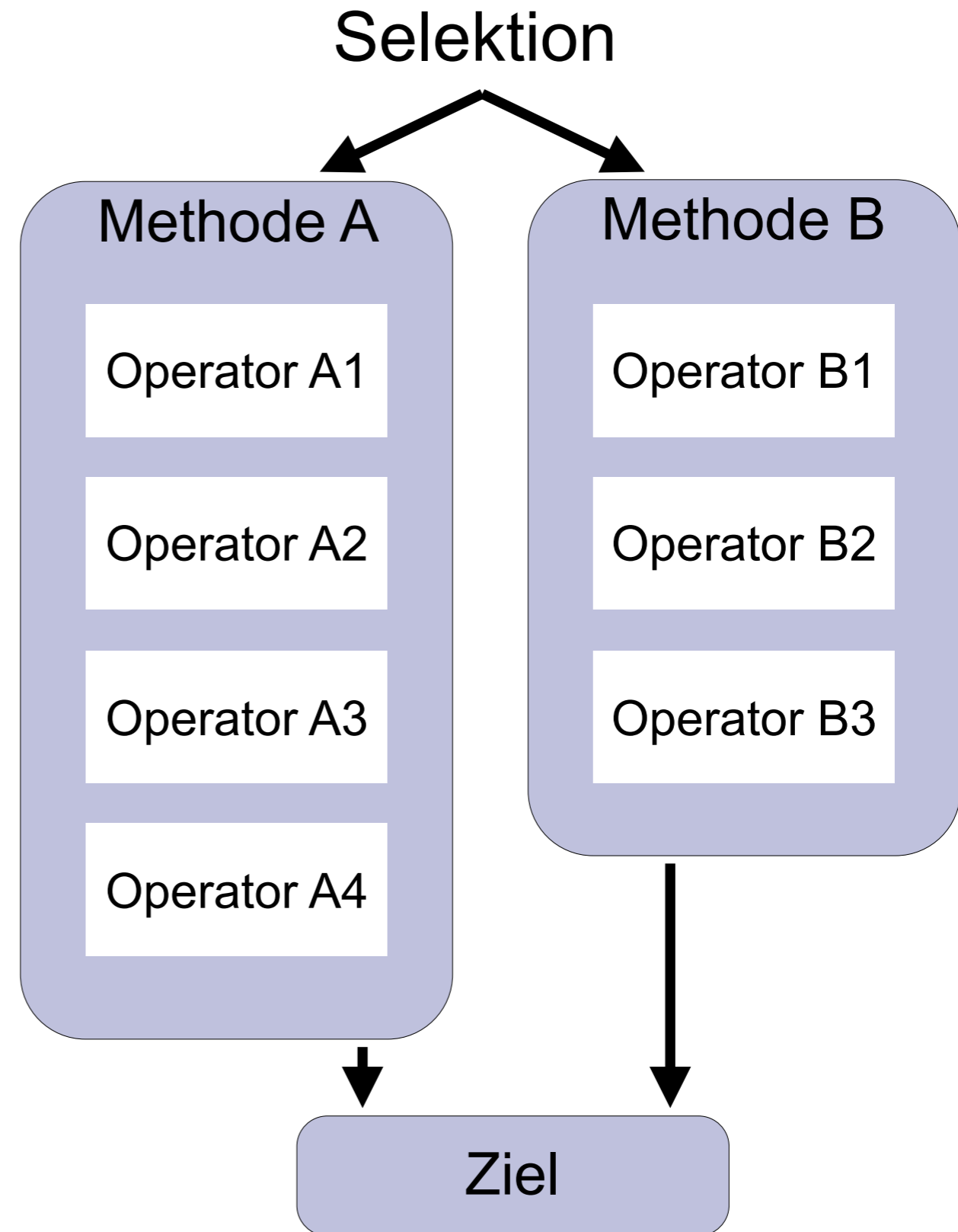
Goals, Operators, Methods & Selection Rules (GOMS)

- **Selection rules**

- **Methods**

- **Operators**

- **Goals**



Keystroke Level Model (KLM)

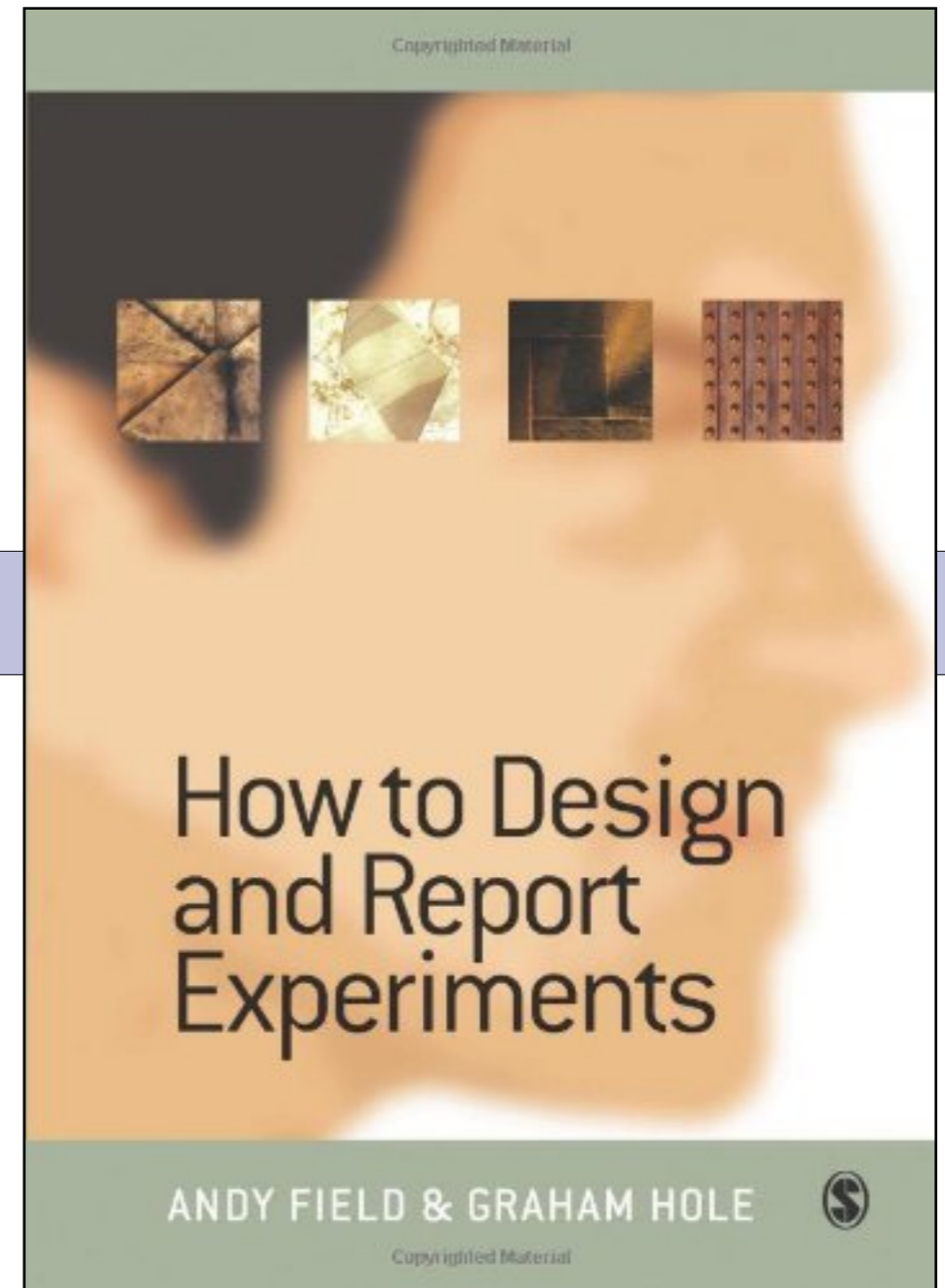
- Benötigte Zeiten im experimentellen Durchschnitt:
- **K** (Keystroke): Drücken einer Taste: **$t_K = 0.28s$** .
- **P** (Pointing): Zeigen auf eine Bildschirmposition: **$t_P = 1.1s$**
- **H** (Homing): Wechsel zw. Tastatur und Maus: **$t_H = 0.4s$**
- **M** (Mental preparation): Geistiges Vorbereiten einer nachfolgenden Operation: **$t_M = 1.35s$**
- **R(t)** (Response time t by the system): Antwortzeit des Systems von **t** Sekunden.

KLM Beispiel

- Welche der Methoden M1 oder M2 ist schneller?
- **M1:** Wechsel zur Maus, Bewegen des Mauszeigers zur Datei, Anklicken der Datei, Ziehen zum Mülleimer und Loslassen, Wechsel zur Tastatur
- **M2:** Wechsel zur Maus, Selektieren der Datei, Wechsel zur Tastatur, Taste Entf.
- **$t_{M1} = t_H + t_P + t_K + t_P + t_H = 0.4 + 1.1 + 0.28 + 1.1 + 0.4 = 3.28s$**
- **$t_{M2} = t_H + t_P + t_H + t_K = 0.4 + 1.1 + 0.4 + 0.28 = 2.18s$**

Kapitel 13 - Evaluation

- Arten der Evaluation
 - Formativ vs. Summativ
 - Quantitativ vs. Qualitativ
 - Analytisch vs. Empirisch
- Analytische Methoden
 - Cognitive Walkthrough
 - Heuristische Evaluation
 - GOMS und KLM
- Empirische Methoden
 - Variablen und Werte
 - Probanden
 - Beobachtungsstudien
 - Kontrollierte Experimente
 - Darstellung der Ergebnisse
 - Statistische Auswertung
 - Feldstudien und Laborstudien
 - Langzeit- und Tagebuch-Studien



<http://www.amazon.de/dp/0857028294>

Qualitätsmerkmale Empirischer Methoden

- Objektivität

- Reproduzierbarkeit

- Validität

 - intern

 - extern

 - ökologisch

- Relevanz



<http://www.schwimmvereinapolda.de/images/Webelemente/Stoppuhr.jpg>



http://www.bgr.bund.de/DE/Themen/Endlagerung/Bilder/end_nfpro_hyperf_g.jpg?__blob=normal&v=2

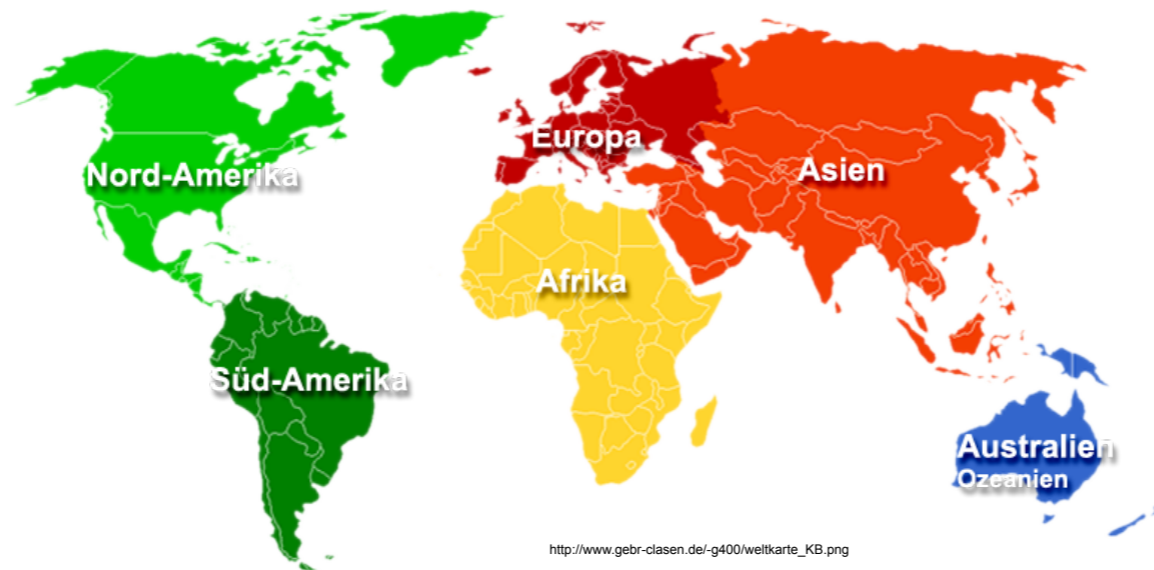
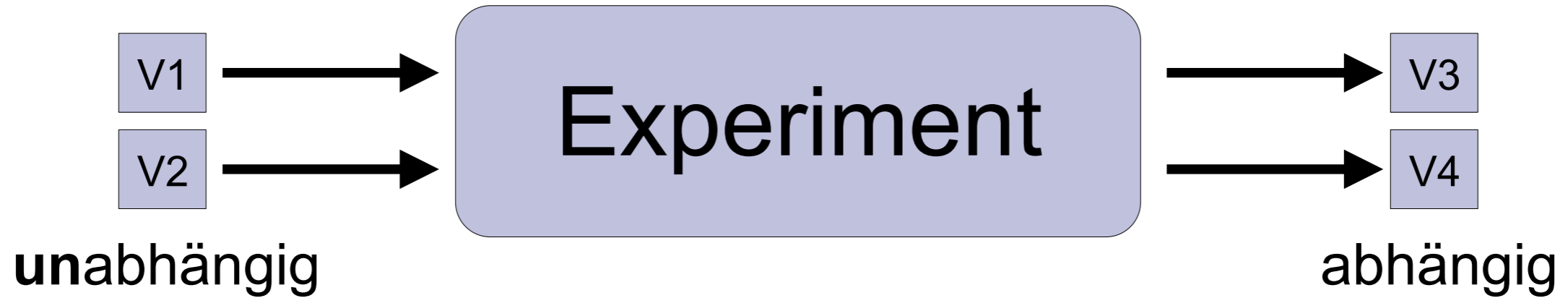


<http://w115www815.webland.ch/travelinfos/images/mensch/gehirn4.jpg>



http://bilder.n3po.com/cache/Photos/Bach-Fluessend-Bergab_w475_h230_cw475_ch230_thumb.jpg

Variablen und Werte

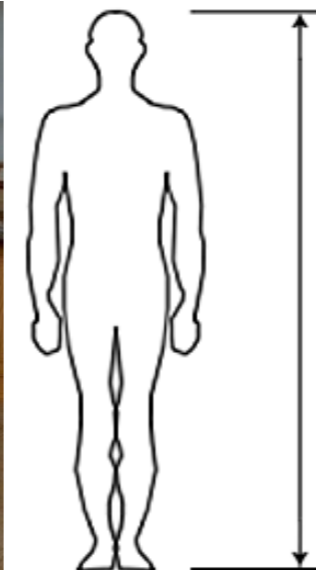


TABELLE

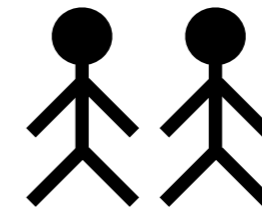
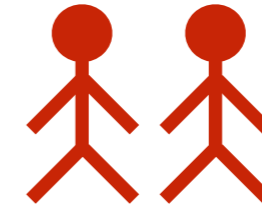
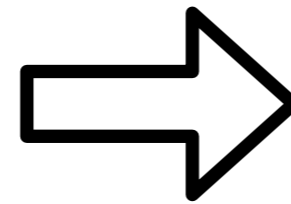
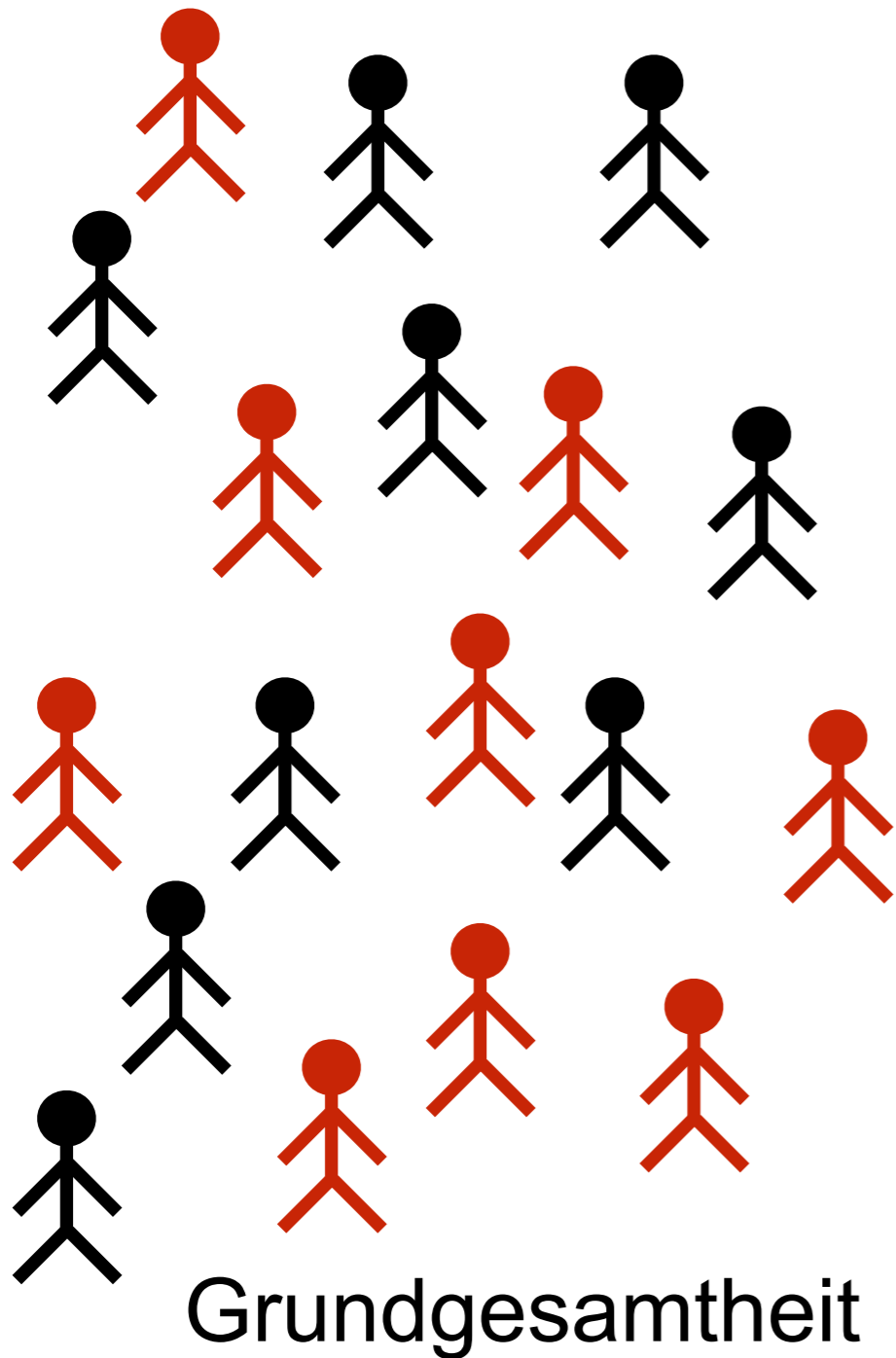
BUNDESLIGA		2. BUNDESLIGA		
PL	VEREIN	SPIELE	TD	PKT
1	FC Bayern	0	0	0
	Borussia Dortmund	0	0	0
	Schalke 04	0	0	0
	Bayer 04	0	0	0
	VfL Wolfsburg	0	0	0
	Borussia M'gladbach	0	0	0
	Mainz 05	0	0	0
	FC Augsburg	0	0	0
	1899 Hoffenheim	0	0	0
	Hannover 96	0	0	0
	Hertha BSC	0	0	0
	SV Werder	0	0	0
	Eintracht Frankfurt	0	0	0
	SC Freiburg	0	0	0
	VfB Stuttgart	0	0	0
	Hamburger SV	0	0	0
	1. FC Köln	0	0	0
	SC Paderborn	0	0	0

<http://www.bundesliga.de/>

- Nominal
- Ordinal
- Kardinal



Probanden



Stichprobe

- Alter
- Geschlecht
- Vorkenntnisse
- Händigkeit
- Sehkraft
- Bildung
- Nationalität ...

Beobachtungsstudie (Bsp.)



- 1 Unabhängige Variable: Übungsteilnahme (Ja / Nein)
– freiwillig
- 1 Abhängige Variable: Note

- 108 Probanden, 54 Teilnehmer, 54 Nicht-Teilnehmer
- Messung ergibt: Note positiv korreliert mit Teilnahme

Kontrolliertes Experiment



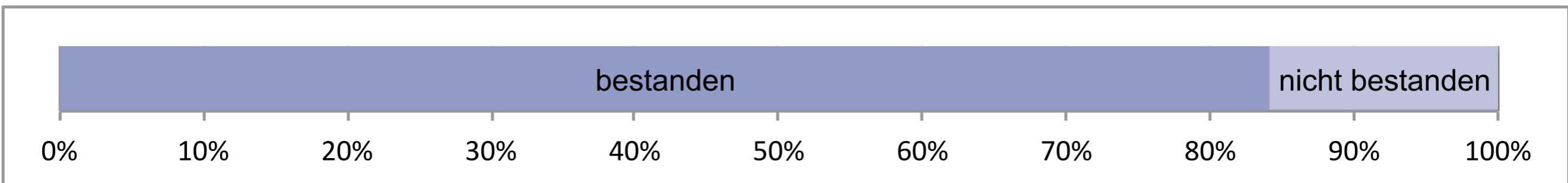
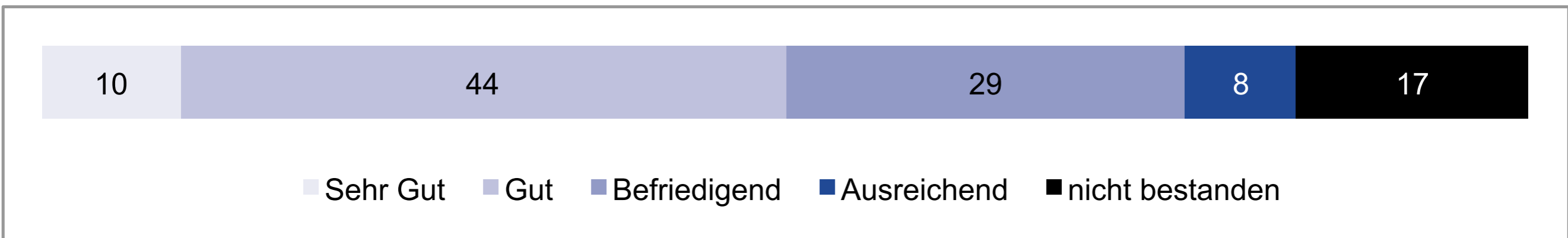
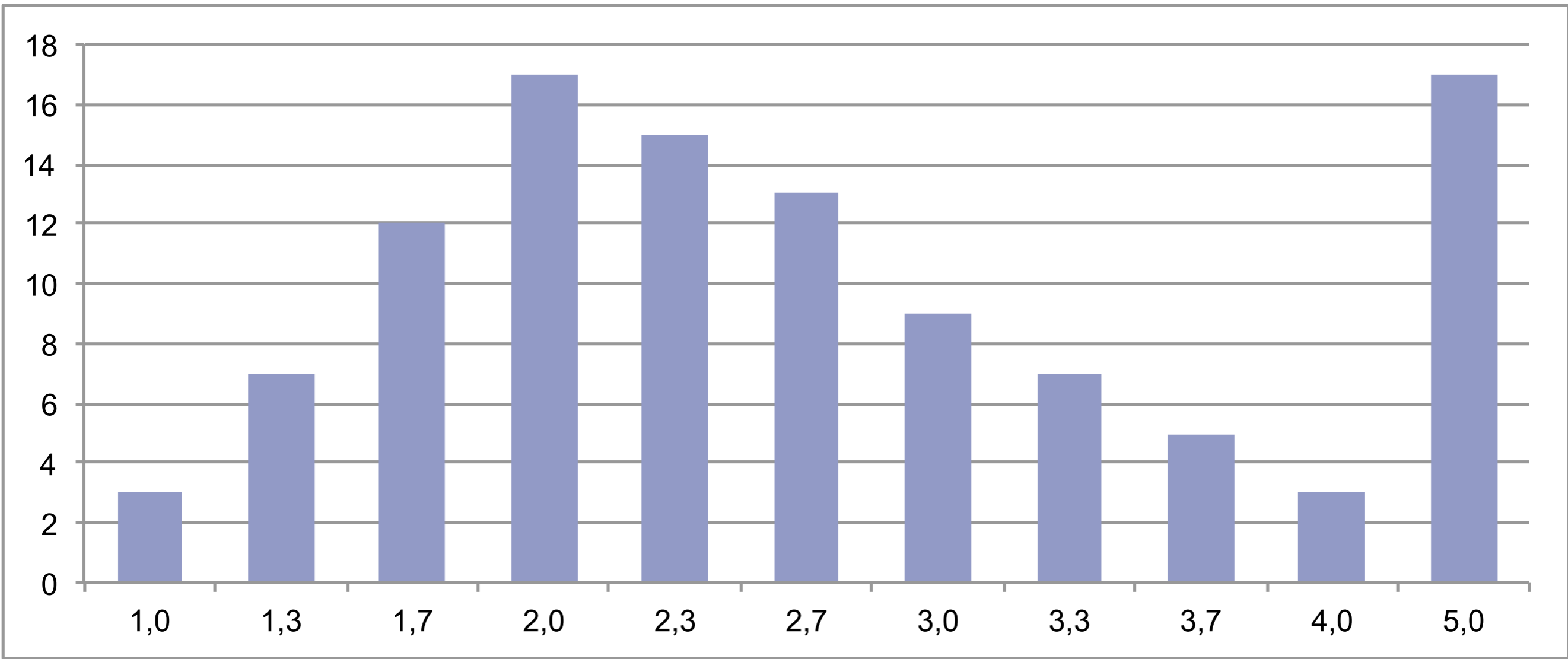
- 1 Unabhängige Variable: Übungsteilnahme (Ja / Nein)
– zufällig zugewiesen !!!
- 1 Abhängige Variable: Note
- 108 Probanden, 54 Teilnehmer, 54 Nicht-Teilnehmer
- Messung ergibt: Note positiv korreliert mit Teilnahme
- Nun folgt: Übungsteilnahme begründet bessere Note!

Experimentaldesign

	MMI1	Analysis	Algebra
Ja	Bedingung 1	Bedingung 2	Bedingung 3
Nein	Bedingung 4	Bedingung 5	Bedingung 6

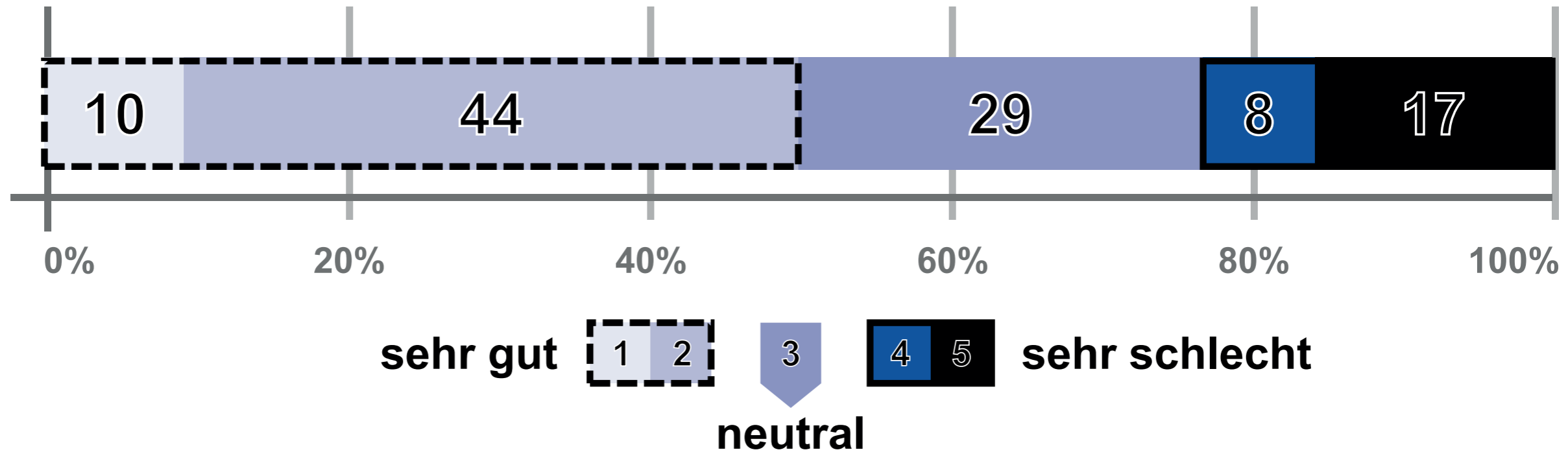
- 2 Variablen mit 2 und 3 Werten: $2 \times 3 = 6$ Bedingungen
- **within-subjects**: jeder macht alles
- **between-groups**: je eine Gruppe macht 1 Bedingung
- Reihenfolge abwechseln wg. **Lern-** und **Ermüdungseffekt**
 - Randomisierung, Permutation, lat. Quadrat

Bed. 6	Bed. 1	Bed. 5	Bed. 2	Bed. 4	Bed. 3
Bed. 5	Bed. 6	Bed. 4	Bed. 1	Bed. 3	Bed. 2
Bed. 2	Bed. 3	Bed. 1	Bed. 4	Bed. 6	Bed. 5
Bed. 1	Bed. 2	Bed. 6	Bed. 3	Bed. 5	Bed. 4
Bed. 4	Bed. 5	Bed. 3	Bed. 6	Bed. 2	Bed. 1
Bed. 3	Bed. 4	Bed. 2	Bed. 5	Bed. 1	Bed. 6



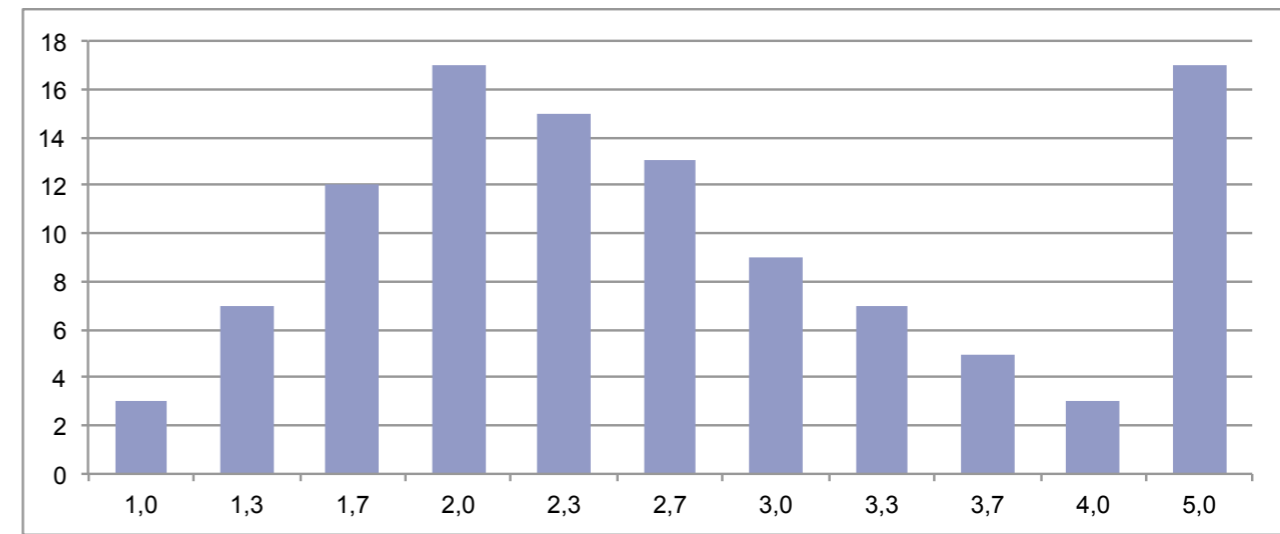
Wie fanden Sie die Vorlesung?

MMI1



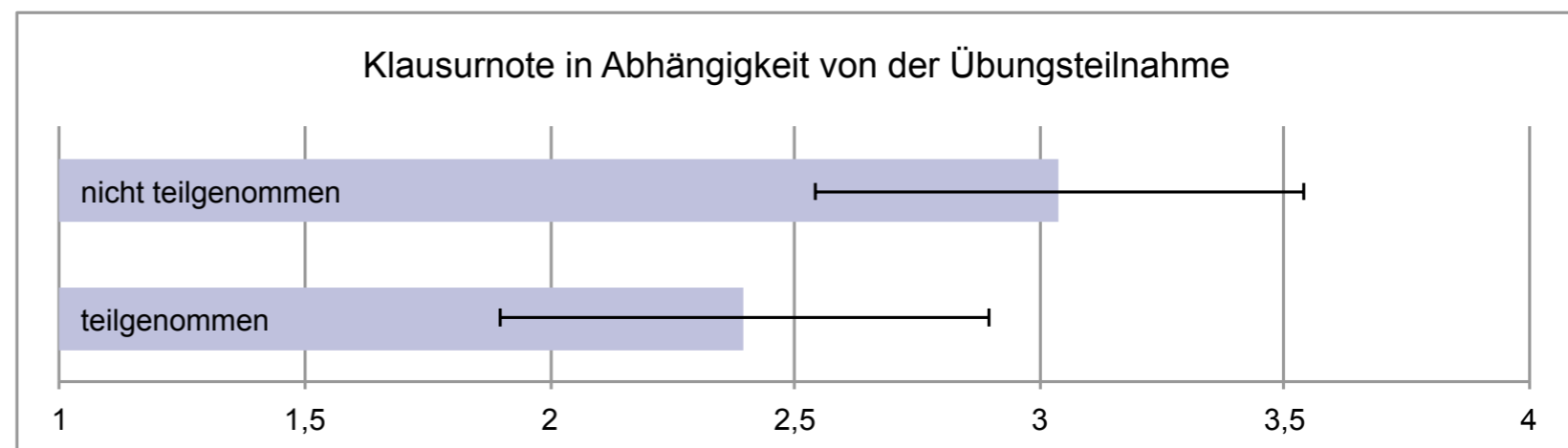
Online-Tool <http://www.likertplot.com>

Deskriptive Statistik



- nominale Daten: Modus $(2, 4, 5, 5, 5, 5, 5) = 5$
- ordinale Daten: Median $(2, 4, 5, 5, 5, 5, 5) = 5$
- kardinale Daten: Mittelwert $(2, 4, 5, 5, 5, 5, 5) = 31/7 = 4,42$
- Standardabweichung:
 - Median(1,2,3,4,5) = Median(3,3,3,3,3) = 3
 - Mittelwert(1,2,3,4,5) = Mittelwert(3,3,3,3,3) = 3
 - $\sigma(1,2,3,4,5) = 1,58$
 - $\sigma(3,3,3,3,3) = 0,0$

$$\sigma = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}$$



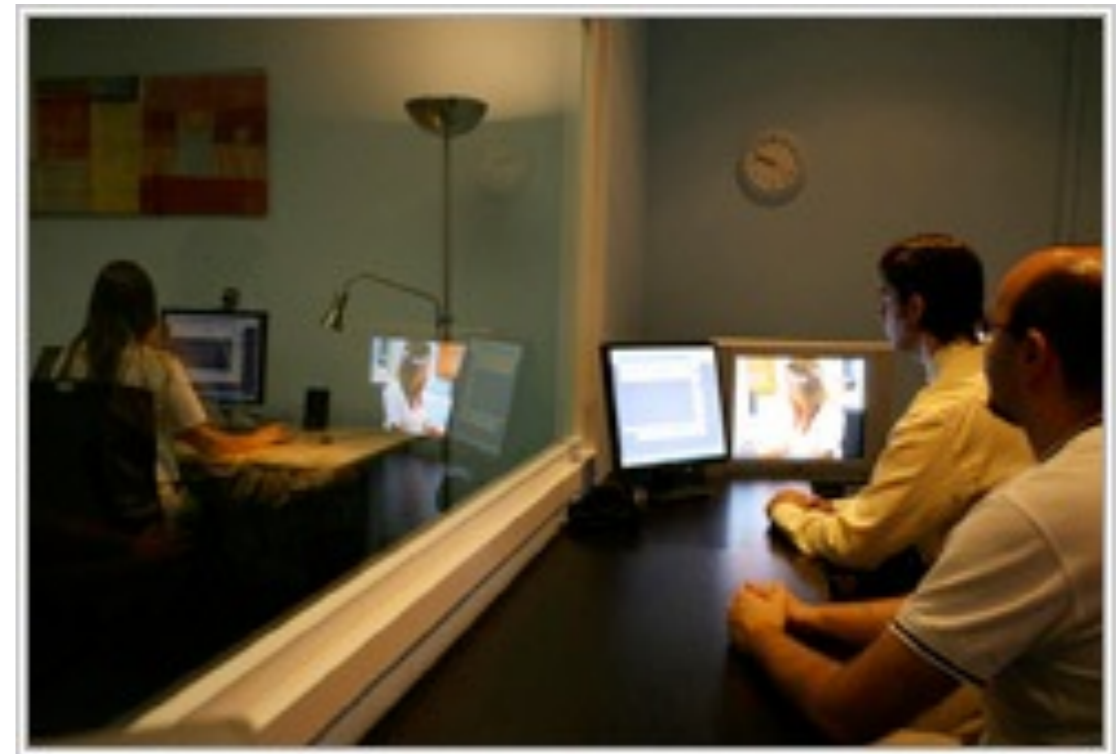
Hypothesen und Signifikanz

- H: Übungsteilnehmer erzielen bessere Noten in der Klausur.
- H_0 : Übungsteilnehmer und Nichtteilnehmer erzielen im Mittelwert die gleichen Noten in der Klausur.
- Effektgröße = Unterschied der Mittelwerte (vorher unbekannt)
- Problem: Effektgröße nicht vorhersagbar, daher H schwer exakter zu formulieren
- Trick: statt H zu beweisen H_0 widerlegen, damit ist automatisch H bewiesen, unabh. von der Effektgröße.

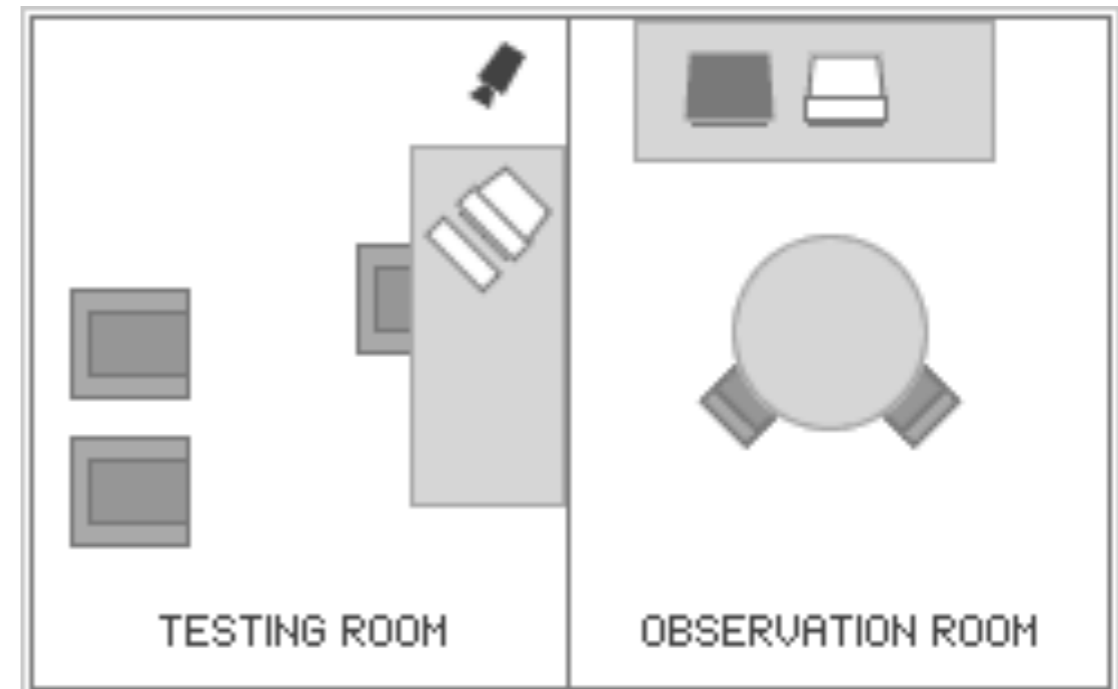
Signifikanztests (Bsp. t-Test)

- Eingabe: 2 Datenreihen
- Ausgabe: Wahrscheinlichkeitswert p zw. 0 und 1
 - Wahrscheinlichkeit, dass beide Reihen in Wirklichkeit den gleichen Mittelwert haben
- Signifikanzniveau z.B. 0,05 (= 5%) oder auch 0,01, 0,001
- Falls $p < 0,05$ spricht man von einem signifikanten Unterschied zwischen den Datenreihen.
- verschiedene Tests für versch. Exp. Designs

Feldstudie vs. Laborstudie



- Externe Validität
- Interne Validität
- Ekologische Validität
- Aufwand



Source: www.xperienceconsulting.com

Langzeit- und Tagebuchstudien



http://www.hcii.cmu.edu/M-HCI/2011/BOA-PlanningTools/images/diary_study.jpg