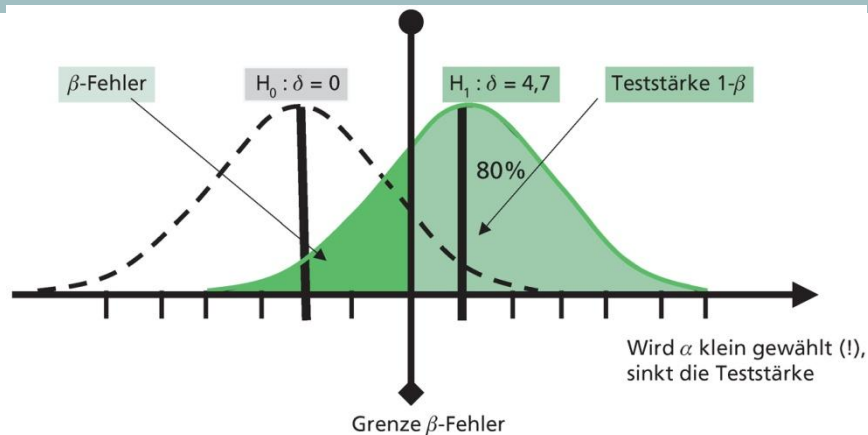


Veranstaltung im Rahmen der Ringvorlesung

(Quasi)-experimentelle Forschungsdesigns in der empirischen Unterrichtsforschung



Prof. Dr. Anne Levin

(Quasi)-experimentelle Forschungsdesigns in der empirischen Unterrichtsforschung

Themen der Vorlesung

- Experimentelle Designs als Möglichkeit empirischer Unterrichtsforschung
- Bedeutung der Gütekriterien beim Aufbau eines experimentellen Designs
- Herausforderungen bei der Konzeption und Durchführung quasi-experimenteller Forschungsdesigns im Unterricht

Das Experiment als Möglichkeit der Hypothesenprüfung - Auswahlkriterien

Voraussetzung für die Wahl der Untersuchungsmethode „Experiment“:

- Auf Grundlage der Literaturrecherche bzw. der Forschungsfrage, lassen sich begründete Hypothesen formulieren.
- Diese Hypothesen können spezifiziert werden:

„Während eine unspezifische Hypothese nur behauptet, dass ein irgendwie geartete Effekt vorliegt und allenfalls noch die Richtung des Effekts angibt, konkretisiert eine spezifische Hypothese auch den Betrag des Effekts bzw. die Effektgröße.“ (Bortz & Döring, 2006, S. 52).

Beispiel:

Sie erwarten, dass der Einsatz der Gruppenarbeitstechnik „Gruppenpuzzle“ (Treatment A) bei Ihren Schülern zu einer tieferen Verarbeitung von dargebotenen Lerninhalten führt als die individuelle Bearbeitung der Lerninhalte (Treatment B).

Unspezifische Hypothese: Die Schüler der Treatmentgruppe A und die Schüler der Treatmentgruppe B unterscheiden sich hinsichtlich ihres erreichten Lernstandes.

Spezifische Hypothese: Fragen, die auf eine Verknüpfung der Lerninhalte abzielen, werden in Gruppe A signifikant besser gelöst als in Gruppe B.

Gültigkeit

Zentral für die Auswahl der Untersuchungsmethode ist die Frage, welche Aussagekraft oder welche Gültigkeit die Untersuchungsergebnisse haben:

Interne Validität:

Sind die Ergebnisse eindeutig kausal interpretierbar, dann ist die Untersuchung intern valide.

Externe Validität:

Sind die Ergebnisse generalisierbar (z.B. auf andere Personen, Zeitpunkte und Situationen), dann ist die Untersuchung extern valide.

Beispiel:

Es gelingt Ihnen, alternative Erklärungen (Motivation, Vorwissen, Lernzeit, genutzte Materialien) für das Auftreten von Unterschieden bezüglich des Lernstands zu minimieren

→Hohe interne Validität

Das führt dazu, dass die Untersuchungssituation sehr „spezifisch“ gestaltet ist

→Geringe externe Validität, da kaum generalisierbar.

Faktoren, die interne Validität gefährden

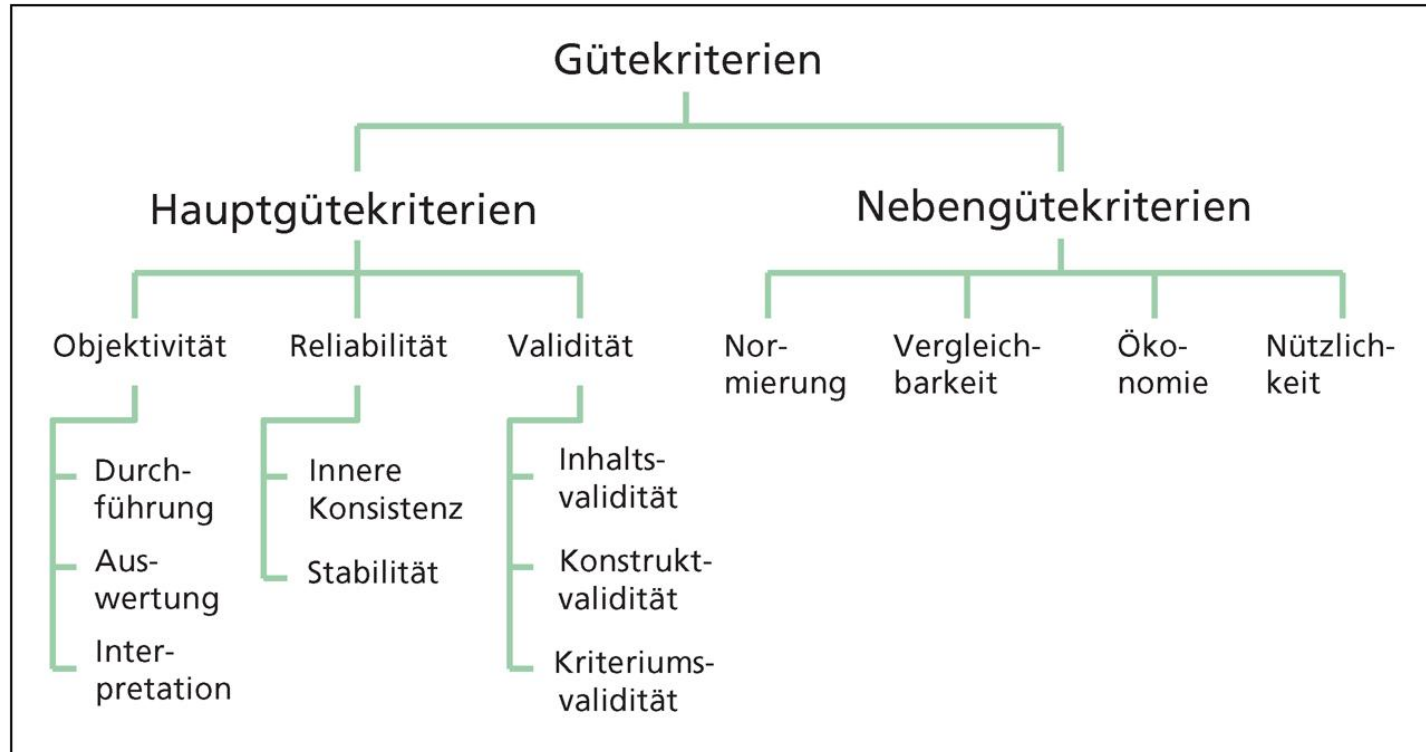
Interne Validität:

- zeitlicher Wandel: andere Einflussgrößen verändern sich im Laufe der Untersuchung
- Reifungsprozesse
- Übung in den Verfahren
- Mangelnde Reliabilität der Messinstrumente
- Selektionseffekte
- Experimentelle Mortalität
- Treatmentdiffusion
- Kompensatorischer Ausgleich
- Kompensatorischer Wettstreit

Faktoren, die externe Validität gefährden

Externe Validität:

- Stichprobenfehler
- Experimentelle Reaktivität (z.B. Wunsch nach Selbstdarstellung, Angst)
- Hawthorne-Effekte (Bewusstsein an einer Untersuchung teilzunehmen)
- Mangelnde instrumentelle Validität (Instrument erfasst nicht, was es erfassen soll)



Zusammenfassung der Haupt- und Nebengütekriterien,
© Pearson Studium 2006

Objektivität: Unabhängigkeit der Ergebnisse eines Tests vom Untersucher

- Durchführungsobjektivität
- Auswertungsobjektivität
- Interpretationsobjektivität

Reliabilität: Grad der Genauigkeit, mit dem der Test das Merkmal misst (unabhängig davon, ob der Test beansprucht, das Merkmal auch wirklich zu messen! → vergl. Validität)

- Retest Reliabilität
- Paralleltestreliabilität
- Interne Konsistenz

Validität: Inwieweit misst der Test das, was er zu messen beansprucht?

- Inhaltsvalidität
- Kriteriumsvalidität
- Konstruktvalidität

Inhaltsvalidität: Bildet das Testinstrument das zu messende Merkmal hinreichend ab?

(→ Augenscheinvalidität):

Überlegen Sie:

Wie inhaltssvalide sind Aufsätze, wenn man die sprachliche Ausdrucksfähigkeit von Schülern erfassen will?

Kriteriumsvalidität: Zusammenhang der Testleistung mit anderen Kriterien

- Vorhersagevalidität (Kann Test ein Merkmal, dass zu einem späteren Zeitpunkt auftritt, vorhersagen?)
- Übereinstimmungsvalidität (Korrelation mit gleichzeitig erhobenen Kriterien z.B. Prüfungsangst (Fragebogen) und physische Merkmale von Angst in einer angstausslösenden Situation)
- Differenzielle Validität (kann der Test in Abhängigkeit von der Gruppe (z.B. Mädchen vs. Jungen) das Merkmal besser oder schlechter vorhersagen?)

Konstruktvalidität: wird das erfasst, was erfasst werden soll?

- Konvergente Validität (Korreliert der Test mit gut geprüften Instrumenten, die gleiches erfassen wollen?)
- Diskriminante Validität (Misst ein Rechentest tatsächlich Rechenfertigkeit oder eher Leseverständnis?)

Gründe für mangelnde Validität

- Mangelnde Reliabilität
- Streuungsrestriktion → das Merkmal streut in der untersuchten Population nur gering (mindert Korrelationen)
- Mangelnde Symmetrie zwischen Kriterium und Prädiktor
- Kriterium ist kontaminiert (gewähltes Kriterium wird durch andere Faktoren beeinflusst, z.B. Berufserfolg durch erzielten Umsatz messen) oder defizient (weil es wichtige Aspekte ausklammert, z.B. Qualität von Pflegeeinrichtungen anhand der Hygiene aber ohne Zufriedenheit der Patienten)
- Methodenfaktoren (Unterschiedliche Methoden der Erhebung Kriterium/Prädiktor → z.B. einmal Interview einmal Test)

Kontrolltechniken von Störvariablen

- Randomisierung → zufällige Zuweisung von Versuchsteilnehmern zu den Experimentalgruppen
- Ausschalten → Gewährleistung eines störungsfreien Ablaufs
- Konstanthalten → durch Stichprobenwahl (z.B. alle Teilnehmer haben das gleiche Alter) → durch Gestaltung der Versuchsumgebung
- Registrieren → unerwartete Einflüsse registrieren;
Variablen, die als Einflussgrößen bekannt sind, aufnehmen

Forschungsethik

Empirischer Erkenntnisgewinn ist in aller Regel daran gebunden, dass Probanden gefunden werden, die an den jeweiligen Untersuchungen teilnehmen.

Aus dieser Situation kann sich eine Reihe von ethischen Problemen ergeben:

- die Verletzung der psychischen und/oder physischen Unversehrtheit und Integrität,
- die fehlende Transparenz der Untersuchungssituation,
- das Vorhandensein von Täuschungen,
- die unfreiwillige Untersuchungsteilnahme sowie
- die mangelnde Vertraulichkeit der Untersuchungsergebnisse.

Hypothesen

Eine Forschungshypothese sollte aus einer Theorie mindestens aber einer gut begründeten Überzeugung) abgeleitet werden

Zusammenhangshypothesen postulieren Zusammenhänge zwischen zwei Merkmalen

Beispiel: Die Verbalisierungsfähigkeit von Schülern korreliert positiv mit dem Urteil der Lehrer über deren Intelligenz

Unterschiedshypothesen postulieren Unterschiede oder Effekte einer Maßnahme (Treatment) auf eine abhängige Variable

Beispiel: Gruppenarbeit führt zu größeren Lerneffekten als Individualarbeit

Veränderungshypothesen postulieren eine verändernde Wirkung in Folge eines Treatments

Beispiel: Ängstliche Schüler profitieren von strukturierten Unterrichtsmaterialien stärker als wenig ängstliche Schüler

Einzelfallhypothesen postulieren den Zusammenhang zwischen Ereignissen bzw. Effekte von Treatments bezogen auf Individuen

Beispiel: Das Motivationstraining für Schüler L. reduziert dessen motorische Unruhe im Unterricht.

Hypothesen

„Eine statistische Hypothese wird stets als statistisches Hypothesenpaar, bestehend aus Nullhypothese (H_0) und Alternativhypothese (H_1) formuliert. Die Alternativhypothese postuliert dabei einen bestimmten Effekt, den die Nullhypothese negiert.“

(Bortz & Döring, 2006, S.24-25)

Ein Hypothesenpaar muss immer alle möglichen Ausgänge der Untersuchung abdecken können.

Beispiel:

Der Mittelwert des am Ende der Untersuchung erhobenen Lernstands ist unter experimenteller Bedingung kleiner oder gleich dem in der Kontrollgruppe erzielten Mittelwert (H_0).

Der Mittelwert des am Ende der Untersuchung erhobenen Lernstands ist unter experimenteller Bedingung größer als der in der Kontrollgruppe erzielte Mittelwert (H_1).

Untersuchungsplan

Thema: Effekte der Gruppenarbeitstechnik JIGSAW auf den Lernerfolg.

Vorgehensweise:

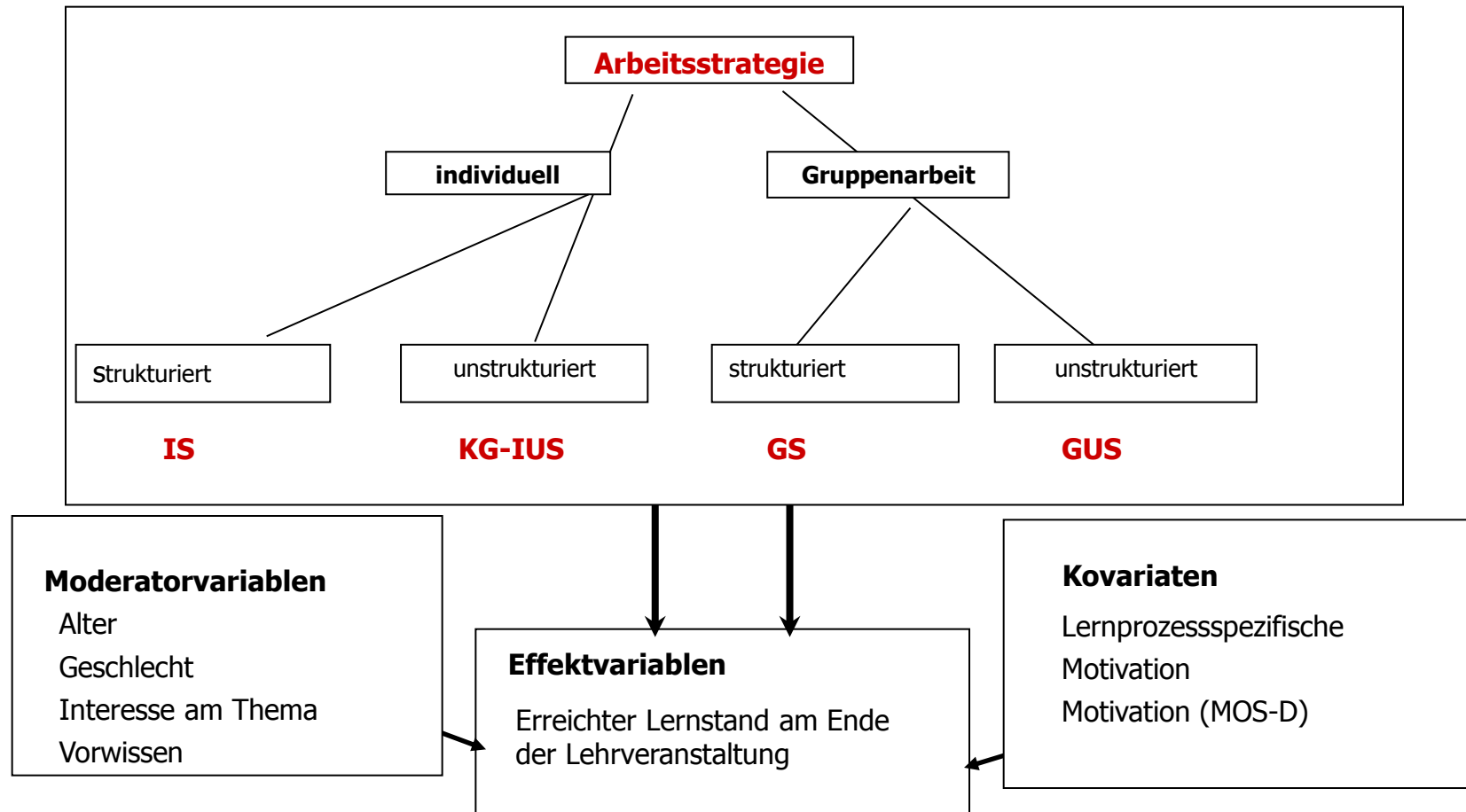
- Spezifikation: Zweigruppendedesign? (Kontrollgruppe vs. Experimentalgruppe) oder Testung verschiedener Treatments?
- (z.B. Individuelles Lernen vs. Gruppenarbeitstechnik 1 vs. Gruppenarbeitstechnik 2)

Auswahl der Stichprobe und Planung des Vorgehens:

- Werden die Treatments innerhalb von Klassen zufällig zugewiesen?
- In welchem Unterricht, über welchen Zeitraum wird das Treatment durchgeführt?
- Wer leitet an?
- Welche Variablen werden wann erhoben? (z.B. Baselinemessung Vorwissen, Erhebung möglicher Einflussfaktoren: Interesse, Geschlecht etc.)

Planung der Durchführung

– Beispiel für ein mögliches Design



Auswahl der Instrumente

- Soziodemographischer Fragebogen
- Erstellung eines Tests zur Abfrage des Vorwissens
- Motivation → spezifisch, unspezifisch
- Konzeption der Lernstandsmessung am Ende → vorzugsweise unterschiedliche Niveaustufen

Konzeption der Treatments

- Inhalte festlegen
- Struktur erstellen
- Ablaufplan (mit genauen Arbeitszeiten: Achtung Parallelisieren sehr wichtig!) erstellen
- Standardisierung der Anweisungen
- Pretest der Materialien!!!

Berücksichtigung von Versuchsleitereffekten

Unerwünschte Effekte, die das Ergebnis verfälschen:

- Nonverbale Signale
- Räumliche Nähe
- Sympathie
- Überzeugungskraft

Generell zu beachten

- Auswertung zunächst planen
- Anonymisierung der Daten
- Gewährleistung von Auswertungsobjektivität
- Trennung von Deskription und hypothesenprüfenden Verfahren
- Richtlinien zur Berichterstellung berücksichtigen

Literatur:

Bortz, J. & Döring, N. (2006). Forschungsmethoden und Evaluation. Springer

Bühner, M. (2006). Einführung in die Test- und Fragebogenkonstruktion. Pearson

Kleining, G. (1995). Lehrbuch Entdeckende Sozialforschung. Band 1. Beltz

Rost, D.H. (2005). Interpretation und Bewertung pädagogisch-psychologischer Studien. Beltz

Rost, J. (2004). Testtheorie - Testkonstruktion. Verlag Hans Huber