

# Evaluationsdesigns

Diskussionsanstoß zum Doktorandenzentrum,  
25.Juni 2010

## Evaluation – warum?

- **Sonderpädagogik ist eine Handlungswissenschaft**
- **Handlungen zielen auf eine Wirkung**
  - Wirkungen treten nicht ein (keine Effekte)
  - Wirkungen können den Zielsetzungen entsprechen (Effekte)
  - Wirkungen können den Zielsetzungen widersprechen (negative Effekte)
  - Wirkungen können ganz andere Wirkungen haben
- **Aktuelles Beispiel (NOZ 25.06.2010)**
  - 9,5 Mio Deutsche trinken zuviel – trotz aller Präventionsmaßnahmen in Schulen etc. = die Präventionsmaßnahmen waren unwirksam (Kommentar)
- **Konklusion: Evaluationsforschung ist notwendig!**
  - Ist die Handlung wirksam?
  - Wirkt die Handlung in die gewünschte Zielrichtung?
  - Gibt es unerwünschte, negative Nebeneffekte?
  - Gibt es wirksamere Maßnahmen?
  - Wirkt die Maßnahme auch in der Praxis in der gewünschten Zielrichtung?

## Evaluation - Begriff

- E.: lat., meint urspr. Bewertung
- Wissenschaftlicher Begriff: fasst eine heterogene Spanne von Tätigkeiten zusammen, aber kaum Bewertung
- Wird oft verstanden als Wirksamkeitsforschung
- Hinweis: man muss konkret sagen, welche Forschung man betreibt!
- Lösungsansatz der Europäischen Kommission 1994: DZME (Daten-Zeit-Methodologie-Evaluator)-Klassifikation der Evaluationsforschung (Uhl 1998) -> Kategorien der Evaluation

## DZME-Klassifikation

- die Datendimension (D)  
Strukturdaten, Prozessdaten, erwartete Ergebnisdaten, unerwartete Ergebnisdaten, Kontextdaten
- die Zeitdimension (Z)  
Konzeptphase = präformative Phase, Entwicklungsphase = formative Phase, Überprüfungsphase = erste summative Phase, Routinephase = zweite summative Phase
- die methodologische Dimension (M)  
Deskriptiv, explorativ, hypothesenprüfend
- die Evaluatordimension (E)  
interne Evaluation, externe Evaluation

## Zeit: Phasen der Evaluationsforschung

1. **Erkundungsphase:** Problemwahl, umfangreiche Aufarbeitung des Kenntnisstands/ Literatur, insbes. Metaanalysen
2. **Theoretische Phase:** Formulierung einer allgemeinen Theorie mit theoretischer Überprüfung, Entwicklung von Ziel-/ Bewährungskriterien, Hypothesenformulierung
3. **Planungsphase:** Untersuchungsplanung, Operationalisierung  
->**Designs!**
4. **Untersuchungsphase:** Datenerhebung
5. **Auswertungsphase:** testtheoretische Datenbewertung, -verarbeitung, Signifikanzniveau festlegen, Signifikanztest, Irrtumswahrscheinlichkeit
6. **Entscheidungsphase:** Signifikanz?, Bewährungskriterium überschritten, Theorie bestätigt,
7. **Überprüfung in Praxis** (evidence studies; Greenberg)

## Planung von Evaluationsstudien - Problem

- Problem der Evaluationsstudien im Feld: hohe Anfälligkeit für Störfaktoren! -> interne Validität (= eindeutige Rückführung der gemessenen Effekte auf die Intervention) ist gefährdet!
- Interne Validität muss also durch das Design in einem möglichst hohen Maß gewährleistet werden
- Prinzipiell ist aber die Ausschaltung aller Störfaktoren in Feldstudien nicht möglich! (nur in Laborexperimenten)
- Zugleich: sehr hohe interne Validität (z.B. im Labor) kann eine sehr geringe externe Validität (Übertragbarkeit, Generalisierbarkeit) zur Folge haben

## Planung von Evaluationsstudien - Lösung

- Lösung ist die Randomisierung: zufällige Zuordnung der Untersuchungsteilnehmer zu den Untersuchungsbedingungen/ alle Pbn haben zu Beginn der Untersuchung die gleichen Chancen in jede der Untersuchungsbedingungen zu gelangen
- Randomisierung gleicht die Störfaktoren in allen Untersuchungsbedingungen (Treatment – Kontrollgruppe) aus, indem sie in allen Untersuchungsbedingungen in gleichem Maß verteilt sind -> erhöht die interne Validität
- Bedingung: eine gewisse, möglichst hohe Anzahl von Pbn – damit der Zufall richtig wirken kann (Rost)
- Untersuchungen mit Randomisierung sind Experimente – bei Verzicht auf Randomisierung sind es quasi-experimentelle Designs

## Designs - Elemente

- Elemente der Versuchsplanung sind also
  - O = Erhebung/ Beobachtung/ Test/ Messung
  - X = Behandlung, Treatment
  - = keine Behandlung
  - R = Randomisierung
  - X' = unspezifische Alternativbehandlung
  - EG = Experimentalgruppe (bei Randomisierung)
  - KG = Kontrollgruppe (bei Randomisierung)
  - VSG = Versuchsgruppe (bei fehlender R)
  - VGG = Vergleichsgruppe (bei fehlender R)
  - (nach Rost 2007, 117)

CARL VON OSSIETZKY universität OLDENBURG

## Designs - Überblick

Gruppe	Randomisierung	Designs
Einmalige Erhebung bei einer Gruppe	-	<b>Deskriptives</b> Design
Eine behandelte Gruppe, mehrfach gemessen	-	„vor“- <b>experimentelles</b> Design
2 unterschiedlich behandelte Gruppen	Ohne Randomisierung	„quasi“- <b>experimentelles</b> Design (Versuchs-/ Vergleichsgruppe)
<b>Starke Designs</b>		
2 unterschiedlich behandelte Gruppen	Mit Randomisierung	<b>Experimentelles</b> Design mit Experimental- und Kontrollgruppe

S  
T  
Ä  
R  
K  
E

CARL VON OSSIETZKY universität OLDENBURG

## Deskriptives Design

- (narrative) Einzelfallstudien: illustrieren, regen zur Ideenbildung (Rost: bei einem Glas Wein zu lesen!), Hinweise für mögliche Hypothesen
- Aber: Keine Prüfung von Hypothesen möglich! = „Mit empirischer Forschung und mit Wissenschaft hat das nichts zu tun.“ (Rost 2007, 119)

## Vor-experimentelle Designs

O<sub>1</sub> O<sub>2</sub> O<sub>3</sub> X O<sub>3</sub> O<sub>4</sub> O<sub>5</sub>

- Zeitreihen- „Experiment“  
(wiederholter) Vortest – Behandlung – (wiederholter) Nachtest
- Interne und externe Validität ist nicht gesichert!
- Besser: Multiple Einzelfallstudien mit ABAB-Designs  
(Julius, Schlosser & Goetze 2000)

## Vor-experimentelle Designs

X O

- Ein-Gruppen-Plan mit Behandlung und Nachtest
- Eigentlich wertlos: Nichts kann wirklich kontrolliert werden
- Bsp.: Stressbewältigungsprogramm – Messung mit Fragebogen: nach Intervention sind die Probanden im Normalbereich = Schluss auf Wirksamkeit des Programms ist unzulässig!

## Vor-experimentelle Designs

O X O

- Ein-Gruppen-Plan mit Vortest - Behandlung -Nachtest
- Leicht besser: Wenigsten können die Veränderungen in der Gruppe dokumentiert werden, aber ist nicht auf die Intervention zurück zu führen (Entwicklungsprozesse? Andere Einflüsse?)

## Quasi-experimentelle Designs

VSG: X O  
VGG: - (oder X') O

- Nichrandomisierter 2-Gruppen-Plan mit Behandlung und Nachtest
- Liegen gemessene Unterschiede an der Auswahl der/ Zuweisung zur jeweiligen Gruppe?

## Quasi-experimentelle Designs

<b>VSG:</b>	<b>O</b>	<b>X</b>	<b>O</b>
<b>VGG:</b>	<b>O</b>	<b>- (oder X')</b>	<b>O</b>

- Nichrandomisierter 2-Gruppen-Plan mit Vortest - Behandlung - Nachtest
- Besser: immerhin könnten Vortestunterschiede erfasst werden
- Liegen gemessene Unterschiede an der Auswahl der/ Zuweisung zur jeweiligen Gruppe?
- Hier liegt entscheidende Klippe zu starken Designs: Randomisierung fehlt! Nicht verwechseln mit experimentellem 2-Gruppen-Design

## Experimentelle Designs

<b>R</b>	<b>EG: O</b>	<b>X</b>	<b>O</b>
<b>R</b>	<b>KG: O</b>	<b>- (oder X')</b>	<b>O</b>

- Randomisierter 2-Gruppen-Plan mit Vortest – Behandlung – Nachtest
- Standardversuchsplan der Evaluationsforschung
- Erlaubt kausale Aussagen
- Hinweis: pro Auswertungszelle sollten mindestens 35 – 40 Pbn enthalten sein!
- Und: Bei geringeren Effekten höhere Zahlen



## Experimentelle Designs

R	EG <sub>1</sub> :	OX <sub>1</sub>	O	
R	EG <sub>2</sub> :	O	X <sub>2</sub>	O
R	KG:	O	- (oder X')	O

- Randomisierter 3-Gruppen-Plan mit Vortest – Behandlung – Nachtest
- Vorteil: erlaubt den Vergleich 2 verschiedener Interventionen mit der KG und der anderen EG = komparative Evaluation
- Erlaubt kausale Aussagen und Einschätzungen, welche Maßnahmen wirksamer als andere sind
- Hinweis: pro Auswertungszelle sollten mindestens 35 – 40 Pbn enthalten sein-> also hier mehr PBN nötig!

## Experimentelle Designs

R	EG:	X	O
R	KG:	- (oder X')	O

- Randomisierter 2-Gruppen-Plan ohne Vortests – Behandlung – Nachtest
- Variante zum klassischen Standardversuchsplan der Evaluationsforschung
- Vorteile
  - Ökonomische Variante: man spart sich den Prätest!
  - Kontrolliert die Wechselwirkung „Testung x Behandlung“
- Anmerkung C.H.: Hohes Risiko in päd.Kontexten!
  - verklumpte Stichproben

Ausreichende Zahl der Pbn

## Experimentelle Designs

R	EG <sub>1</sub> :	OX <sub>1</sub>	O	
R	KG <sub>1</sub> :	O	-	O
R	EG <sub>2</sub> :		X <sub>1</sub>	O
R	KG <sub>2</sub> :		-	O

- Solomon-4-Gruppen-Plan: 1949 von Solomo vorgeschlagen, aufwändig, daher selten genutzt
- Vorteile:
  - Sensibilisierungseffekt durch Prä-Test wird messbar (EG1 – EG2)
  - Testwiederholungseffekt wird messbar (KG1 – KG2)
  - Generalisierbarkeit durch Vergleich Diff (EG1-KG1) zu Diff (EG2-KG2)
- Hinweis zum Aufwand: selbst bei großen zu erwartenden Effektstärken sind mind. 160 Pbn notwendig

## Problem Klumpenstichprobe

- Beobachtung
  - Evaluationsstudien im (sonder-) pädagogischen Feld greifen immer auf bestehende Gruppen zurück
  - Bsp.: KigaGruppen, Schulklassen, Klienten für ein bestimmtes Angebot, Jugendgruppen, Insassen einer JJVA
- Ergibt das Problem „Verklumpung der Stichprobe“: Gruppen zeigen systematische Differenzen = Klumpenstichpoben
- Lösungsvorschläge in der wiss.Diskussion:
  - Durchsetzung der Randomisierung auf Individuumsebene  
→ aber wie? Und: oft sollen Maßnahmen in Gruppe wirken!
  - anspruchsvollere Analysen (Mehrebenenanalysen, erfordern aber hohe statistische Qualität der Daten) oder
  - Erhöhung der Stichprobengröße

## Resumée

- Evaluationsstudien sind eine zentrale Aufgabe der sonderpädagogischen Forschung.
- Qualitativ hochwertige Studien brauchen nicht unbedingt hohe Probandenzahlen, aber
- eine gute Untersuchungsplanung!
- Übrigens: Es fehlt insbesondere an
  - Replikationsstudien und
  - Effectiveness-Studien (Übertrag ins Feld)
- Dies aber ist für die Anerkennung von Intervention als „evidenzbasierte Maßnahmen“ (APA) entscheidendes Kriterium!