

## 7 Designfragen: Planung von Evaluationsuntersuchungen

### Übungsaufgaben

- (1) Frau Dr. A. Wohlgemuth, Geschäftsführerin einer IT-Firma, bietet seit neuestem für ihre Führungskräfte bis zu 20 kostenlose Coachingstunden zu Führungs- und Managementthemen bei externen Coaches an. Frau Dr. Wohlgemuth verfügt über Daten zur Einschätzung der Führungsqualität der Führungskräfte sowohl aus der Zeit vor der Einführung der Coachings als auch aus der Zeit nach der Einführung der Coachings. Um zu entscheiden, ob sie diese Coachings weiter finanzieren soll, vergleicht sie die Führungsqualität ihrer Führungskräfte nach, im Vergleich zu vor, der Einführung der Coachingangebote.
- (a) Um welches Untersuchungsdesign handelt es sich bei dem realisierten Vorgehen? Benennen Sie das Design und zeichnen Sie es formal auf.

#### Musterantwort

*Es handelt sich um ein Eingruppen-Pre-Test-Post-Test Design:*

$O_1 A_1 O_2$

*Dabei beschreibt A1 das Coaching, O1 die Messung der Führungsqualität vor dem Coaching und O2 die Messung der Führungsqualität nach dem Coaching.*

- (b) Welche Gefahren können die interne Validität der geplanten Studie bedrohen?

#### Musterantwort

*Die interne Validität der Studie kann durch Veränderungen über die Zeit bedroht werden, die nichts mit der Einführung der Coachingstunden zu tun haben. Dazu gehören externe zeitliche Einflüsse, Reifung, Testübung, Veränderungen des Messinstruments, Regression zur Mitte und selektiver Drop-out:*

- *Externe zeitliche Einflüsse: Z.B. könnte es sein, dass sich die Führungskräfte in der Zeit zwischen den beiden Messungen der Führungsqualität auch unabhängig von den Coachingstunden zu Führungs- und Managementthemen weitergebildet haben, beispielsweise durch Literatur oder Online-Ressourcen.*
- *Reifung: Es könnte sein, dass sich die Führungsqualität der Führungskräfte über die Zeit aufgrund zunehmender Führungserfahrung verbessert.*
- *Testübung: Die beurteilenden Personen könnten sich bei der zweiten Einschätzung der Führungsqualität an ihre erste Einschätzung erinnern und sich daran orientieren.*
- *Veränderung im Messinstrument: Es könnte sein, dass die beurteilenden Personen ihre Anforderungen für eine gute Bewertung erhöht haben, so dass auch bei gleicher Fähigkeit hinterher eine niedrigere Führungsqualität als vorher eingeschätzt würde.*
- *Regression zur Mitte: Es kann sein, dass die Werte der Führungsqualität zum zweiten Messzeitpunkt allein aufgrund ihrer statistischen Auftretenswahrscheinlichkeit weniger extrem sind als zum ersten, insbesondere wenn sie zum ersten Messzeitpunkt besonders extrem waren (sehr positiv oder auch sehr negativ).*
- *Selektiver Drop-out: Es könnte sein, dass zwischen den beiden Messzeitpunkten Führungskräfte mit einer besonders hohen Führungsqualität von einem anderen Unternehmen abgeworben wurden und darum zur zweiten Erhebung nicht mehr Teil des Unternehmens waren. Dies könnte dazu geführt haben, dass die durchschnittliche Führungsqualität zum zweiten Messzeitpunkt schlechter ausfiel als zum ersten, allein durch das Fehlen der besonders guten Führungskräfte. Dadurch, dass auch Werte vom Vortest vorliegen, könnte ein solcher Drop-out jedoch teilweise statistisch kontrolliert werden.*

- (c) Welche der Gefahren könnten durch eine mehrfache Messung der Führungsqualität vor und nach dem Coaching kontrolliert werden?

### **Musterantwort**

*Wenn die oben genannten Gefahren für die interne Validität aufgrund der Veränderung vor (und nach) dem Coaching für die Zeitspanne während des Coachings vorhergesagt werden können (z.B. durch eine Funktion), kann durch die zusätzlichen Messungen geschätzt werden, welchen Einfluss diese Gefahren auf die gemessenen Veränderungen der Führungsqualität haben. Somit kann der Einfluss dieser Gefahren auch statistisch kontrolliert („herausgerechnet“) werden.*

*Bei externen zeitlichen Einflüssen ist es nicht unbedingt plausibel, dass externe zeitliche Einflüsse während des Coachings aufgrund der Veränderung vor (und nach) den Coachings vorhersagbar sind. Externe zeitliche Einflüsse können somit nicht notwendigerweise durch zusätzliche Messungen kontrolliert werden.*

- (2) Prof. Dr. J. Reichelt möchte untersuchen, inwiefern sich der Erfolg in der Reduktion von sozialen Ängsten zwischen einer Verhaltenstherapie und einer Psychoanalyse unterscheidet. Für seine Untersuchung plant Prof. Reichelt alle an der Studie Interessierten randomisiert entweder einer Verhaltenstherapie oder einer Psychoanalyse zuzuweisen. Er misst die Ausprägung sozialer Ängste bei allen Studienteilnehmenden nach Beendigung der Therapie.
- (a) Um welches Untersuchungsdesign handelt es sich in diesem Beispiel? Benennen Sie es und zeichnen Sie es formal auf.

### **Musterantwort**

*Es handelt sich um ein Nur-Post-Test-Kontrollgruppen-Design:*

*R A<sub>1</sub> O<sub>2</sub>*

*R A<sub>2</sub> O<sub>2</sub>*

*Dabei bezeichnet A1 die Verhaltenstherapie, A2 die Psychoanalyse, O2 die Messung der sozialen Angst und R die randomisierte Zuweisung zu den Gruppen.*

- (b) Welche Aspekte dieser Studie lassen darauf schließen, dass es sich um eine experimentelle Studie handelt?

### **Musterantwort**

- (1) *Zielsetzung: Prof. Reichelt hat das Ziel, den kausalen Effekt der unabhängigen Variable „Therapieverfahren“ (dichotom, Verhaltenstherapie vs. Psychoanalyse) auf die abhängige Variable „soziale Ängstlichkeit“ zu untersuchen.*
- (2) *Manipulation der UV: Prof. Reichelt manipuliert die unabhängige Variable (Therapieverfahren). D.h. er stellt bewusst die beiden Ausprägungen „Verhaltenstherapie“ und „Psychoanalyse“ her.*
- (3) *Kontrolle von Störvariablen: Durch die Randomisierung kann Prof. Reichelt mögliche Störvariablen vollständig kontrollieren, vorausgesetzt seine Stichproben sind groß genug. Er kann sowohl Gefahren über die Zeit, Selektionseffekte, als auch deren Interaktion mit anderen Gefahren durch die Verwendung des Nur-Post-Test Kontrollgruppen-Designs kontrollieren.*
- (c) Prof. Reichelt kann aufgrund finanzieller Ressourcen nur eine sehr kleine Stichprobe untersuchen und vermutet, dass es dadurch durch Zufall zu Unterschieden zwischen den Interventionsgruppen vor der Therapie kommen kann. Was kann Prof. Reichelt tun, um diesem Problem zu begegnen?

**Musterantwort**

Um die Unterschiede zwischen den Interventionsgruppen vor der Therapie zu erfassen, kann Prof. Reichelt einen zusätzlichen ersten Messzeitpunkt vor Durchführung der Therapien planen, bei dem er plausible personengebundene Störvariablen erhebt. Er realisiert somit ein Pre-Test-Post-Test-Kontrollgruppen-Design. Dabei sollte er mögliche Störvariablen idealerweise im Vorhinein recherchieren. In der Regel ist ein Pretest, in diesem Fall die Messung der sozialen Ängstlichkeit vor der Zuweisung zu den Interventionsbedingungen, meist eine relevante Störvariable, die erhoben werden sollte. Die Störvariablen kann er bereits vor Durchführung der Therapien durch Konstanthaltung, Ausbalancierung oder Matching kontrollieren, oder aber im Anschluss bei der Studiauswertung statistisch kontrollieren.

- (3) Sie wollen den Einfluss der Therapieart (Verhaltenstherapie vs. Psychoanalyse) und parallele medikamentöse Behandlung (Medikamente vs. keine Medikamente) auf die Symptombelastung von Patient:innen nach der Therapie mit einem mehrfaktoriellen Design untersuchen. Sie möchten zudem zusätzlich noch den Faktor Ausbildungshintergrund der Therapeut:innen (Medizin vs. Psychologie) in das Design aufnehmen.
  - (a) Nennen Sie zwei Gründe, warum man hier einen mehrfaktoriellen statt drei einfaktorielle Versuchspläne verwenden sollte.

**Musterantwort**

- (1) Mit drei einfaktoriellen Versuchsplänen kann man lediglich die Haupteffekte der Faktoren ermitteln. Mit einem (vollständigen) mehrfaktoriellen Versuchsplan können (alle) Haupt- und vor allem Interaktionseffekte der drei Faktoren untersucht werden. Beispielsweise könnte es sein, dass die Verhaltenstherapie im Vergleich zur Psychoanalyse besonders gut bei 10 Sitzungen abschneidet, während es bei 50 Sitzung genau anders herum ist. Solche Effekte können nur mit Hilfe von Interaktionen untersucht werden.
- (2) Man kann zudem bei der Untersuchung des Effektes eines Faktors (z.B. Therapieart) auf die Symptombelastung den Einfluss anderer Faktoren (z.B. Ausbildungshintergrund der/des Therapeut:in) kontrollieren. Das ist eine Möglichkeit wie für Selektionseffekte kontrolliert werden kann (Siehe auch Ausbalancierung personengebundener Störvariablen in Abschn. 5.2.2).

- (b) Stellen Sie einen entsprechenden vollständigen mehrfaktoriellen Versuchsplan auf.

**Musterantwort**

Es gibt 2 (Therapieart) \* 2 (medikamentöse Behandlung) \* 2 (Ausbildungshintergrund) = 8 Gruppen.

	Verhaltenstherapie		Psychoanalyse	
	Medizin	Psychologie	Medizin	Psychologie
Medikament				
kein Medikament				

Beispielsweise erhält eine Gruppe (oben links) Verhaltenstherapie mit medikamentöser Behandlung von einer/m Psychotherapeut:in mit medizinischem Ausbildungshintergrund. Eine andere Gruppe (oben rechts) erhält Psychoanalyse mit medikamentöser Behandlung von einer/m Psychotherapeut:in mit psychologischem Ausbildungshintergrund.

(c) Stellen Sie einen sinnvollen hierarchischen Versuchsplan für dieses Beispiel auf.

**Musterantwort**

	<b>Medizin</b>	<b>Psychologie</b>
	Medikament	kein Medikament
Verhaltenstherapie		
Psychoanalyse		

*In diesem Beispiel ist die medikamentöse Behandlung unter den Ausbildungshintergrund geschachtelt, da nur Mediziner:innen Medikamente verschreiben dürfen. In dem abgebildeten Versuchsplan wird eine medikamentöse Behandlung nur bei medizinischen Psychotherapeut:innen angeboten und eine nichtmedikamentöse Behandlung nur bei psychologischen Psychotherapeut:innen. Es gibt insgesamt vier Untersuchungsbedingungen.*

(d) Stellen Sie einen quadratischen Versuchsplan für dieses Beispiel auf.

**Musterantwort**

	<b>Verhaltenstherapie</b>	<b>Psychoanalyse</b>
Medikament	Medizin	Psychologie
kein Medikament	Psychologie	Medizin

*Es gibt vier Untersuchungsbedingungen. Jede Faktorstufe eines Faktors ist mit jeder Faktorstufe eines anderen Faktors kombiniert: Beispielsweise wird Verhaltenstherapie sowohl einmal mit als auch ohne medikamentöser Behandlung und sowohl einmal von einer/m Psychotherapeut:in mit medizinischer als auch einmal mit psychologischer Ausbildung angeboten. Es werden jedoch nicht alle Kombinationen aller drei Faktorstufen realisiert; so gibt es beispielsweise keine Gruppe, die Verhaltenstherapie mit medikamentöser Behandlung von einer/m psychologischen Psychotherapeut:in erhält.*

(e) Benennen Sie je einen Vor- und Nachteil eines unvollständigen im Vergleich zu einem vollständigen Versuchsplan.

**Musterantwort**

*Vorteile:*

- *Ein unvollständiger Versuchsplan kann mit weniger Gruppen (und damit weniger Teilnehmenden) umgesetzt werden und ist somit praktikabler.*
- *Ein unvollständiger Versuchsplan kann auch dann umgesetzt werden, wenn einige Kombinationen der Faktoren aus ethischen oder praktischen Gründen nicht möglich sind (z.B. Medikamentenverschreibung von psychologischen Psychotherapeut:innen).*

*Nachteil:*

- *Mit einem unvollständigen Versuchsplan können nicht alle Haupt- und Interaktionseffekte untersucht werden.*