

## Kapitel 3 – Einzelfallforschung

Walter Hussy & Jana F. Bauer

Erst in den letzten Jahrzehnten finden Methoden der **quantitativen Einzelfallforschung** (auch  $N=1$ -Forschung genannt) wieder verstärkt Beachtung. Es handelt sich um die Analyse einer einzelnen Person durch intensives Studium ihres Erlebens und Verhaltens. Dabei hat diese Methodik eine lange Tradition in der psychologischen Forschung. Ausgehend von Wundt (1832–1920), über Ebbinghaus (1850–1909), Pawlow (1849–1936) bis hin zu Skinner (1904–1990) wurde diese Methode in verschiedenen Varianten (z. B. verwendete sich Ebbinghaus selber als Versuchsperson) immer wieder eingesetzt. Dennoch wurde sie – häufig vor dem Hintergrund methodischer Kritik – bis in die 70er Jahre des vergangenen Jahrhunderts stark in den Hintergrund gedrängt und erfuhr erst im Zusammenhang mit der die Einzelfallforschung wieder aufgreifenden **Psychotherapieforschung** neue Würdigung. Auch in der Sonderpädagogik (siehe z. B. Grünke, 2012) und der Medizin (siehe z. B. Vlaeyen et al., 2022) kommt sie zur Anwendung.

Die **Kritik** an der Einzelfallforschung bezog sich zunächst auf ihren **unkontrollierten Charakter**. Diese Kritik resultierte aus den anfangs oft **anekdotischen** Fallstudien und es dauerte einige Jahrzehnte, bis die **streng kontrollierten Formen**, die man als **quasiexperimentell** bezeichnen kann, entwickelt waren.

### Definition

Von quantitativer **Einzelfallforschung** spricht man, wenn eine einzelne Untersuchungseinheit (in der Regel eine Person) bezüglich einer oder mehrerer AV unter möglichst kontrollierten Bedingungen wiederholt beobachtet wird, um die Wirkung einer UV messen und bewerten zu können.

Wie schon angedeutet handelt es sich bei der Untersuchungseinheit meist um eine einzelne Person, aber es sind auch einzelne Gruppen (Familien, Vereine, Unternehmen, Schulen, Dörfer usw.) als Einheiten denkbar. Natürlich kommt vor dem Hintergrund der Forderung nach möglichst großen Stichproben in der **Gruppenforschung** sofort die Frage auf, wie in einem Einzelfallplan denn hinreichende Reliabilität und Validität gewährleistet werden können. Die Antwort gibt der Begriff der **Verhaltensstichprobe**. Die gleiche Person wird hinsichtlich des interessierenden Verhaltens wiederholt beobachtet. In der Einzelfallforschung werden also möglichst viele Verhaltensbeobachtungen an einer Untersuchungseinheit durchgeführt während in der Gruppenforschung angestrebt wird, eine möglichst große Anzahl an Untersuchungseinheiten (Versuchspersonen) zu beobachten (messen, testen usw.).

© Der/die Herausgeber bzw. der/die Autor(en), exklusiv lizenziert an Springer-Verlag GmbH, DE, ein Teil von Springer Nature 2010, 2013, 2023

Aus: Schreier, M., Echterhoff, G., Bauer, J. F., Weydmann, N. & Hussy, W. (2023). *Forschungsmethoden in Psychologie und Sozialwissenschaften für Bachelor* (3. Aufl.). Springer.

### Definition

Unter einer Stichprobe versteht man eine Teilmenge aus einer Grundgesamtheit (Population). Bei der **Personenstichprobe** liegt eine Auswahl aus einer Population von Personen, bei der **Verhaltensstichprobe** eine Auswahl aus einer Population von Verhaltensweisen vor.

Den Begriff der Messwiederholung haben wir bereits im Kontext von Experimenten mit interindividueller Bedingungsvariation, längsschnittlichen Korrelationsstudien oder Panelstudien kennengelernt. In der Einzelfallforschung geht es jedoch nicht um zwei, drei oder vier wiederholte Beobachtungen, sondern es gilt die Forderung nach einer möglichst großen **Zeitreihe** (mit möglichst vielen Messungen in gleichen zeitlich Abständen).

Ausgangspunkt für alle weiteren Überlegungen und Maßnahmen ist die sogenannte **Basisrate** (Grundrate, „baseline“). Überprüft man die Hypothese, dass die verhaltenstherapeutische Maßnahme X (UV) sich positiv auf die Ängstlichkeit (AV) einer Person auswirkt, so wird bei dieser Person zunächst über einen längeren Zeitraum die Ängstlichkeit gemessen, um deren (natürliche) Fluktuation zu bestimmen. Abb. 1 gibt ein Beispiel für eine (fiktive) Basisrate zu diesem Vorgehen.

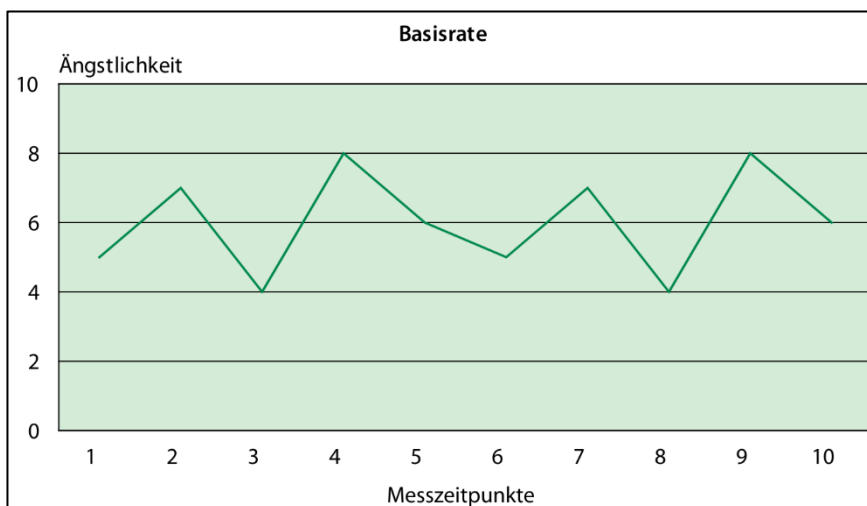


Abb. 1 Beispiel für eine Basisrate aus 10 Messungen zur AV Ängstlichkeit

Abb. 1 zeigt, dass die Ängstlichkeit – fiktiv gemessen auf einer zehnstufigen Skala – in zehn aufeinander folgenden Tagen Werte zwischen 4 und 8 annimmt, also um einen mittleren Wert von etwa 6 Punkten variiert.

## AB-Plan

Die Logik der Einzelfallversuchspläne geht nun dahin, dass eine **deutliche und stabile Niveauperänderung** aufgrund des Einsetzens einer **Intervention** (in unserem Beispiel der Therapie X) an Tag 11 als Effekt dieser Therapie und damit (vorläufig) kausal interpretiert werden kann. Abb. 2 veranschaulicht diese Logik.

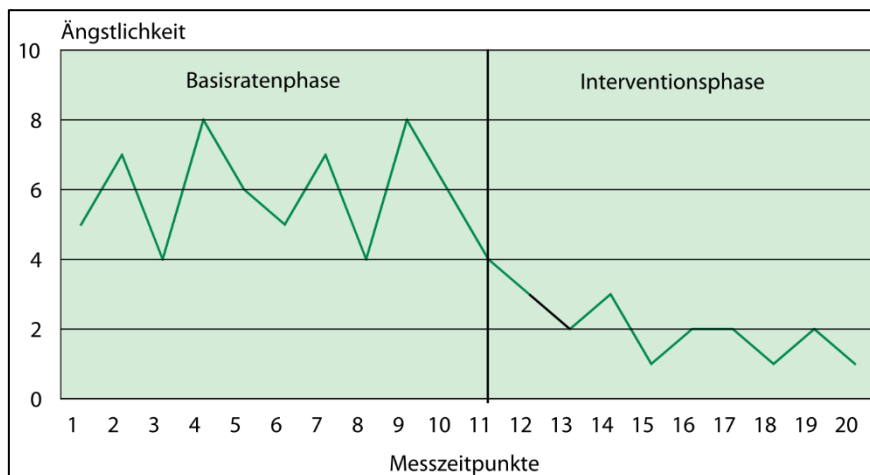


Abb. 2 zeigt zwanzig fiktive Messungen der Ängstlichkeit einer Person, gemessen im Tagesabstand. Die ersten 10 Messungen erfolgten vor und die letzten 10 Messungen nach der Intervention

Mit dieser grundlegenden Vorgehensweise in der quantitativen Einzelfallforschung ist auch bereits der erste Versuchsplan beschrieben. Er nennt sich **AB-Plan**. Die Bezeichnung basiert auf den beiden Untersuchungsphasen. Die Erfassung der Basisrate bildet die A-Phase und die Messungen während des Interventionszeitraums (des Therapiezeitraums) die B-Phase. Aus Abb. 2 ist zu entnehmen, dass mit dem Einsetzen der Therapie die Ausprägung der Ängstlichkeit abnimmt und sich auf einem Niveau zwischen 1 und 2 Skalenpunkten einpendelt.

Diesen Therapieerfolg uneingeschränkt kausal zu interpretieren verbietet sich schon deshalb, weil auch andere Einflüsse (Störeffekte außerhalb der therapeutischen Situation) die Veränderung hervorgerufen haben könnten. Die **interne Validität ist bei diesem Versuchsplan somit eingeschränkt** und die kausale Erfolgsinterpretation vorläufiger Natur. Sie bedarf der Bestätigung durch weitere **Replikationen**, bevor sie als uneingeschränkt kausal formuliert werden kann.

## ABA-Plan

Um diesem Einwand entgegenzutreten wurden weitere Pläne entwickelt, von denen wir hier noch den **ABA- und ABAB-Plan** skizzieren wollen. Beim ABA-Plan wird eine zweite A-Phase an die B-Phase angeschlossen. Es erfolgt also im Anschluss an die Interventions-Phase eine erneute Erfassung der Ängstlichkeit über eine Reihe von Tagen hinweg *ohne* Intervention. Abb. 3 zeigt diesen Fall.

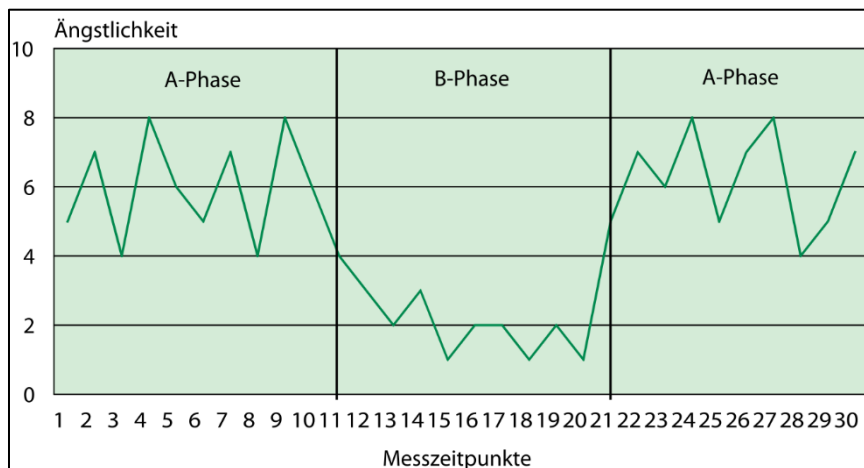


Abb. 3 Die drei Phasen des ABA-Versuchsplans

Dieses Ergebnisbild **stärkt die interne Validität**, weil es relativ unwahrscheinlich ist, dass die äußeren Störeffekte mit dem Ein- und Absetzen der Intervention zusammentreffen, also konfundiert sind. Dennoch ist weiterhin eine Zurückhaltung bei der kausalen Interpretation angezeigt.

## ABAB-Plan

Der ABAB-Plan setzt diese Logik mit einer zweiten B-Phase fort. Bestätigt sich das Ergebnis der ersten Interventionsphase, so **stärkt dieser Sachverhalt die interne Validität** weiter (Abb. 4). Ergibt sich das in der Abbildung dargestellte Ergebnismuster, so kann man (nach entsprechender statistischer Absicherung) von einer nur noch leicht eingeschränkten internen Validität ausgehen. Die Übertragbarkeit auf weitere Personen, Indikationen, Therapeut\*innen usw. ist jedoch noch durch **zusätzliche systematische Replikationen** zu prüfen. Damit wird bei gleichbleibender Befundlage die interne Validität endgültig gesichert und der Geltungs- bzw. Anwendungsbereich der Hypothese bestimmt.

In den besprochenen experimentellen Einzelfallversuchsplänen rückt die untersuchte Person mit ihrem Verhalten in der Vergleichssituation an die Stelle der Kontrollgruppe. Somit fungiert die untersuchte Person sowohl als **Versuchsperson** (B-Phase) als auch ihre **eigene**

Kontrollperson (A-Phase).

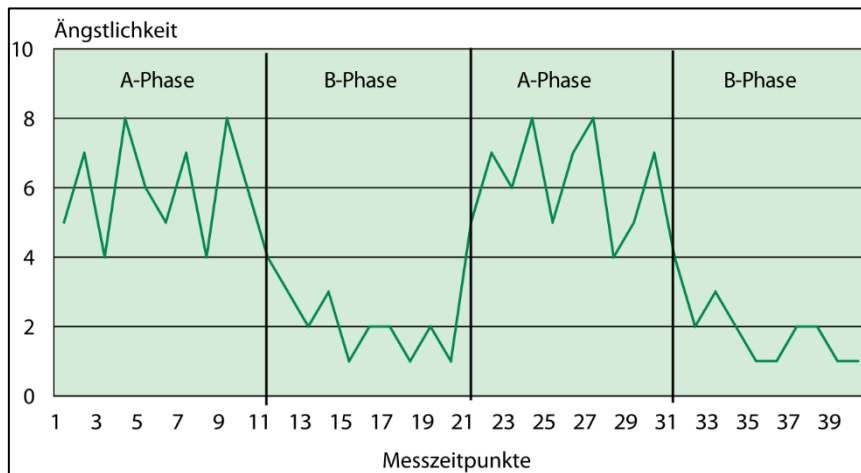


Abb. 4 Die vier Phasen des ABAB-Versuchsplans

Kritik an der quantitativen Einzelfallforschung und den Versuchsplänen wird primär in zwei Punkten geübt:

- die aus Gründen der internen Validität geforderte **Effektzurücknahme** in der zweiten A-Phase sowie
- das Ende der Intervention nach der zweiten A-Phase im ABA-Versuchsplan (**unbehandelter Abbruch**).

**Effektzurücknahme** bedeutet, dass nachzuweisen ist, dass bei ausbleibender Intervention die AV wieder das Basisratenniveau annimmt. Besonders im therapeutischen Kontext ist diese Forderung natürlich kontraindiziert, denn hier wünscht man sich gerade überdauernde Effekte. Der **unbehandelte Abbruch** bezieht sich auf das Ende der Intervention mit der AV auf Basisratenniveau. Auch hier ergibt sich ein klarer Konflikt zum therapeutischen Auftrag, also ein massives ethisches Problem.

Beide Kritikpunkte verlieren an Gewicht, wenn man sich bewusst macht, dass es sich hier um **Forschungspläne** und nicht um **Therapiepläne** handelt. Natürlich wird man im konkreten Fall versuchen, eine therapeutisch wirkungsvolle Maßnahme für die Klient\*innen auch in ihrem Effekt auszuschöpfen. Die erste B-Phase nimmt dann beispielsweise 20 Sitzungen anstelle von 10 Sitzungen ein, um zur Dauerhaftigkeit des Effekts beizutragen. Ebenso einfach ist das Argument zum unbehandelten Abbruch: Der Forschungsplan endet mit der 30. Messung und einer AV auf Basisratenniveau, der Therapieplan geht jedoch mit einer entsprechend umfänglichen Interventionsphase weiter.

## Multiple Grundratenversuchspläne (MGV)

Neben den AB-Versuchsplänen erfreuen sich die **multiplen Grundratenversuchspläne** (MGV) zunehmender Beliebtheit. Dafür gibt es zwei Gründe:

- Bei **irreversiblen Lernprozessen** können (mit Ausnahme des AB-Designs) die bisher besprochenen Versuchspläne nicht eingesetzt werden, da das Absetzen der Intervention nicht zu einer Rückkehr zur Basisrate führt und somit die Kontrolle möglicher Störeffekte auf diese Weise nicht mehr möglich ist.
- Die **ethischen Probleme** sollten möglichst schon im Forschungsplan ausgeräumt sein.

Aus diesem Grund wird das Individuum möglichst **zeitsynchron** mindestens **drei Interventionen** ausgesetzt und bildet dabei seine eigene Kontrolle. Ein konkreter Fall soll das Vorgehen verdeutlichen.

### Beispiel

#### Vorstellungsbasierte Konfrontationsbehandlung

Saigh (1986) erprobte eine vorstellungsbasierte Konfrontationsbehandlung (Flooding-Technik) an Joseph, einem 6½-jährigen Jungen, der in einem Einkaufszentrum eine Bombenexplosion miterlebt und anschließend eine posttraumatische Belastungsstörung entwickelt hatte. Zur Behandlung entwickelte er 5 Szenen, z. B. von verletzten Personen, die behandelt werden oder von Menschen, die sich einem Einkaufszentrum nähern. Die Ausprägung der Störung von Joseph (AV) wurde jeden Tag mit einer fünfstufigen Skala (5 als maximale und 0 als minimale Ausprägung) für alle 5 Szenen erfasst.

Die Therapie bestand darin, dass Joseph zunächst angeleitet wurde sich zu entspannen und sich anschließend die jeweilige Szene für mehr als 20 Minuten intensiv und in allen Einzelheiten vorstellen sollte. In Abb. 5 sind der weitere Therapieverlauf und die Effekte für 3 der 5 Szenen festgehalten.

Vier Messzeitpunkte reichten dem Autor im vorliegenden Fall, um von einer stabilen Basisrate ausgehen zu können. Mit 5 gemessenen Skalenpunkten war der Junge beim Anblick der Szenen maximal belastet. Am Tag darauf setzte die vorstellungsbasierte Konfrontationstherapie für Szene 1 ein und zeitigte im Verlauf der folgenden 12 Sitzungen eine erfolgreiche Reduktion der Belastung auf einen Skalenpunkt. Gleichzeitig blieb die Basisrate in Bezug auf die Szenen 2 und 3 unverändert.

Nachdem sich in Szene 1 ein stabiler und ausgeprägter Therapieeffekt gezeigt hatte, wurde die Intervention auch für Szene 2 gestartet. Über einen Zeitraum von 8 Sitzungen reduzierte sich das Ausmaß der Belastung von 5 auf 0 Skalenpunkte. Das gleiche Bild ergab sich für Szene 3, nachdem in der 25. Sitzung (bei zuvor stabiler, maximal ausgeprägter Basisrate) die Therapie einsetzte.

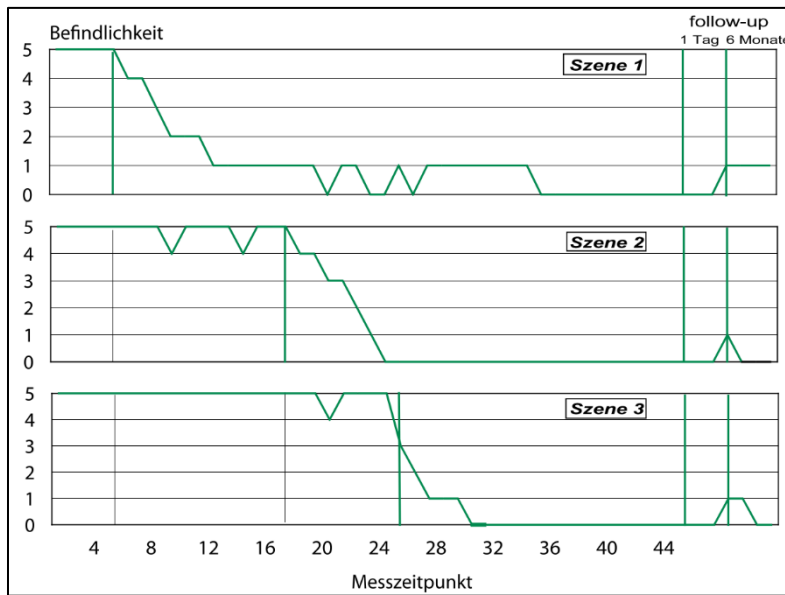


Abb. 5 Ein multipler Grundratenversuchsplan bei einem 6-jährigen Jungen mit einer posttraumatischen Belastungsstörung, behandelt mit einer vorstellungsbasierten Konfrontationstherapie (Befindlichkeit: 5 = maximal gestört, 0 = nicht gestört)

Das Beispiel zeigt, dass bei diesem Vorgehen die **interne Validität** als hoch eingeschätzt werden kann, da es absolut unwahrscheinlich ist, dass ein Störeffekt mit diesem kombinierten Basisraten-Interventions-Muster verbunden sein könnte und die Effekte hervorrufen bzw. verändern könnte. Lern- und Übungeffekte können durch die parallel weiterlaufenden zusätzlichen Basisraten kontrolliert werden. Auch ist Abb. 5 zu entnehmen, dass die Effekte nach Abschluss der Therapie und nach einem halben Jahr stabil bleiben, also kein Rückfall aufzutreten scheint. Dieser Versuchsplan kommt zudem ohne Effektzurücknahme bzw. unbehandelten Abbruch aus.

Gleichzeitig wird an dem Beispiel auch deutlich, dass gewisse **Voraussetzungen** in der Untersuchung erfüllt sein müssen, damit die Ergebnisse kausal interpretiert werden können. Dies sind die Voraussetzungen der **Unabhängigkeit** und der **Ähnlichkeit**.

- Unabhängigkeit bezieht sich auf die beobachteten AVn, die hinsichtlich der Intervention voneinander unabhängig sein müssen. Zu erkennen ist diese Voraussetzung an den stabilen Grundraten bei den nicht direkt beeinflussten AVn.
- Die gleiche Intervention (UV) sollte zu vergleichbaren Veränderungen in allen AVn führen.

Ein Verstoß gegen diese Annahmen führt zur Einschränkung der internen Validität und damit zur notwendigen Zurückhaltung bei der kausalen Ergebnisinterpretation.



## Kriterien quantitativer Einzelfallforschung

Wie kann man die beobachtete Veränderung in einer quantitativen Einzelfallstudie beurteilen? Gängige Kriterien sind die

- statistische Signifikanz,
- pädagogisch-klinische Signifikanz,
- normative Validität und
- soziale Validität.

### Statistische Signifikanz in der Einzelfallforschung

Die statistische Signifikanz beurteilt die Wahrscheinlichkeit des Zustandekommens eines Mittelwertunterschieds oder eines Zusammenhangs in der jeweiligen Stichprobe. Signifikant bedeutet überzufällig und heißt, dass die beobachteten Unterschiede mit dem Zufall nicht mehr erklärt werden können. Die statistische Signifikanz ist von der Größe der Stichprobe abhängig. Deshalb können auch kleine Veränderungen, die im klinischen Kontext unbedeutsam erscheinen, durchaus statistisch signifikant sein.

### Pädagogisch-klinische Signifikanz

Die pädagogisch-klinische Signifikanz abstrahiert vom statistischen Kontext und bezieht sich auf die inhaltliche Bedeutsamkeit der erzielten Effekte. Es wäre weniger missverständlich, dieses Kriterium pädagogisch-klinische Bedeutsamkeit oder Relevanz zu nennen, um es nicht in die Nähe der statistischen Signifikanz zu rücken. Ein\*e Therapeut\*in wird nicht damit zufrieden sein, den Depressionswert von Klient\*innen statistisch signifikant gesenkt zu haben, wenn diese noch immer behandlungsbedürftig sind.

#### Definition

**Pädagogische oder klinische Signifikanz** (je nach dem betroffenen Inhaltsbereich) liegt vor, wenn zwischen den Messwerten aus der Basisratenphase und denen der Interventionsphase keine Überlappungen (Überschneidungen) existieren *und* wenn der Unterschied zwischen dem Mittelwert aus diesen beiden Phasen mindestens 2 Standardabweichungen beträgt.

Von pädagogisch-klinischer Signifikanz geht man auch dann aus, wenn der Prozentsatz nicht überlappender Daten 85% beträgt. Nähere Informationen zur Berechnung dieses Indexes finden sich bei Kern (1997). Abb. 6 illustriert dieses Beurteilungskonzept anhand der Daten zum Fall Joseph (Abb. 3, Szene 3).

Aus diesen Daten wird ersichtlich, dass das Kriterium der pädagogisch-klinischen Signifikanz sehr streng, d. h. schwer zu erreichen ist, denn das Beispiel stellt fast die Ideal- und Maximalform eines Interventionseffekts dar.

© Der/die Herausgeber bzw. der/die Autor(en), exklusiv lizenziert an Springer-Verlag GmbH, DE, ein Teil von Springer Nature 2010, 2013, 2023

Aus: Schreier, M., Echterhoff, G., Bauer, J. F., Weydmann, N. & Hussy, W. (2023). *Forschungsmethoden in Psychologie und Sozialwissenschaften für Bachelor* (3. Aufl.). Springer.



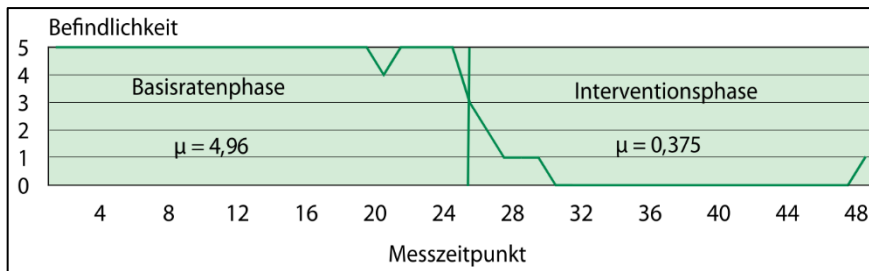


Abb. 6 Im Sinne der pädagogisch-klinischen Signifikanz liegt in diesem Datensatz ein bedeutsamer Therapieeffekt vor, weil es keine Überschneidungen der Daten aus der Basisraten- und Interventionsphase gibt und weil die Veränderung im Mittel 2 Standardabweichungen beträgt (4,96 vs. 0,375)

### Normative Validität

Die normative Validität ergibt sich aus der Übereinstimmung zwischen den Daten einer Person, die an einer Intervention teilgenommen hat und den Daten „unauffälliger“ Personen, die keine Intervention erhalten haben. Unterscheiden sich die Messwerte von Joseph im Anschluss an die Konfrontationstherapie nicht mehr von denen gleichaltriger unbelasteter Kinder, so ist normative Validität gegeben (Abb. 7). Dieses Beurteilungskriterium zieht also das Erreichen der Erlebens- und Verhaltensnorm als Entscheidungsgröße heran: Der Klient bzw. die Klientin ist nach der Intervention „unauffällig“.

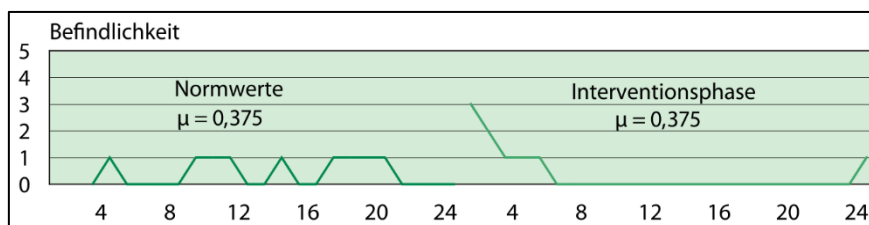


Abb. 7 Im Sinne der normativen Validität liegt ein bedeutsamer Therapieeffekt vor, weil sich die Normwerte und Interventionswerte nicht unterscheiden

### Soziale Validität

Bei der sozialen Validität löst man sich gänzlich von statistischen Kennwerten. Sie ist gegeben, wenn sich nach Meinung von vertrauten Freunden, Bekannten und Verwandten eine starke positive Veränderung durch die Intervention ergeben hat.

Insgesamt lässt sich feststellen, dass die verschiedenen Beurteilungs- und Bewertungskriterien des Interventionserfolgs verschiedene Aspekte der durch die Intervention hervorgerufenen Veränderungen thematisieren. Somit kann es nicht um die Entscheidung gehen, welches



Kriterium heranzuziehen ist, sondern es muss das Ziel sein, durch die gemeinsame Betrachtung der verschiedenen Kriterien zu einem ausgewogenen Gesamturteil zu gelangen.

## Literatur

Grünke, M. (Hrsg.). (2012). Kontrollierte Einzelfallforschung [Schwerpunktheft]. *Empirische Sonderpädagogik*, 4 (3/4).

Kern, H. J. (1997). *Einzelfallforschung. Eine Einführung für Studierende und Praktiker*. Weinheim: Psychologie Verlags Union.

Saigh, P. A. (1986). In vitro flooding in the treatment of a 6-year-old boy's posttraumatic stress disorder. *Behaviour Research and Therapy*, 24, 685–688.

Vlaeyen, J. W. S., Onghena, P., Vannest, K. J. & Kratochwill, T. R. (2022). 3.01 - Single-Case Experimental Designs: Clinical Research and Practice. In G. J. G. Asmundson (Ed), *Comprehensive Clinical Psychology* (2nd Ed., pp. 1–28). Amsterdam: Elsevier. <https://doi.org/10.1016/B978-0-12-818697-8.00191-6>