

Empirische Sonderpädagogik, 2012, Nr. 3/4, S. 211–245

## Versuchspläne der experimentellen Einzelfallforschung

*Anita Jain & Ralf Spieß*

*Universität zu Köln*

### Zusammenfassung

Im vorliegenden Überblicksartikel werden die wichtigsten Versuchspläne experimenteller Einzelfallforschung dargestellt. Hierzu zählen der ABAB-Plan, das Multiple-Baseline-Design, der alternierende Versuchsplan und der Kriterien-Veränderungs-Versuchsplan. Aufbauend auf der Logik experimentellen Vorgehens in Einzelfalluntersuchungen wird beschrieben, wie und unter welchen Voraussetzungen die einzelnen Versuchspläne geeignet sind, interne Validität herzustellen und kausale Interpretationen von Interventionseffekten zu ermöglichen. Zu diesen Voraussetzungen gehört auch eine valide und objektive Datenerfassung. Zur Auswertung von Einzelfallstudien wird das Vorgehen bei der visuellen Inspektion beschrieben. Außerdem erfolgt eine Diskussion verschiedener statistischer Auswerteverfahren mit ihren Vor- und Nachteilen. Eine Generalisierung der Ergebnisse von Einzelfallstudien auf andere Personen oder Situationen wird durch Replikationsstudien ermöglicht.

Schlüsselwörter: Kontrollierte Einzelfallforschung, Versuchspläne, ABAB-Versuchsplan, Multiple-Baseline-Design, Kriterien-Veränderungs-Versuchsplan, alternierender Versuchsplan, visuelle Inspektion, Interventionsforschung, Evaluation

### Research designs for experimental single-case studies

#### Abstract

This overview describes the most important designs for single case experimental studies, which are ABAB-design, multiple baseline design, alternating treatments design and changing criterion design. The logic of experimental control in single case studies is explained and it is described, how these different designs are able to provide internal validity and enable causal interpretations of intervention outcome. An important precondition of valid interpretation is objective and reliable data assessment. Data evaluation by visual inspection is explained and several methods of statistical data analysis are discussed. To establish generability across persons, situations, and settings, the importance of replication studies is highlighted.

Keywords: Single case experimental design, ABAB-design, multiple baseline design, alternating treatments design, changing criterion design, intervention research, visual inspection, evaluation

## Einleitung

„Warum soll ich mich mit Forschung beschäftigen? Ich will ja schließlich Lehrkraft werden. In der Schule lassen sich doch niemals Untersuchungen durchführen.“ So oder so ähnlich lautet häufig der Standardkommentar „forschungskritischer“ Studierender in den Eingangsemestern. Bei näherer Diskussion zeigt sich sehr schnell, dass sie häufig einen bestimmten Typ von Studien im Sinn haben, wenn sie von der Inkompatibilität von Forschung und Schulalltag sprechen: experimentelle bzw. quasi-experimentelle Gruppenuntersuchungen. Der Verweis auf die Möglichkeit, im Schulalltag am Einzelfall orientierte Forschung einzusetzen, löst dann nicht selten Kommentare aus wie: „Ist das überhaupt wissenschaftlich?“, „Können Einzelfallstudien wirklich etwas erklären?“ oder „Sind die Ergebnisse des Einzelfalls denn auf andere Fälle übertragbar?“

Der vorliegende Beitrag soll verdeutlichen, dass Einzelfallforschung keinesfalls Forschung zweiter Klasse ist, sondern experimentelle Gruppenforschung komplementär ergänzt und insbesondere in (sonder-)pädagogischen und psychotherapeutischen Arbeitsfeldern eine notwendige und sinnvolle Alternative darstellt. Einzelfallforschung reicht von anekdotischen Fallberichten (die eher hypothesengenerierenden Charakter haben) bis hin zu experimentellen Anordnungen (die kausale Schlüsse zulassen). Im weiteren Verlauf werden die wichtigsten experimentellen Einzelfalldesigns hinsichtlich ihrer Charakteristika, ihrer Vorteile, aber auch hinsichtlich ihrer Probleme und Begrenzungen anhand illustrierender Beispiele präsentiert. Im Zentrum der Darstellung stehen hierbei das Verständnis der Vorgehensweise und die mit Einzelfallforschung verbundene Logik. Einen vollständigen Überblick zu diesem Themenbereich liefern die Monographien von Barlow, Nock und Hersen (2009), von

Kazdin (2011) oder deutschsprachig von Julius, Schlosser und Goetze (2000).

## Zentrale Konzepte und Begriffsklärungen

Das übergeordnete Ziel wissenschaftlicher Untersuchungen liegt darin, Zusammenhänge zwischen interessierenden Variablen zu beschreiben, zu erklären, ggf. Veränderungen (Interventionen) einzuführen und diese zu bewerten. Hierzu stehen je nach Fragestellung sowohl bei gruppenbasierter als auch bei Einzelfallforschung experimentelle und nicht-experimentelle Designs zur Verfügung. Im Bereich nicht-experimenteller Einzelfalluntersuchungen können beispielsweise Kasuistiken bei der Entwicklung von Hypothesen, Theorien und Interventionen hilfreich sein. Wenn es darum geht, den Erfolg einer Intervention zu bewerten, werden experimentelle Versuchsanordnungen benötigt.

## Experimentelle Designs

Ein experimentelles Design ist durch zwei Kriterien gekennzeichnet:

1. Variation mindestens einer unabhängigen Variablen (UV) und Registrieren der Auswirkung dieser Variation auf eine abhängige Variable (AV)
2. Ausschluss alternativer Erklärungsmöglichkeiten für Veränderungen in der AV (Herstellen interner Validität durch Kontrolle von Störvariablen)

Abbildung 1 verdeutlicht exemplarisch Kriterium 1: Die UV Art der Intervention wird variiert (Interventionsgruppe / Kontrollgruppe) und die Auswirkungen dieser Variation auf die AV Aufmerksamkeitsleistung registriert, indem ihre Messung vor Interventionsbeginn (Prä) und nach Abschluss der Intervention (Post) erfolgt. Eine Möglichkeit, einen Interventionseffekt abzubilden, bietet der Vergleich der Prä-Post-Differenzen von Interventions- und Kontrollgrup-

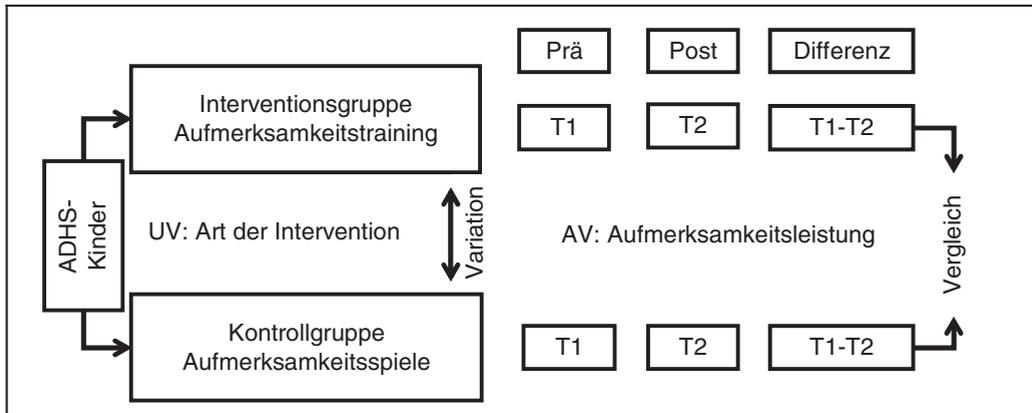


Abbildung 1: Variation der UV und Vergleich der Auswirkung auf die AV bei einem Prä-Post-Kontrollgruppendesign.

pe. Wenn die Differenz in der Interventionsgruppe signifikant besser in Richtung Aufmerksamkeitssteigerung ausfällt als in der Kontrollgruppe, spricht dies für einen Interventionseffekt.

Ein signifikanter Unterschied allein reicht jedoch nicht aus, um einen Interventionseffekt eindeutig zu belegen. Zusätzlich müssen alternative Erklärungen für seine Entstehung ausgeschlossen werden. Maßnahmen, die diesem Zweck dienen, werden oft unter der Bezeichnung „Kontrolle von Störvariablen“ zusammengefasst. Die Kontrolle von Störvariablen ist die Voraussetzung dafür, dass Untersuchungen eine hohe interne Validität aufweisen, dass also die Veränderung in der AV eindeutig auf die Variation der UV zurückgeführt werden kann. Wenn sich in dem oben dargestellten Beispiel in der Interventionsgruppe hauptsächlich unaufmerksame Kinder, in der Kontrollgruppe hingegen eher hyperaktiv/impulsive Kinder befinden, kann nicht ohne Weiteres davon ausgegangen werden, dass der gefundene Unterschied interventionsbedingt ist. Eine plausible Alternativerklärung wäre die unterschiedliche Ausgangslage in Interventions- und Kontrollgruppe. Hier bietet das Prä-Post-Design eine Möglichkeit, derartige Bedingungen zu kontrollieren, indem

die Prä-Post-Differenzen miteinander verglichen werden. In dem vorliegenden Fall reicht das jedoch nicht aus, weil unaufmerksamere Kinder u.U. stärker von einer aufmerksamkeitsfördernden Intervention profitieren könnten als hyperaktive. Bei einer umgekehrten Aufteilung der Mädchen und Jungen auf Interventions- und Kontrollgruppe hätten die aufmerksamkeitschwachen Teilnehmerinnen und Teilnehmer möglicherweise von der Kontrollgruppenbedingung (Aufmerksamkeitsspiele) genauso profitiert wie von dem Aufmerksamkeitstraining. Potentielle Störvariablen, die sich auf Merkmale von Versuchspersonen beziehen (wie Intelligenz, Leistungsbereitschaft oder die Aufmerksamkeitskapazität im oben gewählten Beispiel), werden in experimentellen Gruppenuntersuchungen durch eine Randomisierung oder eine Parallelisierung kontrolliert. Beide Techniken zielen darauf ab, dass potentielle personenbezogene Störvariablen in beiden Untersuchungsgruppen in etwa gleich verteilt sind. Ist dies der Fall, üben sie keinen systematischen Einfluss auf die AV aus und können somit als mögliche Alternativerklärungen nicht mehr in Frage kommen. Randomisierung und Parallelisierung tragen also zur Erhöhung der internen Validität bei. Hierzu

dient auch die Einführung einer Kontrollgruppe, die verhindert, dass Einflüsse, die eigentlich gar nichts mit der Intervention zu tun haben (sogenanntes zwischenzeitliches Geschehen), als potentielle Erklärung für den Interventionseffekt in Frage kommen.

Randomisierte Kontrollgruppenuntersuchungen werden häufig als „Goldstandard“ gruppenbasierter Interventionsforschung bezeichnet, weil mit diesem Design eine hohe interne Validität herstellbar ist, die den Ausschluss potentieller alternativer Erklärungen ermöglicht. Obwohl der Wert randomisierter Kontrollgruppenuntersuchungen in der Interventionsforschung unbestritten ist, sollten auch ihre Begrenzungen bekannt sein. Hierzu zählt u.a., dass nicht jede Fragestellung eine randomisierte Zuweisung zu Experimental- und Kontrollgruppe erlaubt. Viele möglicherweise sinnvolle Interventionen bzw. Interventionsprogramme finden daher niemals den Weg in eine randomisierte Kontrollgruppenuntersuchung. Darüber hinaus hilft es

im konkreten Fall nur bedingt weiter, wenn bekannt ist, dass eine bestimmte Behandlung in einer randomisierten Kontrollgruppenstudie zu signifikanten Verbesserungen in der Interventionsgruppe mit einer Effektstärke von z. B.  $d = 0,8$  geführt hat. Solche Untersuchungen helfen bei der Auswahl von Maßnahmen, sie sagen jedoch wenig darüber aus, ob diese Intervention auch im konkreten Fall (beispielsweise in einer Einrichtung mit einem speziellen Kind oder einer Gruppe von Schülerinnen bzw. Schülern) erfolgreich sein wird. Um solche Fragestellungen zu beantworten, bieten experimentelle Einzelfalluntersuchungen eine sinnvolle Alternative.

### **Experimentelles Vorgehen bei Einzelfalluntersuchungen**

Es wird oft behauptet, dass Einzelfalluntersuchungen keinen experimentellen Charakter aufweisen können. Dies ist, wenn die oben beschriebenen Kriterien heran-

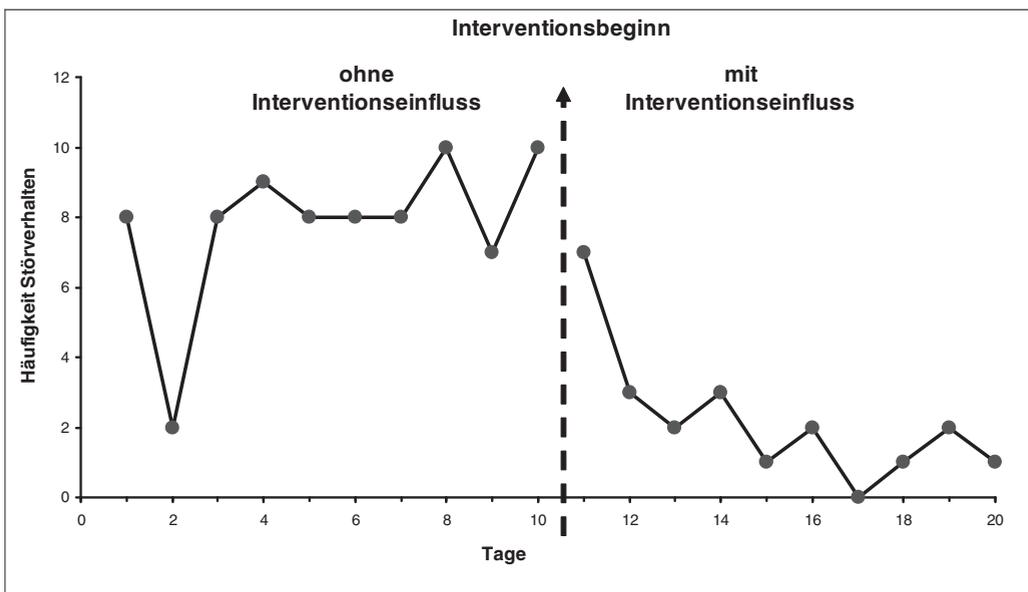


Abbildung 2: Variation der UV in experimentellen Einzelfalldesigns und ihre Auswirkung auf die AV.

gezogen werden, nicht richtig. Sowohl die Variation der UV als auch der Ausschluss alternativer Erklärungsmöglichkeiten sind im Rahmen von Einzelfalldesigns möglich.

### **Variation der UV in Einzelfalluntersuchungen**

Abbildung 2 zeigt die Variation der UV in einem Einzelfalldesign. Während experimentelle Gruppenuntersuchungen (vgl. Abbildung 1) klassischerweise zu zwei Messzeitpunkten eine Interventions- mit einer Kontrollgruppe hinsichtlich des interessierenden Zielverhaltens vergleichen, untersuchen experimentelle Einzelfalldesigns den Effekt eines Treatments an einer Person (bzw. einigen wenigen Personen). Statt zwei Messungen mit vielen Personen (Gruppenuntersuchungen) werden hier viele Messungen bei einer Person durchgeführt. Die Variation der UV entsteht durch die Einführung der Behandlung. Die AV wird also ohne und mit Interventionseinfluss gemessen.

Wiederholte Messungen sind ein zentrales Konzept in der experimentellen Einzelfallforschung. Sie sind notwendig, weil die erfassten Merkmale häufig Schwankungen unterliegen und auch die Messungen selbst fehlerhaft sein können. In Gruppenuntersuchungen kommen wir mit zwei Messzeitpunkten aus, weil sich Schwankungen und Messfehler über Personen ausmitteln. Bei ausreichend großer Stichprobe und Randomisierung verteilen sich die Störgrößen auf die Interventions- und Kontrollgruppe ungefähr gleich.

In Einzelfalluntersuchungen sieht das anders aus. Vergleicht man in Abbildung 2 die Häufigkeit des Störverhaltens vor der Intervention nur an Tag 2 und nach der Intervention nur an Tag 19, wäre der potentielle Interventionseffekt nicht sichtbar geworden. Ein Vergleich von Tag 1 mit Tag 20 zeigt zwar einen potentiellen Effekt, allerdings ist hier nicht auszuschließen, dass einfach nur

ein besonders schlechter mit einem besonders guten Tag verglichen wird. Dies wäre eine plausible Alternativerklärung und somit eine Bedrohung der internen Validität. Wie Abbildung 2 zeigt, sind wiederholte Messungen (s. u., Abschnitt „Baseline und Intervention“) eine Möglichkeit, diese alternative Erklärung auszuschließen.

### **Ausschluss alternativer Erklärungsmöglichkeiten in Einzelfalluntersuchungen**

Kazdin (2011) nennt die in Tabelle 1 dargestellten wesentlichen Bedrohungen der internen Validität, mit denen man sich auseinandersetzen sollte, wenn man eine Einzelfalluntersuchung plant.

Die Kontrolle der Störeffekte History und Maturation sind Gegenstand der verschiedenen Einzelfallversuchspläne, die im übernächsten Kapitel beschrieben werden. Instrumentation und Diffusion of treatment sind Bedrohungen der internen Validität, die generelle Aspekte der Versuchsplanung und Versuchsdurchführung betreffen. Sie können beispielsweise durch Schulung der Beobachterinnen und Beobachter (Instrumentation) oder durch Schulung der Versuchsleiterinnen und Versuchsleiter (Diffusion of treatment) kontrolliert werden. Veränderungen aufgrund wiederholten Messens (Testing) werden beeinflusst durch die Operationalisierung der AV, also durch die Auswahl der Messinstrumente, die bei der Datenerhebung in experimentellen Einzelfallstudien Einsatz findet. Dieser Aspekt wird im folgenden Abschnitt behandelt.

### **Datenerhebung in experimentellen Einzelfallstudien**

Ganz entscheidend für die Qualität einer Einzelfallstudie ist eine angemessene Operationalisierung und Messung der AV. Experimentelle Einzelfallstudien sollten vorzugsweise beobachtbares Verhalten als

Tabelle 1: Bedrohungen der internen Validität in Einzelfallstudien (in Anlehnung an Kazdin, 2011, S. 29).

Bedrohung der internen Validität durch	
<b>History</b> Zwischenzeitliches Geschehen	bezieht sich auf Ereignisse, die im Verlauf der Untersuchung auftreten und eine plausible alternative Erklärung darstellen. Hierzu zählen beispielsweise familiäre Krisen, Lebensveränderungen, Lehrerwechsel etc.
<b>Maturation</b> Reifung	bezieht sich auf Reifungsprozesse im Verlauf der Untersuchung, die möglicherweise eine plausible alternative Erklärung darstellen.
<b>Instrumentation</b> Veränderungen des Messprozederes über die Zeit	entsteht beispielsweise, wenn Beobachter ihre Einschätzungskriterien über die Zeit verändern. Dies wäre der Fall, wenn Störverhalten zu Beginn einer Untersuchung anders definiert wird als im weiteren Verlauf.
<b>Testing</b> Veränderungen aufgrund wiederholten Messens	ist immer dann relevant, wenn Messinstrumente eingesetzt werden, bei denen allein die wiederholte Darbietung des Messinstrumentes zu Veränderungen führt (z. B. Übungseffekte bei einem Leistungstest).
<b>Diffusion of treatment</b> Unzureichende Trennung von Interventions- und Nicht-Interventionsphasen	tritt z. B. dann auf, wenn in Nicht-Interventionsphasen Interventionen weitergeführt werden (vgl. ABAB-Design).

AV verwenden, weil Verhalten objektiver erfasst werden kann und weniger anfällig für Verzerrungen ist als beispielsweise Einschätzungen. Die Verhaltensweisen, die zur Beobachtung ausgewählt werden, sind hierbei oft Bestandteile des zu verändernden Problemverhaltens.

Möchte etwa eine Sonderschullehrerin registrieren, wie häufig bei einem Schüler bestimmte störende Verhaltensweisen auftreten, wäre es eine sinnvolle Möglichkeit, diese Variable innerhalb eines vorher definierten Zeitintervalls (von beispielsweise einer Unterrichtsstunde) mit einer Strichliste zu erfassen. Dies ist ein deutlich reliableres Vorgehen als eine nachträgliche Einschätzung des Störverhaltens am Ende der Unterrichtsstunde.

In Abhängigkeit vom Zielverhalten können verschiedene Aspekte wie Häufigkeit oder Dauer einer Verhaltensweise gemessen werden. Neben direkter Verhaltensbeobachtung ist je nach Fragestellung auch

die Erfassung von Verhaltenskonsequenzen (z.B. ein nasses Bett bei einem Kind mit Enuresis), von objektiven Maßen (z.B. Körpergewicht) oder von physiologischen Variablen (z.B. Elektromyografie bei Verspannungen) sinnvoll.

Wenn Selbst- oder Fremdeinschätzungsdaten im Rahmen einer Einzelfallstudie verwendet werden, sollte man sich darüber im Klaren sein, dass diese anfällig für Verzerrungstendenzen wie z.B. soziale Erwünschtheit sein können. In Bereichen, in denen Verhalten einer direkten Beobachtung nicht zugänglich ist (z. B. Schmerzwahrnehmung, Stimmung) sind sie unerlässlich. In anderen Fällen, in denen solche Einschätzungen erhoben werden sollen, ist zu überlegen, ob sich diese Variablen nicht durch weitere, objektivere Maße ergänzen lassen. Generell gibt es einige wichtige Anforderungen an eine AV in experimentellen Einzelfalluntersuchungen. Hierzu gehören kontinuierliches Messen, reliable Erfassung

und Änderungssensitivität (vgl. Kazdin, 2011, S. 59).

Wie oben bereits dargelegt, ist eine der Grundlagen experimenteller Einzelfallforschung das Konzept des wiederholten Messens. Die Aufzeichnung der AV sollte hierbei möglichst kontinuierlich (z.B. jeden Tag) stattfinden. Daraus folgt, dass eine AV in einer Einzelfalluntersuchung so beschaffen sein muss, dass sie keinen oder nur geringen Übungs- oder Gewöhnungseffekten unterliegt (vgl. Tabelle 1: Testing).

Darüber hinaus ist es notwendig, dass eine AV reliabel erfasst werden kann und änderungssensitiv ist. Reliable Erfassung bezieht sich darauf, dass die Fehler bei der Datenerhebung möglichst gering sein sollten. Fehler bei der Datenerhebung können beispielsweise durch unklare Beobachtungskriterien entstehen. Eine genaue Beschreibung, was bei der Erstellung von Beobachtungsplänen zu beachten ist, findet sich bei Kazdin (2011, S. 56ff).

Änderungssensitiv bedeutet, dass die ausgewählten Variablen Veränderungen in dem durch die Intervention anvisierten Zielbereich angemessen abbilden können. Hierzu sollte eine AV so gestaltet sein, dass sie nach Möglichkeit Werte auf einer kontinuierlichen oder mehrstufigen Skala statt dichotomer Einordnungen erlaubt. In einer Schulstunde ist es beispielsweise sinnvoller, die Häufigkeit des Störverhaltens bei einem Schüler zu erfassen (kontinuierliche Skala), als einfach nur zu registrieren, ob Störverhalten aufgetreten ist oder nicht (dichotome Erfassung). Wenn durch eine Intervention das mittlere Störverhalten von zwölf Mal pro Stunde auf zwei Mal pro Stunde gesenkt wird, wäre dieser Interventionserfolg durch eine dichotome Erfassung nicht abbildbar. Die AV hätte in diesem Fall also keine ausreichende Änderungssensitivität.

Der Aufwand einer Einzelfallstudie wird deutlich reduziert, wenn keine externe Beobachterin bzw. kein externer Beobachter erforderlich ist, sondern die Datenerfassung

entweder durch diejenige Person, die das Training durchführt (Lehrkraft, Therapeutin oder Therapeut, Trainerin oder Trainer), durch die Betroffenen selbst oder durch Angehörige (z. B. Eltern) durchgeführt werden kann. Technische Geräte (z. B. eine Stoppuhr) können dabei die Beobachterinnen und Beobachter unterstützen. In manchen Fällen ist es möglich, das Zielverhalten durch Audio- bzw. Videoaufnahmen aufzuzeichnen und später auszuwerten. Hierdurch wird die Objektivität der Datenerfassung erhöht. Sie lässt sich überprüfen, indem das zur Verfügung stehende Videomaterial durch mindestens zwei unabhängige Beobachterinnen bzw. Beobachter ausgewertet wird. Stimmen die Beobachtungsergebnisse weitgehend überein, ist dies ein Beleg für die Objektivität der Datenerfassung.

## **Die wichtigsten Versuchspläne der experimentellen Einzelfallforschung**

Im Folgenden werden die wichtigsten Versuchspläne (Designs) der Einzelfallforschung beschrieben und anhand von fiktiven Beispielen und Literaturbeispielen illustriert. Das Ziel der verschiedenen Versuchsdesigns besteht darin, durch die Art der Versuchsanordnung für unterschiedliche Praxisbedingungen ein möglichst hohes Maß an interner Validität herzustellen. Da sich Einzelfallforschung dadurch auszeichnet, dass Designs flexibel den Erfordernissen der Praxis angepasst werden können, existiert eine Vielzahl von Abwandlungen und Kombinationen. Die nachfolgenden Abschnitte versuchen, die wesentlichen Eigenschaften der wichtigsten Standarddesigns darzustellen. Weitere Details finden sich u.a. bei Barlow (2009) und Kazdin (2011).

## Der AB-Plan

Alle Einzelfallversuchspläne basieren auf dem sogenannten AB-Plan. Unter der A-Phase versteht man die Baseline oder Grundrate. In der Baseline werden die Ausprägungen der AV vor dem Einsetzen einer Intervention registriert. Die B-Phase bezeichnet die Interventionsphase. Hier werden die Ausprägungen der AV unter Interventionsbedingungen erfasst.

Die Erhebung der Baseline hat drei Ziele: Deskription, Prognose und Vergleich (siehe Abbildung 3). Deskription bedeutet, dass die Ausprägung des Verhaltens ohne Intervention beschrieben wird. Hieraus lässt sich eine Prognose über die voraussichtliche Weiterentwicklung des Zielverhaltens ohne Intervention erstellen. Der Vergleich dieser Prognose mit den tatsächlich erhobenen Daten aus der Interventionsphase bildet die Grundlage zur Abschätzung des Interventionseffektes. Wenn sich das beobachtete Verhalten nach Einsetzen der Inter-

vention von dem Verhalten unterscheidet, das man aufgrund der Baseline erwarten kann (Prognose), deutet dies auf eine Wirkung der Intervention hin.

Abbildung 3 verdeutlicht diesen Sachverhalt für das Beispiel einer Sonderschullehrerin, die durch eine Intervention das Störverhalten eines Schülers reduzieren möchte: Nach einer Baseline, die 10 Messzeitpunkte umfasst und relativ konstante Werte aufweist, erfolgt die Interventionsphase, in der sich eine rasche und deutliche Reduktion des Problems beobachten lässt.

### Baselinevoraussetzungen zur Beurteilung von Interventionseffekten

Nur wenn die Baseline genügend Information darüber liefert, wie das Verhalten ohne Intervention beschaffen wäre, kann ein möglicher Interventionseffekt verlässlich beurteilt werden. Dies ist besonders dann gegeben, wenn die Baseline genügend Messzeitpunkte umfasst und stabil ist,

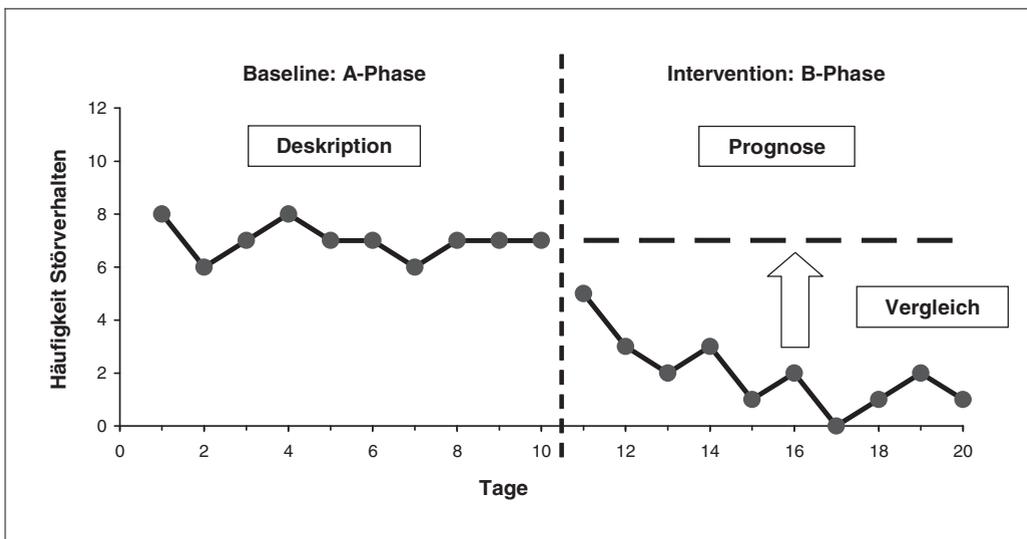


Abbildung 3: Funktionen der Baseline. Die durchgezogene Linie zeigt das tatsächlich aufgetretene Verhalten während Baseline und Intervention, die gestrichelte Linie das aufgrund der Baseline prognostizierte Verhalten ohne Intervention.

d.h. wenige Schwankungen sowie keine Tendenz nach oben oder unten (Trend) aufweist.

### **Anzahl der Messungen in der Baseline**

Wie viele Messungen eine Baseline aufweisen sollte, um ihrer deskriptiven und prognostischen Funktion gerecht zu werden, hängt sehr stark von ihrem Verlauf ab. In manchen Fällen kann sie sehr kurz sein oder man kann sogar darüber diskutieren, ob überhaupt eine Baseline erhoben werden muss, wenn bereits verlässliche Informationen über die Ausprägung des Zielverhaltens zur Verfügung stehen. Ein solcher Fall liegt z. B. bei einer Schülerin mit einer Agoraphobie vor, von der man aufgrund von Selbst- und Fremdauskünften weiß, dass sie ihre Wohnung seit drei Monaten nicht mehr ohne Begleitung verlassen hat. Als ein Erfolgskriterium für eine verhaltenstherapeutische Behandlung soll die Zeit erfasst werden, die das Mädchen täglich alleine außerhalb ihrer Wohnung verbringt. Hier reicht sicher eine sehr kurze Baseline aus, welche die Aussage der Schülerin bestätigt. Der hier skizzierte Fall zeichnet sich dadurch aus, dass die Baseline stabil ist (das Mädchen geht nie allein aus dem Haus) und es zudem Hinweise darauf gibt, dass sie auch vorher (zumindest in den letzten drei Monaten) stabil war. Auf der Basis dieser Daten ist die Prognose angemessen, dass die Schülerin ohne Intervention auch zukünftig die Wohnung nicht allein verlassen wird. Diese günstigen Baselinebedingungen für eine angemessene Prognose finden sich leider eher selten. Um eine Einzelfalluntersuchung zu planen, wird man vor ihrem Beginn die Anzahl an Messungen in der Baseline festlegen wollen, um die aufzubringende Zeit und den nötigen Aufwand abschätzen zu können. Nach Abschluss der Baselinephase (egal, ob nach drei Tagen oder nach drei Wochen) sollte

man sich in jedem Fall fragen, ob die hierbei erhobenen Daten ihrer prognostischen Funktion gerecht werden können.

### ***Stabilität der Baseline***

Prognostisch ideal ist eine möglichst stabile Baseline, die wenig Schwankungen zeigt und trendfrei ist (also keine Tendenz nach oben oder unten aufweist). Bei starken Abweichungen ist eine Interventionswirkung viel schwieriger zu erkennen als bei einer stabilen Baseline, weil sich dann die Werte der Interventionsphase nicht so eindeutig unterscheiden. Dies zeigt der Vergleich von Abbildung 4a und Abbildung 4b. Trotz identischer Mittelwerte beider Baselinephasen spricht der Verlauf in Abbildung 4a deutlicher für einen Interventionseffekt, weil die Prognose des weiteren Verlaufs hier eindeutiger ausfällt.

### ***Trends in der Baseline***

Auch Trends in der Baseline erschweren die Interpretation einer Interventionswirkung. Zu unterscheiden sind hier Trends in Interventionsrichtung von solchen, die gegenläufig zur beabsichtigten Wirkung verlaufen. Geht es um die Reduktion von störendem Verhalten im Unterricht und weist die Baseline wie in Abbildung 4c einen Trend in diese Richtung auf, ist eine Interpretation potentieller Behandlungseffekte nicht möglich. Zwar sinkt die Häufigkeit des Störverhaltens in der Interventionsphase weiter ab; es kann jedoch nicht gesagt werden, ob dies eine Interventionswirkung oder einfach nur die Fortsetzung des bereits bestehenden Trends ist. Abbildung 4d zeigt hingegen einen Anstieg des Störverhaltens in der Baseline. Da hier der Trend gegenläufig zum beabsichtigten Treatmenteffekt ist, kann die Interventionswirkung interpretiert werden. Eine Fortsetzung des bestehenden Trends wäre hier keine plausible Alternativerklärung.

### Maßnahmen zur Erhöhung der Baseline Stabilität

Generell gilt also, dass sich auf der Basis einer wenig variablen und trendfreien Baseline eine besonders zuverlässige Prognose darüber erstellen lässt, wie sich das Verhalten ohne Intervention entwickelt hätte. Aus diesem Grund ist es sinnvoll,

sich bereits vor Untersuchungsbeginn Gedanken darüber zu machen, wie die Stabilität der Baseline erhöht werden kann. Je besser es gelingt, durch eine gute Untersuchungsplanung potentielle Varianzquellen auszuschalten, die unabhängig von der Intervention einen Einfluss auf die Variabilität der Baseline haben, desto stabiler wird die Baseline sein. Potentielle Varianzquel-

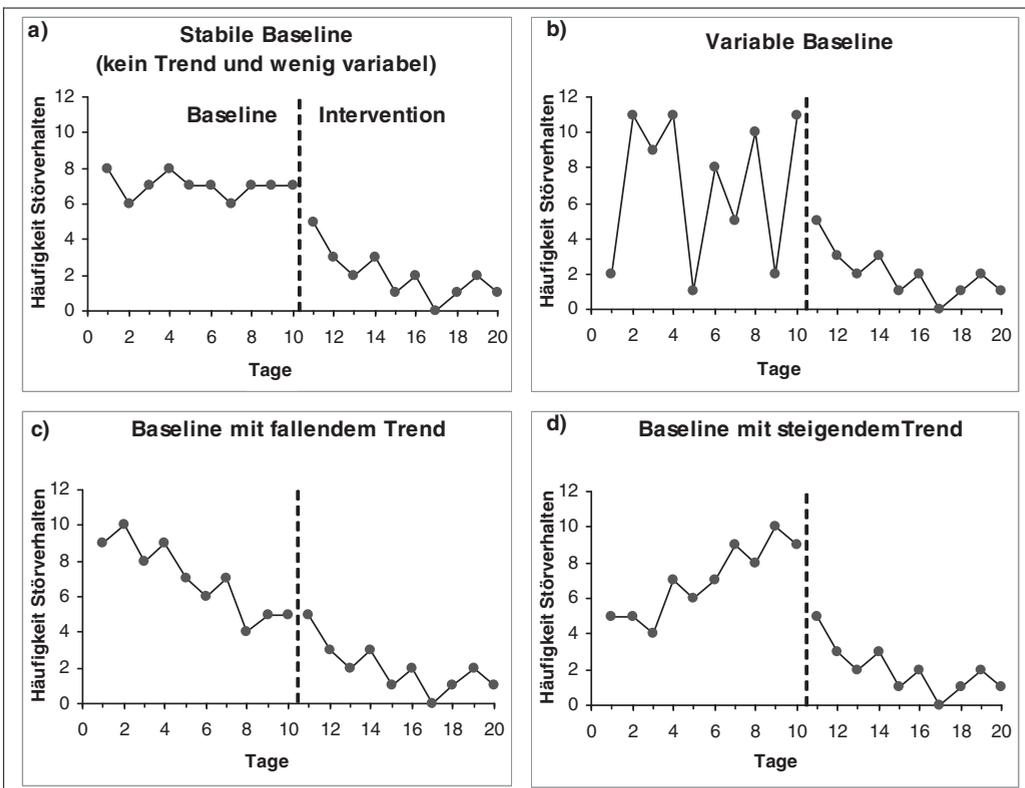


Abbildung 4: Ablauf eines AB-Plans und unterschiedliche Baselineverläufe.

Abbildung 4a: Eine stabile Baseline erlaubt eine eindeutige Beurteilung des Interventionseffektes.

Abbildung 4b: Bei einer variablen Baseline ist die Beurteilung des Interventionseffektes schwieriger, auch wenn sich A- und B-Phase im Mittel genauso stark unterscheiden wie in Abbildung 4a.

Abbildung 4c: Bei einem Baseline-Trend in die erwünschte Richtung kann der Interventionseffekt nicht beurteilt werden.

Abbildung 4d: Bei einem Baseline-Trend in die unerwünschte Richtung kann man die Interventionswirkung dann eindeutig erkennen, wenn sie stark ausgeprägt ist.

len sind in Einzelfalluntersuchungen beispielsweise Merkmale der Situation (z. B. Störverhalten wird in unterschiedlichen Schulfächern aufgezeichnet), der Tageszeit (z. B. einmal wird das Störverhalten morgens, einmal nachmittags registriert) oder Merkmale der Versuchsleiterin bzw. des Versuchsleiters (z. B. wird die Intervention einmal von der Lehrkraft X, einmal von der Lehrkraft Y durchgeführt). Diese Variablen sollten nach Möglichkeit konstant gehalten werden (gleiches Fach, gleiche Tageszeit, gleiche Lehrkraft), um die Variabilität der Baseline zu reduzieren.

Diese Vorgabe gilt natürlich nur dann, wenn die oben genannten Variablen nicht aufgrund des Versuchsdesigns selbst verändert werden müssen. Geht es beispielsweise darum, zu zeigen, dass eine Intervention sowohl bei der Lehrkraft X als auch bei der Lehrkraft Y eine zeitnahe Verbesserung mit sich bringt, ist diese aus versuchsplanerischen Gründen von zwei unterschiedlichen Lehrkräften durchzuführen (vgl. beispielsweise Multiple-Baseline-Designs über Situationen, s. u.).

Wenn die Baseline trotz sorgfältiger Planung sehr variabel ist, besteht auch während der Untersuchung noch die Möglichkeit, weitere Variationsquellen zu identifizieren, auszuschalten und die Baselinephase so lange fortzusetzen, bis sich eine ausreichende Stabilität einstellt. Zeigt sich während der Baseline ein Trend in Richtung der angezielten Interventionswirkung, sollte die Baseline-Erhebung nach Möglichkeit fortgesetzt werden, bis sich ein trendfreier Verlauf ergibt. Hier ist individuell zu entscheiden, ob es vertretbar ist, mit dem Beginn der Intervention zu warten. In Einzelfällen (z. B. bei besonders hoher Symptombelastung) kann es sein, dass man sich trotz eines entsprechenden Trends in der Baseline (vgl. Abbildung 4c) aus ethischen Gründen dazu entschließt, eine Behandlung zu beginnen, obwohl schon ohne Treatment eine Besserung eintritt. Nur las-

sen sich dann, wie oben beschrieben, keine Rückschlüsse auf die Wirksamkeit der Intervention ziehen.

### **Interventionsphase**

Auch die Behandlungsphase (B-Phase) eines AB-Plans sollte über genügend Messzeitpunkte verfügen, damit eine zuverlässige Beurteilung des Interventionseffektes erfolgen kann. Die Anzahl der Messzeitpunkte, die hierfür notwendig sind, hängt von den Eigenschaften der Baseline und von der Stärke des Effektes ab. Wenn sich das beobachtete Verhalten auf der Basis einer stabilen Baseline mit Einsetzen der Maßnahme möglichst direkt und deutlich in die vorhergesagte Richtung verändert, so spricht dies für einen Interventionseffekt (siehe auch Abschnitt Datenauswertung).

### **Nachteile des AB-Plans**

Der Nachteil eines AB-Plans besteht darin, dass seine interne Validität eingeschränkt ist. Auch wenn sich eine deutliche Veränderung der Daten in der Interventionsphase gegenüber der Baseline zeigt, lässt sich nicht ausschließen, dass diese Veränderung möglicherweise auf Störvariablen zurückgeht. Dabei kann es sich um externe Ereignisse, um individuelle Entwicklungen oder um Reifungsprozesse handeln, die auch ohne ein Treatment stattgefunden hätten. Der AB-Plan wird deshalb auch als quasi-experimenteller Versuchsplan bezeichnet. Seine interne Validität ist gering, da externe Ereignisse (History) und individuelle Reifungs- und Entwicklungsprozesse (Maturation) oft plausible Alternativerklärungen darstellen. Im Beispiel zu Abbildung 3 könnte es möglicherweise sein, dass sich etwa zeitgleich mit der Einführung des Verstärkerplans durch die Lehrerin das Erziehungsverhalten der Eltern unabhängig von der Intervention verändert hat.

### **Varianten des AB-Versuchsplans**

An die B-Phase kann sich noch eine Follow-up-Messung anschließen, die einige Zeit nach Ende der Intervention durchgeführt wird. Ziel ist es dabei, die Stabilität der Verhaltensänderung über die Zeit zu beurteilen (Nachhaltigkeit des Interventionseffektes).

### **Der ABAB-Plan**

Ein ABAB-Plan stellt eine Wiederholung eines AB-Plans dar, d. h. nach der ersten Interventionsphase (B1-Phase) kehrt man zur Baseline zurück (A2-Phase) und führt dann nach einer gewissen Zeit, in der die Messung fortgesetzt wird, die Intervention zum zweiten Mal ein (B2-Phase). Alle Versuchspläne, bei denen auf eine Interventionsphase wieder eine Baseline folgt, bezeichnet man als Umkehrpläne oder Ausblendungsdesigns. Hier wird eine bereits dargebotene Intervention vorübergehend wieder ausgesetzt. Abbildung 5 verdeutlicht den idealtypischen Verlauf eines ABAB-Plans.

Als Beispiel für einen sinnvollen Einsatz des ABAB-Plans sei an dieser Stelle das Vorgehen einer Lehrerin angeführt, die die Häufigkeit aktiver Unterrichts-beteiligung eines Schülers fördern möchte. Zunächst registriert sie während einer Baseline von acht Tagen, wie häufig sich der Junge in den ersten zwei Unterrichtsstunden eines Tages meldet. Am neunten Tag der Untersuchung beginnt sie mit einem vorher festgelegten Verstärkerplan zu arbeiten, um die Unterrichts-beteiligung zu erhöhen. Diese Intervention wird 8 Tage durchgeführt. Während dieser Zeit zeigt sich ein rascher und deutlicher Anstieg in der Anzahl der Wortmeldungen im Unterricht. Da aber nicht zweifelsfrei belegbar ist, dass der Anstieg an erwünschtem Verhalten wirklich auf den Verstärkerplan zurückgeht, sondern auch andere Ursachen haben könnte, die zufällig mit der Einführung des Verstärkerplans zusammenfallen, wird die Maßnahme in der Phase A2 wieder zurückgenommen. Da sich hier die Anzahl der Wortmeldungen deutlich reduziert und mit dem erneuten Einsetzen der Intervention (B2) wieder

Table 2: Zusammenfassung AB-Plan.

<b>Kurzbeschreibung des Designs</b>	Auf eine Baseline folgt die Interventionsphase.
<b>Vorteile</b>	Einfach umzusetzen, wenige praktische und ethische Probleme (siehe im Gegensatz dazu „Probleme des ABAB-Plans“), wenige Voraussetzungen.
<b>Nachteile</b>	Geringe interne Validität: Eine Veränderung im Zielverhalten kann in der Regel nicht sicher auf die Intervention zurückgeführt werden. Keine Kontrolle von History und Maturation.
<b>Voraussetzungen</b>	Eine ausreichend lange Baselinemessung muss möglich sein.
<b>Bewertung</b>	Ein AB-Plan ist von allen Einzelfallplänen in der Praxis am einfachsten durchzuführen. Mit einem AB-Plan kann aber eine Interventionswirkung nicht zweifelsfrei nachgewiesen werden, da es in der Regel mögliche und plausible Alternativerklärungen für Veränderungen der AV gibt. Er wird deshalb meist dann eingesetzt, wenn sich aufwändigere Versuchspläne aus praktischen oder ethischen Gründen nicht umsetzen lassen. Durch Wiederholung ( <i>Replikation</i> , s. u.) gewinnt ein AB-Plan deutlich an Aussagekraft.

ansteigt, ist es sehr wahrscheinlich, dass tatsächlich das Treatment das Verhalten beeinflusst hat und die Veränderungen nicht auf andere Umstände (wie z.B. Reifungsprozesse, Änderung im Erziehungsverhalten der Eltern, Wetter, Schlaf, soziale Konflikte) zurückführbar sind.

Nach Barlow et al. (2009) ist das ABAB-Design vermutlich der bekannteste und am häufigsten eingesetzte experimentelle Einzelfall-Versuchsplan. Sein großer Vorteil liegt in folgender Eigenart: Es lässt sich ggf. zweimal zeigen, dass eine Intervention das relevante Verhalten in die gewünschte hin Richtung verändert. Außerdem ist feststellbar, ob sich ein Verhalten verschlechtert, wenn die Intervention ausgesetzt wird. Dieser Versuchsplan besitzt somit eine sehr hohe interne Validität. Es wäre nämlich sehr unwahrscheinlich, dass die Wirkung von Störvariablen zeitlich dreimal mit einer Veränderung der UV zusammenfällt.

Wenn sich eine deutliche Verbesserung im Verhalten zwischen A1 und B1 sowie zwischen A2 und B2 und gleichzeitig eine

Verschlechterung im Vergleich von B1 zu A2 zeigt, gilt eine funktionale Beziehung von Intervention und Verhalten gemeinhin als gesichert. Dabei ist es nicht unbedingt erforderlich, dass die Messungen in der A2-Phase das Niveau der A1-Phase erreichen; wichtig ist vielmehr, dass es eine deutliche Verschlechterung von B1 nach A2 gibt, um die Wirksamkeit der Intervention zu belegen.

**Probleme bei ABAB-Plänen**

Problematisch an allen Umkehrplänen ist die Tatsache, dass sie in der Regel nur eingesetzt werden können, wenn die in der Interventionsphase aufgetretenen Verhaltensänderungen zumindest teilweise reversibel sind und wenn zu erwarten ist, dass sich mit Absetzen der Intervention wieder eine Veränderung in Richtung auf das Ausgangsniveau einstellt. Damit sind Umkehrpläne dann geeignet, wenn man mit keinen oder nur mit geringen Transfereffekten zwischen der ersten Interventions- und

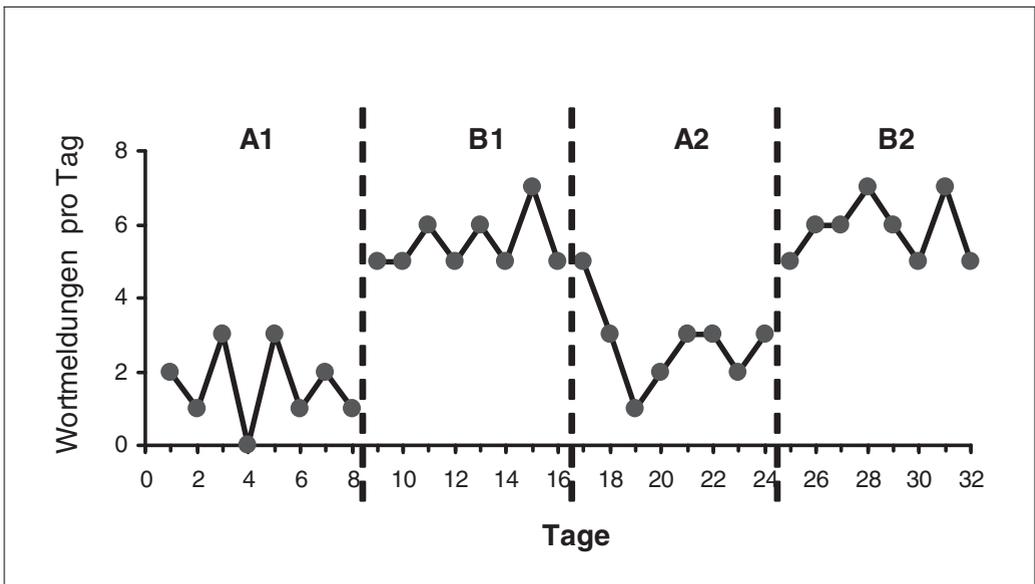


Abbildung 5: Idealtypischer Verlauf eines ABAB-Plans.

der zweiten Baselinephase rechnet. Dabei handelt es sich allerdings oftmals um eine Frage des optimalen Timings: Manchmal sind Interventionseffekte nach einer kurzen Förderung noch reversibel, nach einer längeren Behandlung hingegen dauerhaft. In solchen Fällen bietet es sich an, die erste B-Phase relativ kurz durchzuführen. Diese erste Interventionsphase und die zweite Baselinephase hätten dann den Zweck, die Wirksamkeit der Maßnahme zu belegen, die (längere) zweite Behandlungsphase hätte das Ziel, den Erfolg der Maßnahme zu stabilisieren.

Ein weiteres Problem, das mit Umkehrdesigns verbunden ist, besteht darin, dass man mit dem Ausblenden der Intervention eine Verschlechterung des Zielverhaltens in Kauf nimmt, da eine erfolgreiche Intervention nicht fortgesetzt wird. Eine Verschlechterung des Zielverhaltens ist in diesem Fall aus versuchsplanerischen Gründen sogar erwünscht, um die Wirksamkeit der Intervention belegen zu können. Sie führt aber oft zu erheblichen Akzeptanzproblemen seitens der Probandin bzw. des Probanden, des sozialen Umfelds oder derjenigen Personen, welche die Intervention durchführen. Es ist oft schwer zu vermitteln, warum ein erfolgreiches Treatment vorübergehend ausgesetzt wird. Sind die Folgen einer Verschlechterung in der AV in der zweiten Baseline so negativ, dass man sie ethisch nicht vertreten kann (z. B. bei selbstverletzendem Verhalten, Suizidgefahr oder extremem Untergewicht), muss ein anderer Versuchsplan gewählt werden.

### **Varianten und Erweiterungen des ABAB-Designs**

In Abwandlung eines herkömmlichen ABAB-Plans kann man zusätzlich eine Generalisierungsphase einbauen, die sich an die letzte Interventionsphase anschließt. Hierbei werden wichtige Merkmale der Untersuchungssituation verändert, um ei-

nen Transfer auf andere Situationen zu fördern. In der am Ende dieses Abschnittes beschriebenen Untersuchung von Degner und Nußbeck (2011) wurde beispielsweise ein strukturiertes Arbeitssystem von dem Autismuszentrum, in dem es eingeführt wurde, in die Schule der Kinder übertragen.

Des Weiteren ist es denkbar, eine Follow-up-Phase mit dem Ziel einzuplanen, die Stabilität einer Interventionswirkung zu überprüfen. Hierzu werden in einem festgelegten zeitlichen Abstand zum Interventionsende (z. B. nach sechs Monaten) erneut Messungen über einen vorher definierten Follow-up-Zeitraum durchgeführt.

Schließlich wäre es nach Maßgabe eines so genannten ABC-Designs (Multi-Treatment-Designs) möglich, nach dem Ende der ersten Intervention (B) eine zweite Intervention (C) einzuführen, wobei die erste entweder rückgängig gemacht wird (A-B-C-Design) oder bestehen bleibt (A-B-BC-Design). Sinnvoll kann ein A-B-BC-Design beispielsweise dann sein, wenn es um die Beantwortung der Frage geht, ob die Intervention C zusätzlich zur Intervention B einen Effekt ausübt (z. B. bei einem Vergleich von einem Verstärkerplan alleine mit einem Verstärkerplan plus einem Selbstinstruktionstraining). Problematisch an ABC-Designs sind potentielle Reihenfolgeeffekte. So kann etwa die Wirkung der Intervention C anders ausfallen, wenn zuvor Intervention B durchgeführt wurde, als wenn dies nicht der Fall war.

Zur Abfolge von A- und B-Phasen gibt es zahlreiche Varianten, z. B. ABA-Design, BAB-Design, ABABAB-Design, deren Eigenschaften, Vor- und Nachteile hier nicht dargestellt werden können (siehe stattdessen z. B. Barlow et al., 2009; Julius et al., 2000; Kazdin, 2011).

### **Literaturbeispiel für ein ABAB-Design**

Degner und Nußbeck (2011) untersuchten an zwei Jungen mit Autismus die Wirksam-

keit strukturierter Arbeitssysteme zur Förderung der Selbständigkeit dieser Kinder. Dabei kam ein ABAB-Versuchsplan mit Generalisierungsphase und Follow-up-Phase zum Einsatz. Als AV dienten Einschätzungen der Aufgabenbezogenheit und der Personenunabhängigkeit als Indikatoren für selbständiges Arbeiten, die aus Video-Aufzeichnungen beim Lösen von Aufgaben mit oder ohne Strukturierungshilfen angefertigt und von zwei unabhängigen Beobachterinnen bzw. Beobachtern kodiert wurden. Abbildung 6 zeigt den Verlauf der AV Aufgabenbezogenheit bei einem der beiden Jungen. Der Verlauf der AV folgt eng dem Wechsel von Baseline und Intervention, so dass die Wirksamkeit der Behandlung zur Steigerung des aufgabenbezogenen Arbeitsverhaltens bei diesem Kind als gut belegt gelten kann. Wie sich erkennen lässt,

war hier die Umkehrphase (A2) deutlich kürzer als die erste Baseline (A1) und die beiden Interventionsphasen. Aus methodischer Sicht ist diese A2-Phase dennoch sehr wichtig, weil durch das Absinken der AV in dieser Phase und den erneuten Anstieg in der nächsten Interventionsphase (B2) der Einfluss der Intervention auf das Arbeitsverhalten eindeutig nachgewiesen wird.

### Multiple-Baseline-Designs

Bei einem Multiple-Baseline-Design (auch multipler Grundratenversuchsplan genannt) handelt es sich im Prinzip um einen AB-Plan, der mehrmals mit unterschiedlichen Verhaltensweisen, in unterschiedlichen Situationen oder mit verschiedenen Personen wiederholt wird. Er ist dann besonders sinnvoll, wenn ein Umkehrplan aus den

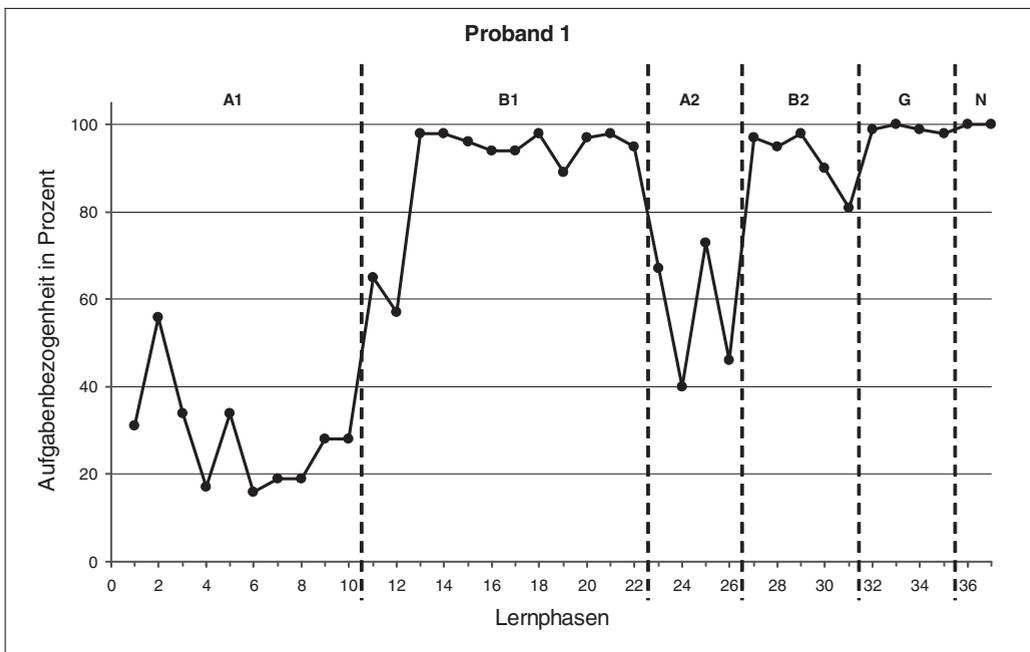


Abbildung 6: Literaturbeispiel für ein ABAB-Design zur Überprüfung der Wirksamkeit strukturierter Arbeitssysteme auf das Arbeitsverhalten eines autistischen Jungen. A1 und A2: Baseline; B1 und B2: Interventionsphasen; G: Generalisierungsphase; N: Follow-up nach sechs Monaten (modifiziert nach Degner & Nußbeck, 2011, S. 64).

oben beschriebenen Gründen nicht durchführbar ist.

### **Das Multiple-Baseline-Design über Personen**

Ein Multiple-Baseline-Design mit einer Variation über Personen wird unter den Multiple-Baseline-Designs bei weitem am häufigsten eingesetzt, weil es meist am einfachsten zu realisieren ist. In Abbildung 7 ist ein fiktives Multiple-Baseline-Design über Probandinnen und Probanden dargestellt.

Eine Grundschullehrerin möchte bei drei Schülern störendes Unterrichtsverhalten durch ein Response-Cost System verbessern, bei dem erwünschtes Verhalten verstärkt und grob regelverletzendes Verhalten durch den Entzug von Verstärkern bestraft wird. Sie plant eine Baselinephase von zwei Wochen und einen zeitlich versetzten Beginn der Intervention um jeweils eine Woche. Schüler A beginnt also mit der In-

tervention nach zwei, Schüler B nach drei und Schüler C nach vier Wochen. Eine hierfür geschulte Praktikantin registriert die Häufigkeit des störenden Verhaltens bei allen drei Schülern an jedem Schultag in den mittleren 25 Minuten der vierten Unterrichtsstunde.

Multiple-Baseline-Designs haben das Ziel, zwischenzeitliches Geschehen (History) und Reifung (Maturation) als mögliche Alternativerklärungen auszuschließen. Wenn, wie in Abbildung 7 zu sehen, bei zeitversetztem Interventionsbeginn das beobachtete Verhalten unter Baselinebedingungen relativ konstant bleibt und sich das Zielverhalten jeweils in kurzem zeitlichen Abstand zum Interventionsbeginn verändert, ist es extrem unwahrscheinlich, dass die Ursache hierfür in interventionsunabhängigen Faktoren liegt. Wären beispielsweise bei Schüler A externe schulbedingte Einflüsse unabhängig von der Intervention für eine Verhaltensverbesserung verantwortlich, hätte sich das auch in den Base-

*Tabelle 3: Zusammenfassung ABAB-Plan.*

<b>Kurzbeschreibung des Designs</b>	Auf eine Baseline (A1) folgt eine Intervention (B1), diese wird wieder zurückgenommen (A2) und setzt dann erneut ein (B2).
<b>Vorteile</b>	Sehr hohe interne Validität durch effektive Kontrolle von History und Maturation.
<b>Nachteile</b>	Praktische und ethische Probleme durch das Aussetzen der Intervention in der zweiten Baseline.
<b>Voraussetzungen</b>	Der Interventionseffekt muss umkehrbar sein.
<b>Bewertung</b>	Der ABAB-Plan ist ein methodisch anspruchsvoller Plan und erlaubt es, die Wirksamkeit einer Intervention zweifelsfrei nachzuweisen, wenn sich das Verhalten in den B-Phasen jeweils in die erwünschte Richtung verändert und in der zweiten Baselinephase hin zum Ausgangsverhalten. Das Aussetzen der Intervention in der A2-Phase ist allerdings problematisch (ethisch, praktisch, Complianceprobleme) und deshalb nicht immer durchführbar. Ist mit starken Transfereffekten von der ersten Interventionsphase zur nachfolgenden Baseline zu rechnen, ist ein Umkehrplan ungeeignet.

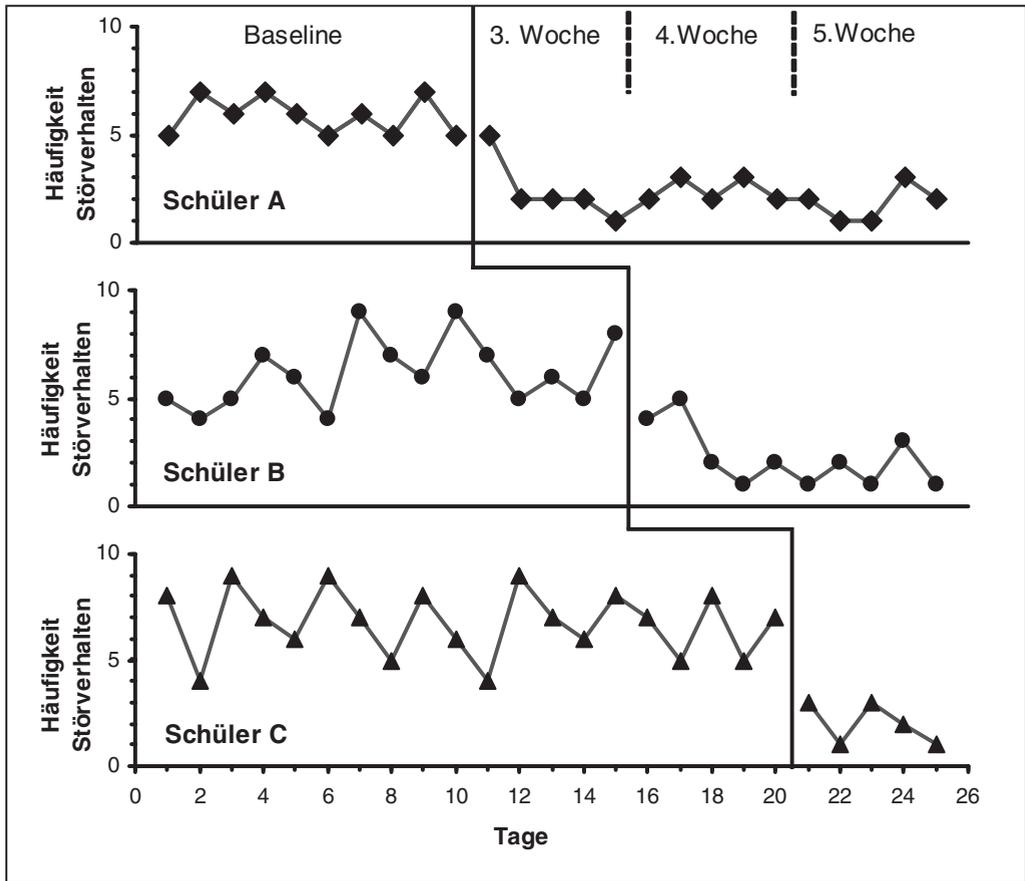


Abbildung 7: Fiktives Beispiel für ein Multiple-Baseline-Design über Personen.

lines von Schüler B und C zeigen müssen. Häusliche Veränderungen bei den einzelnen Schülern sind als Alternativerklärung gleichermaßen unplausibel. Denn dann hätten diese bei den einzelnen Kindern immer zeitgleich mit dem Beginn des Treatments eintreten müssen.

Multiple-Baseline-Designs führen im Vergleich zum einfachen AB-Design zu einer deutlichen Erhöhung der internen Validität. Ein multipler Grundratenversuchsplan über Personen ist dann besonders aussagekräftig, wenn die Probandinnen oder Probanden eine große Überschneidung in ihren sozialen Kontexten aufweisen (im Beispiel in Abbildung 7 die gleiche Klasse), weil

dann Einflüsse zwischenzeitlichen Geschehens (History) besonders gut kontrollierbar sind. Viele Beispiele in der Literatur, die als multiple Baseline-Designs über Personen bezeichnet werden, weisen dieses Kriterium nicht auf. Sie stellen daher Replikationen eines AB-Plans dar.

Die interne Validität eines Multiple-Baseline-Designs ist umso höher, je mehr unabhängige Baselines erfasst werden. Grundsätzlich genügen zwei Baselines, um von einem Multiple-Baseline-Design sprechen zu können und die interne Validität eines AB-Plans deutlich zu erhöhen. In den Standardwerken zu Einzelfalldesigns (Barlow et

al., 2009; Kazdin, 2011) werden mindestens drei Baselineerhebungen empfohlen.

### **Das Multiple-Baseline-Design über Verhaltensweisen**

Bei einem Multiple-Baseline-Design über Verhaltensweisen wird eine ähnliche Intervention bei derselben Person nacheinander auf unterschiedliche Verhaltensweisen oder Problembereiche angewandt. Dies können z.B. unterschiedliche Fertigkeiten sein, die im Rahmen eines umfassenderen Programms erlernt werden sollen. Wichtig dabei ist, dass man die Erfassung der Ausprägung jeder Verhaltensweise von Anfang an separat vornimmt und dass sich Verbesserungen jeweils genau dann zeigen, wenn eine Intervention in Bezug auf die jeweilige Verhaltensweise erfolgt.

Bei einem sozialen Kompetenztraining werden z.B. nacheinander die Bereiche „eigene Wünsche verbalisieren“, „Kritik äußern“, „Kritik annehmen“, „Selbstbehauptung“ und „Sensibilisierung für eigene Gefühle“ durch spezifische Trainings-Bausteine geübt. Die Erfassung der Ausprägung jeder dieser Fertigkeiten erfolgt von Anfang an kontinuierlich. Eine Änderung in der jeweiligen Fertigkeit in zeitlicher Abhängigkeit vom jeweiligen Trainingsbaustein spricht für die spezifische Wirksamkeit der einzelnen Interventionen.

### **Das Multiple-Baseline-Design über Situationen**

Bei einem Multiple-Baseline-Design über Situationen wird die gleiche Intervention zeitversetzt unter unterschiedlichen Umständen eingeführt. Im Schulkontext können die diversen Situationen z.B. unterschiedliche Unterrichtsfächer, unterschiedliche Lehrkräfte oder unterschiedliche Umgebungen (wie Schule, Hausaufgabenbetreuung oder häusliches Umfeld) sein. Wie bei den Multiple-Baseline-Designs über Perso-

nen und Verhaltensweisen ist es auch hier für den Nachweis der Wirksamkeit der Intervention wichtig, dass sich die Wirkung jeweils nur in der Situation zeigt, in der die Intervention eingeführt wurde, während die Werte der AV in den anderen Situationen, für die noch Baselinebedingungen gelten, auf Baselineniveau bleiben.

### **Literaturbeispiel für ein Multiple-Baseline-Design**

Ruiz-Olivares, Pino und Herruzo (2010) untersuchten in einem Multiple-Baseline-Design über Situationen die Wirksamkeit des GoodBehavior Game, das hierzulande unter der Bezeichnung „KlasseKinderspiel“ (Hillenbrand & Pütz, 2008) bekannt ist. Bei diesem Programm können die Kinder in Teams durch regelkonformes Verhalten Belohnungen erhalten, um störendes Verhalten der gesamten Klasse in einer normalen Grundschulklasse zu reduzieren. Ein weiterer Trainingsbaustein, in dem die Mädchen und Jungen die Regeln und das Einhalten dieser Regeln verbalisieren sollten (Say-Do-Correspondence Training), hatte darüber hinaus das Ziel, die Effekte der Intervention über einen längeren Zeitraum aufrechtzuerhalten.

Das GoodBehavior Game wurde zu Beginn der Intervention mehrmals an jedem Schultag für jeweils zehn Minuten gespielt. In einem Multiple-Baseline-Design über zwei Situationen, nämlich „Unterricht beim Klassenlehrer“ und „Unterricht bei einer Fachlehrerin“, wurde das Störverhalten der Kinder zunächst an sechs Tagen ohne Intervention von zwei externen Beobachtern erfasst (Baseline; vgl. Abbildung 8). Danach führte der Klassenlehrer das Spiel ein. Nach sechs Spieltagen bei dem Klassenlehrer wurde das Spiel auch in der zweiten Situation, nämlich dem Unterricht bei einer Fachlehrerin, eingeführt.

Abbildung 8 zeigt, dass sich das Störverhalten der Klasse direkt mit Beginn der In-

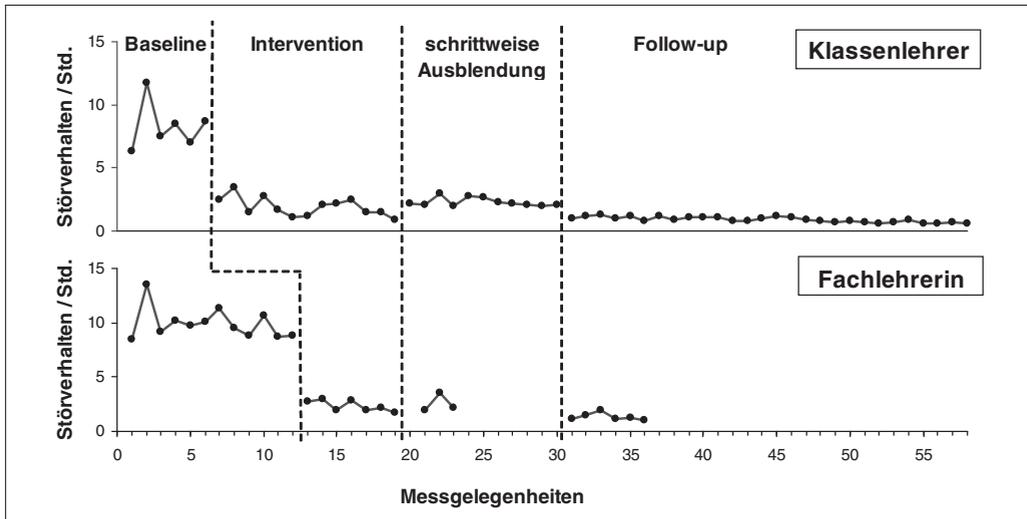


Abbildung 8: Literaturbeispiel für ein Multiple-Baseline-Design über Situationen zur Wirksamkeit des GoodBehavior Game in einer andalusischen Grundschulklasse (modifiziert nach Ruiz-Olivares et al., 2010, S. 1054) (die unvollständige Kurve der Fachlehrerin entstand durch Krankheit).

tervention beim Klassenlehrer sehr schnell und deutlich in die erwünschte Richtung veränderte, während es bei der Fachlehrerin weiterhin auf hohem Niveau blieb und sich erst mit Einführen der Maßnahme durch die zweite Lehrerin auch dort schnell und deutlich änderte. Nach einer Interventionsdauer von 13 Tagen (Klassenlehrer) bzw. 7 Tagen (Fachlehrerin) wurde die Maßnahme graduell ausgeblendet, indem die Spielphasen von ihrer Dauer her immer länger gestaltet und die materiellen Verstärker schrittweise abgebaut wurden (vgl. Abbildung 8, Ausblendung). Auch in einer mehrmonatigen Follow-up-Phase, in der keine Intervention mehr stattfand, blieb das Störverhalten der Kinder stabil auf niedrigem Niveau.

Im Gegensatz zu den bisher präsentierten Beispielen besteht in der hier vorgestellten Studie der „Einzelfall“ nicht aus einer einzelnen Person, sondern aus der ganzen Klasse. Es handelt sich dennoch um eine Einzelfalluntersuchung, weil das Verhalten der Klasse als Ganzes beobachtet wird und die AV, nämlich die Häufigkeit störender

Verhaltensweisen, für die gesamte Klasse (und nicht für einzelne Kinder) aufgezeichnet wurde.

Kritisch lässt sich zu dieser Untersuchung anmerken, dass das Verhalten hier nur in zwei unterschiedlichen Situationen (Unterricht beim Klassenlehrer und bei einer Fachlehrerin) beobachtet wurde; die Datenerfassung in drei oder mehr Situationen erlaubt generell eindeutigere Schlussfolgerungen. Allerdings ließ sich in dieser Studie die Wirksamkeit der Intervention auf die AV aufgrund der engen zeitlichen Korrespondenz von Interventionen und Verhaltensänderungen trotzdem demonstrieren.

### Der Kriterien-Veränderungs-Versuchsplan

Bei einem Kriterien-Veränderungs-Versuchsplan werden im Laufe der Intervention unterschiedliche Kriterien spezifiziert, die in jeder einzelnen Phase zu erreichen sind. Der Anspruch dieser Kriterien steigt mit der Zeit an. Am Schluss sollte dann

das letztendliche Behandlungsziel erreicht sein. Auch durch diesen Ablauf kann mehrfach gezeigt werden, dass eine Veränderung der UV eine Veränderung der AV zur Folge hat.

Abbildung 9 verdeutlicht den Ablauf eines Kriterien-Veränderungs-Versuchsplans am Beispiel eines Kindes, das lernen soll, längere Zeit konzentriert an seinen Hausaufgaben zu arbeiten. Als AV registriert die Mutter mit einer Stoppuhr jeden Tag die Zeit vom Beginn der Hausaufgabenbearbeitung bis zur ersten Unterbrechung. Eine Baselineerhebung an acht aufeinanderfolgenden Tagen zeigt, dass das Kind nur zwei bis vier Minuten ohne Unterbrechung bei der Sache bleibt. Da die Werte der Baseline sehr niedrig sind, wird das erste Kriterium auf 5 Minuten festgelegt. Ein Kriterien-Veränderungs-Versuchsplan beinhaltet häufig einen Verstärkerplan. Eine Verstärkung ist stets dann zu vergeben, wenn das Kriterium erreicht oder übertroffen wurde. In diesem Beispiel bekommt das Kind immer

dann eine Münze, wenn es das Kriterium an einem Tag erreicht hat. Später kann es die Münzen dann gegen kleine Belohnungen eintauschen. Um eine hohe Motivation des Kindes sicherzustellen und es nicht zu überfordern, werden nicht allzu große Änderungen im Kriterium von Phase zu Phase festgelegt. Ein bestimmtes Kriterium bleibt hierbei immer eine Woche bzw. vier Hausaufgabentage lang das jeweilige Maß der Dinge. Die letzte Phase dauert zwei Wochen, um die Stabilität prüfen zu können.

Abbildung 9 verdeutlicht auch eine besonders aussagekräftige Variante eines Kriterien-Veränderungs-Versuchsplans, nämlich die Einführung einer Mini-Umkehrphase, in der ein schon erreichtes Kriterium wieder gelockert wird. Im Beispiel wurde das Kriterium von 15 Minuten konzentrierter Arbeit in der entsprechenden Phase dreimal erreicht bzw. übertroffen. Danach wird das Kriterium wieder auf 10 Minuten herabgesetzt, um noch deutlicher die kausale Beziehung zwischen Kriterium

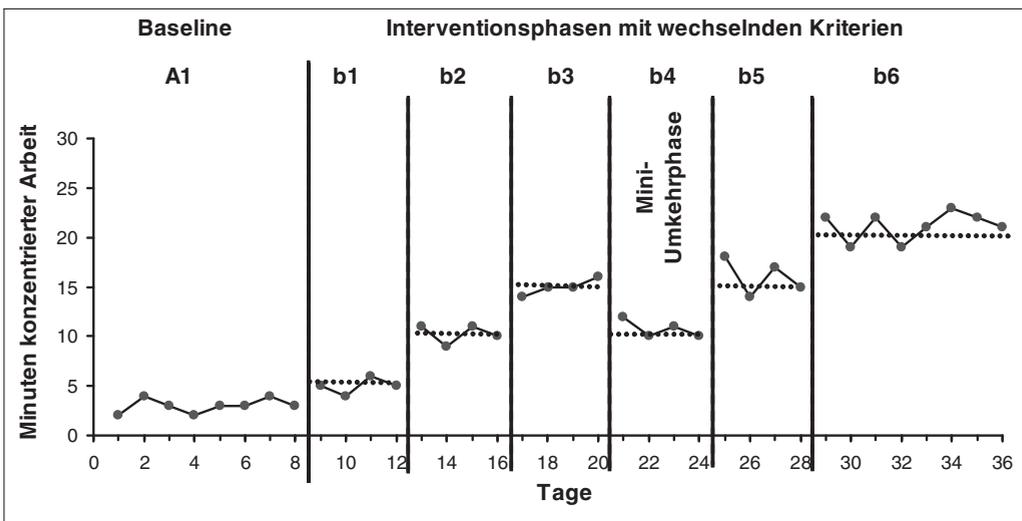


Abbildung 9: Ablauf eines Kriterien-Veränderungs-Versuchsplans zur Verbesserung des Arbeitsverhaltens (die gestrichelten horizontalen Linien markieren die jeweiligen Kriterien für die einzelnen Phasen, die gestrichelten vertikalen Linien die einzelnen Phasen; bei b4 wurde das Kriterium im Rahmen einer Mini-Umkehrphase wieder herabgesetzt).

Tabelle 4: Zusammenfassung *Multiple-Baseline-Design*.

<b>Kurzbeschreibung des Designs</b>	Ein AB-Plan wird entweder beim selben Probanden über verschiedene Situationen oder Verhaltensweisen oder bei verschiedenen, ähnlichen Probanden zeitversetzt durchgeführt.
<b>Vorteile</b>	Hohe interne Validität, keine Umkehrphase nötig.
<b>Nachteile</b>	Zeitaufwändig, lange Baselinemessungen erforderlich.
<b>Voraussetzungen</b>	Die Intervention muss in unterschiedlichen Situationen, für verschiedene Verhaltensweisen oder Personen zeitversetzt durchführbar sein. Die Wirkung für die verschiedenen Situationen, Verhaltensweisen oder Personen muss unabhängig voneinander sein.
<b>Bewertung</b>	Ein Multiple-Baseline-Design vermeidet die mit den Umkehrplänen verbundenen Nachteile des Aussetzens einer erfolgreichen Intervention. Es ist dann vorteilhaft, wenn sich verschiedene Verhaltensweisen oder Verhaltensweisen in unterschiedlichen Situationen unabhängig voneinander, aber in ähnlicher Weise behandeln lassen oder eine Intervention zeitversetzt bei mehreren Personen durchgeführt werden kann, die gleichen Störvariablen ausgesetzt sind und sich hinsichtlich wesentlicher Merkmale ähneln.

und Verhalten dokumentieren zu können. In dieser Mini-Umkehrphase zeigt sich, dass das Verhalten eng dem Kriterium folgt: Das Kind arbeitet nur 10–12 Minuten konzentriert, obwohl es zuvor schon 15 Minuten „geschafft“ hatte. Im weiteren Verlauf der Untersuchung wird das Kriterium wiederum schrittweise auf 15 und 20 Minuten erhöht. Das Kind erreicht diese Ziele ohne Probleme.

Folgt das Verhalten eng den jeweils sich ändernden Kriterien, kann von einer kausalen Wirkung der Intervention auf das Zielverhalten ausgegangen werden. Dies zeigt sich in der grafischen Darstellung durch einen treppenförmigen Verlauf. Das Verhalten soll also keine allmähliche und gleichmäßige Veränderung über die Zeit zeigen, denn dies könnte auch ein Trend sein, der durch andere Ursachen bedingt

ist. Vielmehr soll es den wechselnden Kriterien in Stufen folgen. Dies kann in der Regel dadurch erreicht werden, indem die Unterschiede zwischen aufeinanderfolgenden Kriterien nicht zu klein ausfallen und die einzelnen Phasen nicht zu kurz gestaffelt sind.

Ein Kriterien-Veränderungs-Versuchsplan eignet sich also besonders dann, wenn das Verhalten graduell verändert werden soll, was bei vielen Problemen sinnvoll ist. Sein Vorteil liegt darin, dass keine Umkehrphasen nötig sind, in denen eine Aussetzung der Intervention erfolgt. Außerdem müssen weder verschiedene Verhaltensweisen noch gleiche Verhaltensweisen in unterschiedlichen Situationen behandelt werden.

Wichtige Entscheidungen bei der Planung eines Kriterien-Veränderungs-Designs betreffen die Dauer der jeweiligen Phasen,

die Größe der Unterschiede zwischen den wechselnden Kriterien und die Anzahl der Phasen. Dabei müssen die Abstände zwischen den Kriterien und die Dauer der einzelnen Phasen nicht gleich sein, sondern sollten so gewählt werden, dass eine möglichst eindeutige Interpretation möglich ist (z. B. ist das jeweils nächste Kriterium erst dann einzuführen, wenn sich eine gewisse Stabilität bezüglich des vorigen Kriteriums eingestellt hat).

Eine erhebliche Stärkung der internen Validität ergibt sich, wenn Mini-Umkehrphasen in den Versuchsplan integriert werden, in denen eine Lockerung des Kriteriums erfolgt (siehe Beispiel in Abbildung 9). Wenn sich in diesen Umkehrphasen das Verhalten wieder hin zur unerwünschten Richtung ändert, ist dies ein besonders eindeutiger Beleg dafür, dass die Intervention wirkt und die Verhaltensänderung nicht auf einen ein allgemeiner Trend zurückführbar ist, der nichts mit der Intervention zu tun hat.

### ***Varianten des Kriterien-Veränderungs-Versuchsplans***

Statt als Kriterium einen einzelnen Wert anzugeben, kann auch ein Bereich spezifiziert werden, innerhalb dessen sich der Wert des Kriteriums in den einzelnen Phasen bewegen soll. Dies ist dann sinnvoll, wenn eine zu starke Veränderung des Ausgangsverhaltens negative Konsequenzen haben könnte (z. B. zu plötzliche Steigerung der Kalorienaufnahme bei Magersucht).

### ***Literaturbeispiel für einen Kriterien-Veränderungs-Versuchsplan***

Singh et al. (2011) untersuchten die Wirksamkeit von Achtsamkeitsbasierter Stressreduktion (einem Verfahren, das Elemente von Yoga und Meditation beinhaltet) zur Raucherentwöhnung. Proband war ein 31-jähriger Mann mit leichter intellektueller Minderbegabung und diversen Verhaltens-

auffälligkeiten, der seit 17 Jahren rauchte. In den vergangenen sechs Jahren hatte er schon erfolglos verschiedene Interventionen zur Raucherentwöhnung durchlaufen. Zu Therapiebeginn rauchte er jeden Tag 12 Zigaretten, weil dies das Maximum war, das man ihm in seiner Wohngruppe zugestand. Als Kriterium wurde festgelegt, dass er jeweils eine Zigarette weniger rauchen sollte als in der vorhergehenden Phase (beginnend mit elf Zigaretten/Tag) und dass das jeweilige Kriterium an drei aufeinanderfolgenden Tagen erreicht werden musste, um zur nächsten Phase überzugehen. Mit Hilfe des Meditationsverfahrens lernte er, seinen Zigarettenkonsum sukzessive um jeweils eine Zigarette zu verringern (siehe Abbildung 10). Nach 79 Interventionstagen (90. Tag der Aufzeichnung) hatte er sein Ziel, nämlich gar nicht mehr zu rauchen, erreicht. Auch bei Nacherhebungen nach 12 Monaten und drei Jahren wurde kein Zigarettenkonsum vom Probanden berichtet und diese Angaben wurden von seinen Betreuern und den Nachbarn bestätigt.

### ***Der alternierende Versuchsplan***

Ein alternierender Versuchsplan ist dann sinnvoll, wenn es zu überprüfen gilt, welche von zwei Interventionen stärkere oder schnellere Wirksamkeit zeigt. Dazu werden diese Interventionen in schnellem (z. B. im täglichen) Wechsel durchgeführt. Es resultieren dann zwei getrennte Verläufe für die beiden Interventionen, die vergleichbar sind (vgl. Abbildung 11). Unterschiede der beiden Kurven lassen darauf schließen, welche der beiden Interventionen erfolgreicher ist oder welche schneller wirkt.

Im Beispiel in Abbildung 11 soll überprüft werden, ob Lob oder Süßigkeiten als Verstärker für Unterrichtsbeteiligung wirksamer sind. Nach einer (relativ kurzen) Baselinephase von vier Tagen, in der eine Registrierung der Häufigkeit von Wortmeldungen im Unterricht erfolgt, beginnt die

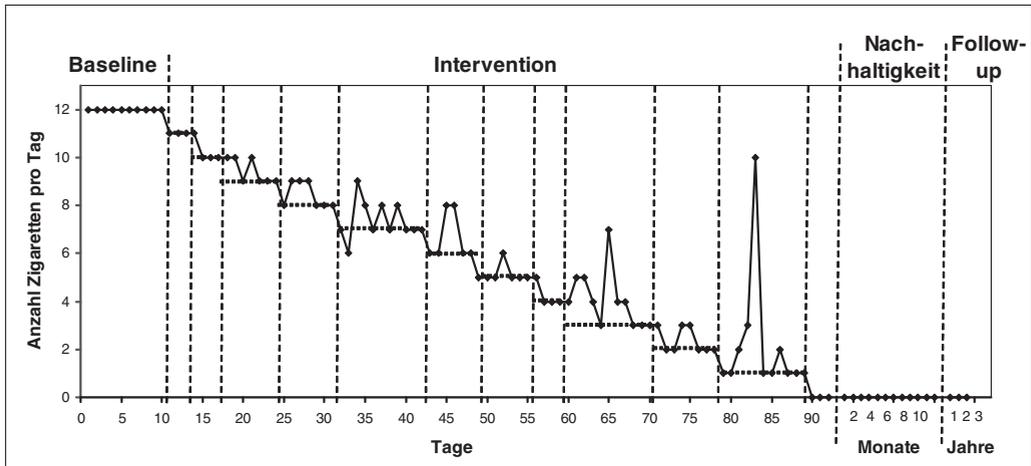


Abbildung 10: Literaturbeispiel für einen Kriterien-Veränderungs-Versuchsplan zur Raucherentwöhnung (modifiziert nach Singh et al., 2011, S. 1184) (die gepunkteten horizontalen Linien markieren die jeweiligen Kriterien für die einzelnen Phasen, die gestrichelten vertikalen Linien die einzelnen Phasen).

erste Intervention. Jede der beiden Interventionen wird an jeweils acht Tagen durchgeführt, wobei sich die beiden Maßnahmen in zufälliger Reihenfolge abwechseln, eine Intervention aber höchstens an zwei aufeinanderfolgenden Tagen stattfinden soll. Die Daten dieses Beispiels zeigen eine schnelle und sehr deutliche Wirkung von Intervention 1 (Lob), aber keine oder nur geringe Wirksamkeit von Intervention 2 (Süßigkeiten).

Typisch und wichtig für einen alternierenden Versuchsplan ist, dass die Reihenfolge der Interventionen zufällig festgelegt wird und kein regelmäßiger Wechsel erfolgt (z. B. nicht: ABABABAB etc.). Dies dient dazu, alle möglichen Störeinflüsse möglichst konstant zu halten bzw. zufällig über die Bedingungen zu verteilen. Wesentlich ist außerdem, dass mit der randomisierten Einteilung der Reihenfolge auch alle denkbaren weiteren Einflüsse zufällig und möglichst gleichmäßig auf die beiden Interventionen verteilt werden (z. B. die Tageszeit, die therapierende Person, der Raum etc.). Es ist also zu vermeiden, dass bestimmte

Rahmenbedingungen immer oder gehäuft mit einer der beiden Interventionen auftreten (Konfundierung).

Die Reihenfolge der beiden Interventionen wird in der Regel aber nicht völlig zufällig festgelegt, sondern man führt bestimmte Restriktionen ein (z. B. dass jede Intervention gleich häufig vorkommen soll und dass nicht mehr als zweimal oder dreimal nacheinander dieselbe Maßnahme durchgeführt wird). Ansonsten wäre die Anzahl an Beobachtungen unter einer Intervention oder die Anzahl an Wechseln zu gering.

Wesentlich für die Anwendbarkeit eines alternierenden Versuchsplans ist, dass Behandlungseffekte sehr schnell mit Einsetzen und Wechsel der Intervention sichtbar werden und ebenso schnell wieder aufhören. Nur wenn es geringe Transfers von einer Phase zur nächsten gibt, lässt sich ausschließen, dass das Verhalten während einer Intervention auf Effekte der vorherigen Intervention zurückzuführen ist. Problematisch bei einem alternierenden Versuchsplan sind mögliche Reihenfolge- oder

Tabelle 5: Zusammenfassung Kriterien-Veränderungs-Versuchsplan.

<b>Kurzbeschreibung des Designs</b>	In mehreren Phasen werden unterschiedliche Kriterien spezifiziert, die sich sukzessive dem eigentlichen Interventionsziel annähern.
<b>Vorteile</b>	Hohe interne Validität, keine Umkehrphase nötig, graduelle Verhaltensänderung ist für viele Problembereiche angemessen.
<b>Nachteile</b>	Relativlange Interventionsphase.
<b>Voraussetzungen</b>	Das Interventionsziel muss abstufbar sein.
<b>Bewertung</b>	Wenn sich das Interventionsziel in Teilschritte unterteilen lässt, stellt der Kriterien-Veränderungs-Versuchsplan eine praktisch sinnvolle und intern valide Methode dar, um einerseits dem Interventionsziel schrittweise näher zu kommen und andererseits mehrfach belegen zu können, dass auf eine Veränderung der UV eine Änderung in der AV folgt.

Übertragungseffekte. Dabei geht es um die Frage, ob die Wirkung eines Treatments auch dann eingetreten wäre, wenn die andere Behandlung nicht vorausgegangen wäre.

Ein großer Vorteil eines alternierenden Versuchsplans besteht darin, dass eine Ba-

seline nicht unbedingt erforderlich ist und somit direkt mit einer Intervention begonnen werden kann, wenn es primär um den Vergleich zweier Maßnahmen geht und nicht hauptsächlich darum, ob eine Behandlung überhaupt wirksam ist. In Fällen, in denen eine schnelle Änderung im Ziel-

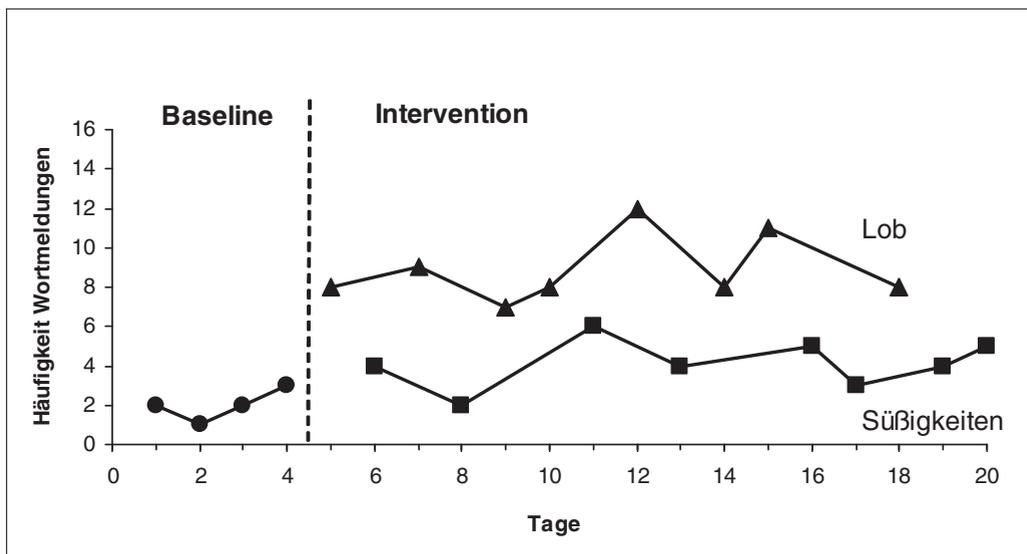


Abbildung 11: Alternierender Versuchsplan mit Baseline zum Vergleich der Wirksamkeit zweier Interventionen (Lob vs. Süßigkeiten) zur Steigerung der Mitarbeit im Unterricht.

verhalten unter ethischen Gesichtspunkten besonders wichtig ist, kann dies ein großer Pluspunkt sein. Eine vorangehende Baseline erhöht allerdings die Aussagekraft der Untersuchung deutlich, da dann auch ein Hinweis vorliegt, ob jede der Interventionen überhaupt wirksam ist.

Ein weiterer Vorteil dieses Designs liegt darin, dass Datentrends die Interpretation der Ergebnisse nicht erschweren, da diese in der Regel auf beide Interventionen gleichermaßen wirken. Generell erfolgt bei einem alternierenden Versuchsplan eine optimale Kontrolle von Störvariablen dadurch, dass bei gelungener Ausbalancierung der Reihenfolge) davon auszugehen ist, dass alle externen Einflüsse gleichermaßen auf beide Interventionen wirken und der Vergleich durch Störvariablen nicht verfälscht werden kann.

### **Varianten des alternierenden Versuchsplans**

Alternierende Versuchspläne können auch ohne eine Baseline konzipiert werden. Es ist außerdem möglich, mehr als zwei Treatments einander gegenüberzustellen. Eine weitere Abwandlung vom üblichen Vorgehen besteht darin, statt zweier Interventionen eine Intervention mit der Baselinebedingung oder mit einer Placebomaßnahme zu vergleichen.

### **Literaturbeispiel für einen alternierenden Versuchsplan**

Price, Martella, Marchand-Martella und Cleanthous (2002) verglichen die Wirksamkeit zweier unterschiedlicher Feedbackvarianten auf die Rate unangemessener Verbalisierungen im Unterricht bei einem zehnjährigen Jungen mit ADHS (siehe Abbildung 12). Hierbei überprüften sie die

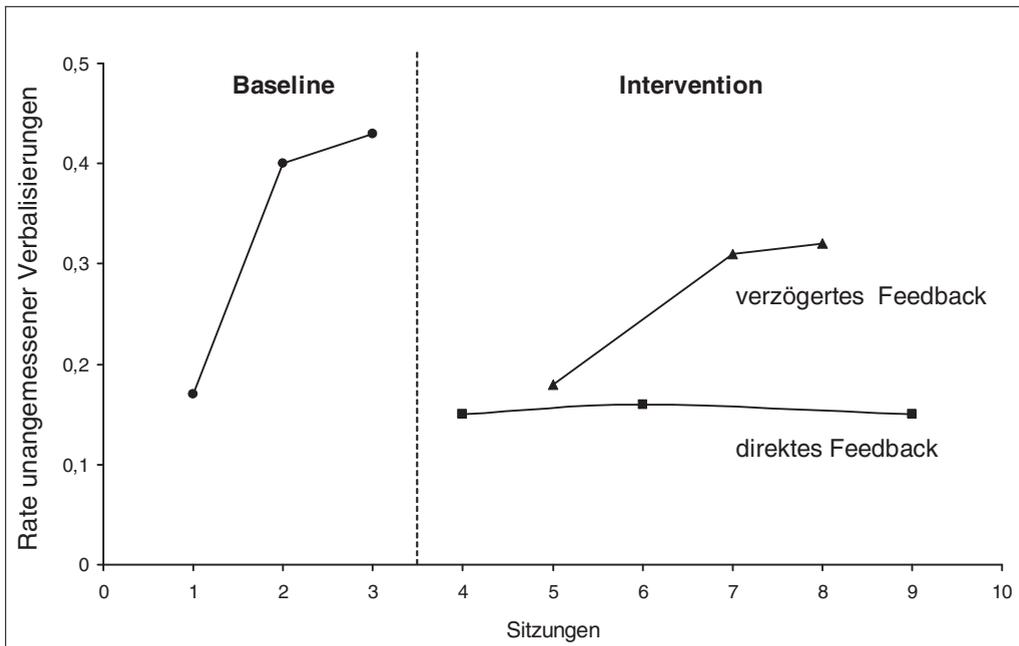


Abbildung 12: Literaturbeispiel für einen alternierenden Versuchsplan zum Vergleich der Wirksamkeit von direktem und verzögertem Feedback (nach Price et al., 2002, S. 166).

Hypothese, dass unangemessenes Verhalten durch direkte Rückmeldung (über Kopfhörer innerhalb von fünf Sekunden) effektiver reduziert wird als durch zeitverzögertes Feedback am Ende der jeweiligen Stunde. Im Rahmen eines alternierenden Versuchsplans wechselten beide Feedbackvarianten in der Interventionsphase tageweise in zufälliger Reihenfolge.

Abbildung 12 zeigt, dass der Anteil unangemessener Verbalisierungen bei direktem Feedback hypothesenkonform deutlich niedriger liegt als im Baselinemittel und bei verzögertem Feedback. Zusätzlich wird offenbar, dass ein Wechsel von direktem zu verzögertem Feedback mit einem Anstieg und ein Wechsel von verzögertem zu di-

rektem Feedback mit einem Absinken des problematischen Verhaltens einhergeht. Kritisch bleibt bei diesem Literaturbeispiel anzumerken, dass trotz der grundsätzlich sinnvollen Versuchsanordnung die Ergebnisse nicht eindeutig zu interpretieren sind. Dies liegt hauptsächlich daran, dass sowohl in der Baseline- als auch in der Interventionsphase die Anzahl an Datenpunkten zu gering ist. Darüber hinaus ist die Baseline nicht ausreichend stabil und weist Datenüberschneidungen mit der Interventionsphase auf. Bei so wenigen Messungen und der Gestalt der vorliegenden Baseline kann man nicht sicher sein, dass die Daten der Interventionsphase Ausdruck eines Interventionseffektes sind. Plausibel wäre hier

*Tabelle 6: Zusammenfassung Alternierender Versuchsplan.*

<b>Kurzbeschreibung des Designs</b>	Zwei unterschiedliche Interventionen werden in schnellem Wechsel durchgeführt und ihre Wirkung wird miteinander verglichen.
<b>Vorteile</b>	Baseline nicht unbedingt erforderlich; auch bei Trends in der Baseline einsetzbar; keine Umkehrphasen; ermöglicht den Vergleich zweier Interventionen; zeitsparend, da pro Phase nur ein Datenpunkt benötigt wird.
<b>Nachteile</b>	siehe Voraussetzungen.
<b>Voraussetzungen</b>	Schnell einsetzende und abklingende Interventionseffekte; keine Transfereffekte; zwei Interventionen müssen im schnellen Wechsel verabreicht werden können.
<b>Bewertung</b>	Soll die Effektivität von zwei Interventionen miteinander verglichen werden und können diese beiden Interventionen im schnellen Wechsel verabreicht werden, ohne dass sich ihre Wirkungen gegenseitig beeinflussen, stellt ein alternierender Versuchsplan eine elegante Methode dar, um zu überprüfen, welche von zwei Interventionen bei einer Person erfolgreich ist bzw. bei welcher der Erfolg stärker ist oder schneller eintritt. Dieser Versuchsplan ist einer der wenigen, der ohne Baselinephase auskommt, und ist deshalb auch geeignet, wenn eine Intervention schnell einsetzen muss. Die Voraussetzung des schnellen Wechsels beider Interventionen und ihrer Wirkungen lässt sich allerdings nur in wenigen Situationen erfüllen.

auch, dass sie eine zufällige Variation der Baselinedaten darstellen. Um den Effekt der Behandlung eindeutig abzusichern, wären mehr Messzeitpunkte in Baseline- und Interventionsphase notwendig gewesen.

### **Kombinationen der vorgestellten Versuchspläne**

Alle genannten Einzelfalldesigns lassen sich je nach praktischen Anforderungen kombinieren, um die Schwächen einzelner Designs zu überwinden und zu praktikablen und validen Versuchsanordnungen zu kommen (vgl. z. B. Kern, 1997).

### **Replikationen**

Das Ziel von Replikationen besteht darin, durch wiederholte Durchführung ähnlich angelegter Studien Ergebnisse abzusichern (Petermann, 1996). Sie dienen dem Nachweis der Reliabilität, der Stärkung der internen Validität und ermöglichen Generalisierungen. Durch ein Einzelfalldesign mit hoher interner Validität wird gezeigt, dass die durchgeführte Intervention tatsächlich für die eingetretene Veränderung in der AV bei der untersuchten Person, im jeweiligen Untersuchungszeitraum und im untersuchten Setting verantwortlich war. Darüber hinaus ist es aber auch von großem Interesse, ob sich die gefundenen Ergebnisse auf andere Personen, Therapeutinnen bzw. Therapeuten und Settings übertragen lassen. Hierbei geht es also um die Frage nach der Verallgemeinerbarkeit bzw. nach der externen Validität. Für die Generalisierung auf andere Personen, Behandlerinnen bzw. Behandler oder Situationen ist die Wiederholung des Einzelfallexperimentes eine notwendige Voraussetzung.

Wenn eine Intervention in einem bestimmten Setting (z. B. in der Schule) bei mehreren Personen mit ähnlichen Merkmalen (z. B. bei sechsjährigen Kindern mit ADHS) ein ähnliches Ergebnis zeigt, ist es

wahrscheinlich, dass sich dieses Ergebnis auch auf weitere Personen dieses Settings mit denselben Merkmalen generalisieren lässt. Eine Untersuchung unter möglichst identischen Umständen bezeichnet man als direkte Replikation (Sidman, 1960). Hierbei werden Personen- und Situationsmerkmale sowie Merkmale der intervenierenden Person (Therapeutin bzw. Therapeut oder Lehrkraft) konstant oder möglichst ähnlich gehalten. Ziel der direkten Replikation ist es zu zeigen, dass ein Interventionseffekt in einem bestimmten Setting wiederholbar ist. Jede erfolgreiche Replikation ist ein Beleg für die Reliabilität der Wirksamkeit, also dafür, dass das Ergebnis nicht durch spezifische Umstände der jeweiligen Untersuchungsdurchführung zustande gekommen ist. Jede erfolgreiche Replikation stärkt außerdem die interne Validität, da es mit steigender Anzahl an Replikationen zunehmend unwahrscheinlicher wird, dass Störvariablen eine Veränderung in der AV bewirkt haben (siehe auch Hussy & Jain, 2002).

Generalisierungen der Interventionseffekte auf andere Personengruppen (z. B. 12-jährige Kinder mit ADHS) oder andere Settings sollten nicht vorschnell erfolgen, sondern Gegenstand eigener Untersuchungen sein. Solche Studien bezeichnet man als systematische Replikation. Hierbei werden gezielt und planvoll wesentliche Merkmale der Untersuchung variiert (Personenmerkmale, Situationsmerkmale, Person und Eigenschaften der Behandlerinnen bzw. Behandler). Zeigt sich auch bei veränderten Bedingungen ein Erfolg der Intervention, so stellt dies eine zusätzliche Stärkung der internen Validität dar. Hauptziel einer systematischen Replikation ist es aber, die Verallgemeinerbarkeit der Ergebnisse zu erhöhen bzw. den Anwendungsbereich einer Maßnahme auszutesten. Wenn eine Intervention bei unterschiedlichen Personen, Störungsbildern, Settings und Behandlerinnen bzw. Behandlern erfolgreich ist, weist

die Interventionswirkung eine hohe Allgemeingültigkeit auf.

Erweist sich die Intervention bei Personen mit anderen Eigenschaften oder Störungsbildern oder in anderen Situationen als nicht erfolgreich, können die relevanten Moderatorvariablen für einen Interventionserfolg identifiziert werden, wenn die jeweiligen Gegebenheiten (Personen- oder Situationsmerkmale usw.) in den einzelnen Replikationsstudien differenziert erfasst und systematisch variiert werden.

## **Datenauswertung in Einzelfalldesigns**

Die Standardauswertungsmethode für experimentelle Einzelfalldesigns ist die so genannte visuelle Inspektion, die auf der grafischen Darstellung der erhobenen Daten beruht. Sie kann durch statistische Verfahren ergänzt werden, bei denen jedoch im Einzelfall zu überprüfen ist, ob die Voraussetzungen für deren sinnvollen Einsatz gegeben sind (siehe unten).

### **Visuelle Inspektion**

Wie die Beispiele dieses Artikels zeigen (Abbildungen 3–12), erfolgt die grafische Darstellung der Messwerte in Einzelfalluntersuchungen in der Regel als Liniendiagramm. Die Beobachtungszeitpunkte werden im Verlauf auf der Abszisse aufgetragen und die zugehörigen Werte der AV auf der Ordinate. Zusätzlich erfolgt eine Kennzeichnung der Baseline- und Interventionsphasen im Diagramm. Das Ziel der visuellen Inspektion besteht darin, mögliche Treatmenteffekte zu erkennen, indem Unterschiede in den Datenmustern zwischen Baseline- und Interventionsphasen analysiert werden. Der Vorteil dieses Vorgehens besteht darin, dass große Effekte direkt ins Auge springen. Dies ist insbesondere dann der Fall, wenn sich die Messwerte von Baseline- zu Interventionsphase von

einem Ende des Kontinuums zum anderen bewegen. Hierzu zählt beispielsweise ein starker Anstieg eines gewünschten Zielverhaltens, das unter Baselinebedingungen nie aufgetreten ist, oder das Verschwinden eines Problemverhaltens unter Interventionsbedingungen, das in der Baseline sehr häufig gezeigt wurde. Dermaßen deutliche Wirkungen („slam bang“ effects, vgl. Kazdin, 2011, S. 287) treten jedoch eher selten auf.

In Fällen, in denen die Ergebnislage nicht so eindeutig ist, werden verschiedene Kriterien herangezogen, anhand derer Interventionseffekte mithilfe visueller Inspektion beurteilbar sind. Diese Kriterien beziehen sich auf das Ausmaß der Veränderung (Veränderungen im Mittelwert und im Level), auf die Geschwindigkeit, mit der eine Veränderung eintritt (Latenz der Veränderung) und auf Veränderungen im Trend.

### **Veränderungen im Mittelwert und im Level**

Abbildung 13 (linke Grafik) veranschaulicht Mittelwertveränderungen (gestrichelte Linie) in einem ABAB-Plan bei nicht stabiler Baseline. Trotz überlappender Datenpunkte in Baseline- und Interventionsphasen zeigen sich deutliche Mittelwertunterschiede zwischen den Phasen. In Kombination mit klar sichtbaren Richtungswechseln im Trend bei Ein- und Aussetzen der Intervention lässt sich dieses Muster eindeutig als Interventionseffekt interpretieren. Die rechte Grafik in Abbildung 13 offenbart neben Veränderungen im Mittelwert auch Veränderungen im Level. Veränderungen im Level sind deutliche Sprünge bei Ein- bzw. Aussetzen der Intervention, die in der Abbildung durch Pfeile gekennzeichnet sind. Treten sie bei Phasenwechseln auf, stellen sie einen zusätzlichen Beleg für die Wirksamkeit einer Intervention dar. Dies ist insbesondere dann der Fall, wenn Veränderungen im Level bereits vor Beginn ei-

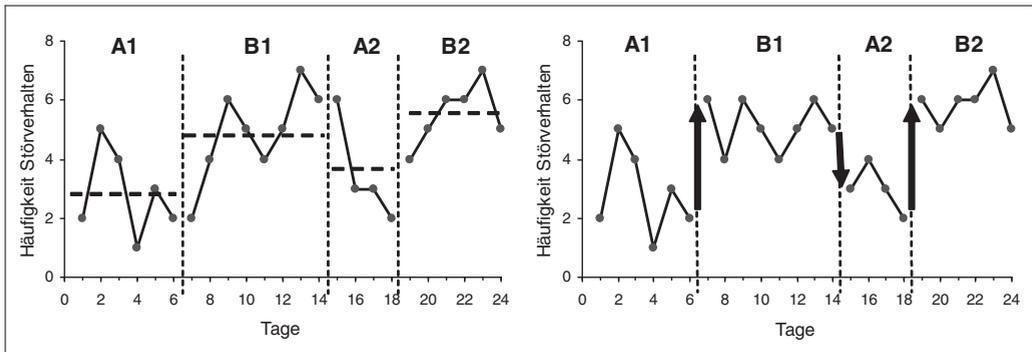


Abbildung 13: Beispiele für eindeutige Veränderungen zwischen einzelnen Phasen eines ABAB-Plans im Mittelwert und im Trend (links) sowie im Level (rechts).

ner Untersuchung als Hypothese formuliert worden sind. Eine solche Hypothese wäre beispielsweise bei Kindern mit ADHS sinnvoll, deren hyperaktiv/impulsives Verhalten durch eine medikamentöse Behandlung mit Methylphenidat reduziert werden soll. Da bekannt ist, dass bei optimaler Dosis das Wirkmaximum dieses Präparats bereits nach ca. zwei Stunden eintritt und der Wirkstoff innerhalb von 24 Stunden völlig abgebaut wird, erwartet man hier rasche Veränderungen bei Ein- und Absetzen der Medikation.

### Latenz und Veränderungen im Trend

In Abbildung 13 zeigen beide Grafiken zeitnahe Veränderungen nach Einsetzen und Ausblenden der Behandlung. Grundsätzlich spricht eine kurze Latenz zwischen Interventionsbeginn und Veränderungen des Verhaltens in die prognostizierte Richtung für einen Effekt des Treatments. Je größer die Latenz, desto wahrscheinlicher ist es, dass externe Faktoren für eine Veränderung verantwortlich sind. Diese Faustregel sollte jedoch nicht schematisch eingesetzt werden, da es viele Interventionen gibt, bei denen eine schnelle Wirkung nicht erwartet wird (z. B. bei der Behandlung von Depressionen mittels Antidepressiva oder bei Programmen zur Gewichtszunahme

bei Anorexie). Hier ist der Treatmenteffekt umso eindeutiger zu interpretieren, je präziser der Zeitpunkt einer Veränderung vorhergesagt werden kann.

Veränderungen im Trend sind ein weiterer Indikator für die Effektivität einer Maßnahme. Sie können wie in Abbildung 13 (links) in Form von schnellen Richtungswechseln zwischen Baseline- und Interventionsphasen auftreten und sind dann ein klarer Hinweis für eine Interventionswirkung. In vielen Fällen sind Veränderungen im Trend nicht so eindeutig wie in diesem Beispiel.

In Abbildung 14 ist der ansteigende Baselinetrend aufgrund der hohen Variabilität nicht direkt zu erkennen. Um den Trend sichtbar zu machen, hilft das Einzeichnen von Trendlinien für Baseline- und Interventionsphasen, die von aktuellen Tabellenkalkulationsprogrammen automatisch generiert werden können. Eine Verlängerung der Trendlinie der Baseline in die Interventionsphase kann als plausible Prognose des Verhaltens ohne Intervention dienen.

### Weitere Hilfen für die visuelle Inspektion: Blocking und Gleitmittelwerte

Die Berechnung von Mittelwerten über mehrere Datenpunkte kann eine Hilfe sein, um eine klarere Struktur in Daten mit sehr variablem Verlauf zu erkennen (Blocking;

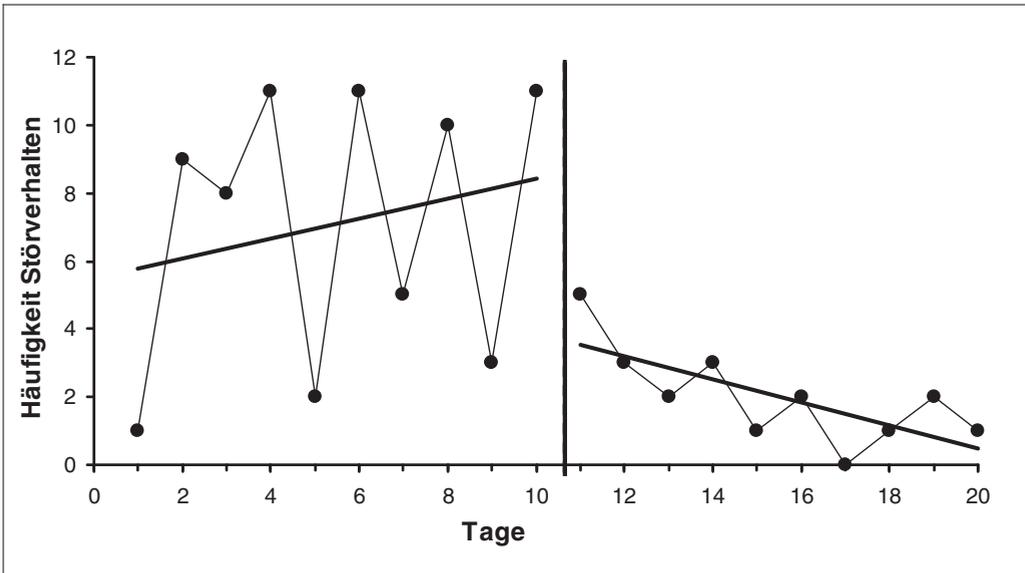


Abbildung 14: Variable Baseline mit Trendlinie für Baseline (leicht steigend) und Interventionsphase (fallend).

vgl. Kazdin, 2011). Sie hat aber den Nachteil, dass sich damit die Anzahl verfügbarer Datenpunkte stark reduziert (bei einem Mittelwert über z.B. vier Werte auf ein Viertel) und kommt deshalb nur bei einer ausreichenden Anzahl an Werten in Frage.

Dem gleichen Zweck dient die Berechnung und Darstellung eines gleitenden Mittelwertes (Moving Average), der mit einem Tabellenkalkulationsprogramm entweder automatisch generiert wird oder einfach selbst zu erstellen ist, indem der Mittelwert aus einem Wert und seinen „Nachbarwerten“ berechnet wird. Solche Gleitmittelwert-Kurven werden auch in anderen Bereichen (z. B. für Börsenkurse) ermittelt, um eine klarere Struktur in variablen Datenverläufen erkennen zu können. Dabei ist die Anzahl an Werten, die für die Darstellung und Beurteilung verlorengehen, in der Regel geringer als beim Blocking. Abbildung 15 zeigt den Verlauf einer durch Mittelung geglätteten Baseline.

### **Vor- und Nachteile visueller Inspektion**

Bei der visuellen Inspektion werden die Veränderungen im Mittelwert, im Level und im Trend unter Berücksichtigung der Latenz einer Veränderung zu einem komplexen Urteil über die Wirksamkeit der Intervention zusammengefasst. Hierbei ist auch die Berücksichtigung von Besonderheiten im Untersuchungsablauf möglich, die durch eine schematisch vorgenommene statistische Auswertung nicht einbezogen werden können. Auch die Konsistenz von Datenverläufen bei einer Wiederholung (z. B. im Rahmen eines ABAB-Plans oder eines Multiple-Baseline-Designs) lässt sich beurteilen. Ist sie vorhanden, spricht dies für die Wirksamkeit einer Intervention. Bei nicht eindeutigen Datenverläufen kann die Auswertung mittels visueller Inspektion aber auch Einschränkungen in der Objektivität und Reliabilität aufweisen. Verschiedene Auswerterinnen und Auswerter können aufgrund visueller Inspektionen zu unter-

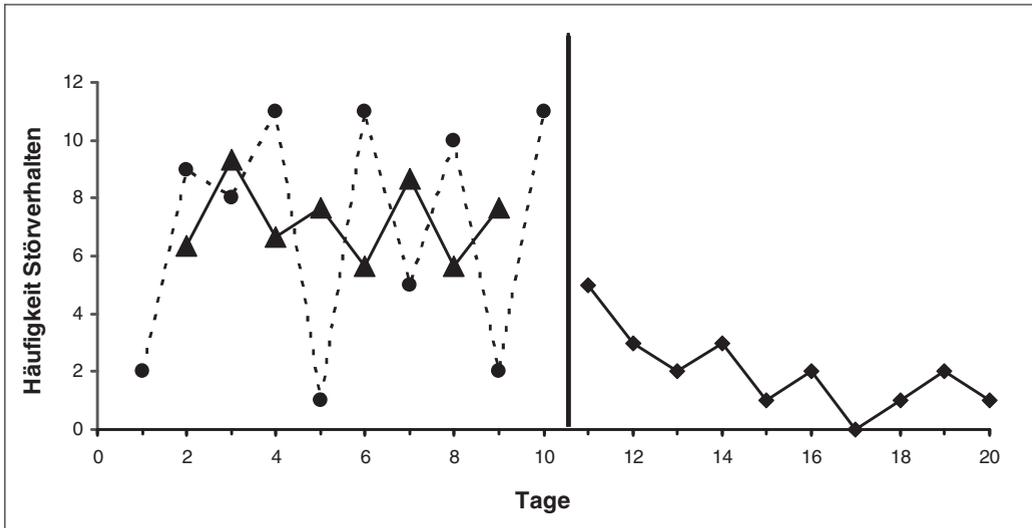


Abbildung 15: Variable Baseline mit Gleitmittelwertkurve über drei Werte. Die gestrichelte Linie ist die Originalbaseline, die durchgezogene Linie ist die Baseline mit gleitendem Mittelwert über jeweils drei Datenpunkte.

schiedlichen Einschätzungen bezüglich des Behandlungserfolges kommen. Die Datenanalyse ist für Urteilsfehler und Verzerrungstendenzen anfällig, da keine klaren Entscheidungsregeln existieren. So können Unterschiede zwischen Phasen übersehen werden. Des Weiteren ist es denkbar, vermeintliche Unterschiede zu erkennen, die objektiv gar nicht nachweisbar sind.

Vor diesem Hintergrund erscheint es sinnvoll, die visuelle Inspektion durch statistische Berechnungen zu ergänzen. Für Einzelfallstudien gibt es hierfür allerdings bisher noch keine Standardmethode. Die existierenden Verfahren haben alle auch Nachteile oder Einschränkungen, die in den statistischen Eigenschaften der Verfahren selbst, in den Voraussetzungen oder in der Komplexität der Verfahren liegen.

Statistische Datenanalyse kann dann besonders vorteilhaft sein, wenn die Verläufe nicht eindeutig sind oder wenn nur kleine Effekte vorliegen, die bei der visuellen Inspektion nicht eindeutig erkennbar sind. Ein weiterer Vorteil statistischer Größen

liegt darin, dass die Auswertungsobjektivität höher ist als bei einer Beurteilung aufgrund visueller Inspektion. Des Weiteren lassen sich statistische Werte leichter kommunizieren. Liegen mehrere Untersuchungen zum selben Themenbereich vor, können statistische Größen über mehrere Studien besser aggregiert werden, z. B. durch Berechnung von Effektgrößen. Als Effektgröße zieht man in Einzelfallstudien z. B. den durch die Phasen aufgeklärten Varianzanteil heran. Ein weiteres Maß stellt die Überlappung der Werte in Baseline und Interventionsphasen dar. Für beide Konzepte existieren mehrere Berechnungsformeln (siehe z. B. Parker et al., 2005 oder Parker, Vannest & Davies, 2011).

### **Einfache deskriptiv-statistische Auswertung: Prozentsatz nicht-überlappender Datenpunkte (PND)**

Ein einfaches deskriptiv-statistisches Maß für die Quantifizierung einer Interventionswirkung stellt der Prozentsatz nicht-

überlappender Datenpunkte (PND) dar, denn ein Treatmenteffekt ist umso größer und deutlicher, je weniger sich die Werte von Baseline und Behandlungsphase überschneiden. Der PND ist ein deskriptives Maß und somit an keine Annahmen inferenzstatistischer Verfahren wie Normalverteilung oder Unabhängigkeit der Werte geknüpft. Zur Berechnung des PND wird (bei beabsichtigtem Anstieg der Werte durch die Maßnahme) der größte Wert der Baseline bestimmt und ausgezählt, wie viele Datenpunkte der Interventionsphase größer als dieser Wert sind. Die Anzahl dieser nicht-überlappenden Datenpunkte wird durch die Anzahl aller Datenpunkte in der Interventionsphase geteilt und mit 100 multipliziert. Je höher dieser Prozentsatz, desto deutlicher und verlässlicher der Interventionseffekt (Scruggs, Mastropieri, Cook & Escobar, 1986). Der große Vorteil des PND ist seine einfache Berechnung und Interpretation. Er ist deshalb auch besonders verbreitet. Nachteilig an diesem Maß ist, dass die Grenze für den Überschneidungsbereich nur durch einen einzigen Wert der Baseline (den Höchsten oder je nach beabsichtigter Interventionsrichtung den Niedrigsten) gebildet wird und er deshalb nicht sehr reliabel ist. Parker et al. (2011) berichten über ähnliche Maße mit besseren statistischen Eigenschaften, um den Nicht-Überschneidungsbereich aufeinanderfolgender Phasen zu bestimmen. Der PND liefert dann keine sinnvollen Ergebnisse, wenn es einen Trend in den Daten gibt, die Baseline starke Schwankungen oder einzelne Ausreißer aufweist oder wenn die Werte in der Baseline Decken- oder Bodeneffekte zeigen.

Bei Vorliegen einer stabilen Baseline kann der PND die durch eine visuelle Analyse erhaltenen Informationen sinnvoll quantifizieren und dadurch objektiver und besser kommunizierbar machen. In der Studie von Degner und Nußbeck (2011), die im Kapitel zu den ABAB-Plänen vorgestellt wurde,

beträgt der PND für die in Abbildung 6 dargestellten Daten 91%. Dieser Wert spricht nach einer Klassifikation von Scruggs et al. (1986) für eine sehr effektive Intervention.

### **Probleme inferenzstatistischer Tests im Rahmen der Einzelfallforschung**

Die für die Gruppenforschung standardmäßig angewandten inferenzstatistischen Verfahren können bei Einzelfallplänen nicht eingesetzt werden, weil in der Regel wesentliche Voraussetzungen dieser Verfahren nicht erfüllt sind. Parametrische Verfahren (z. B. t-Test oder Varianzanalyse) stellen Bedingungen im Hinblick auf Verteilung der Werte oder Gleichheit der Varianzen, die oftmals nicht gegeben sind. Auch nonparametrische Verfahren (z. B. Randomisierungstests oder U-Test), die diese Annahmen nicht machen, haben als Voraussetzung die Unabhängigkeit der einzelnen Werte. Ist diese Voraussetzung nicht gegeben, sind die Ergebnisse inferenzstatistischer Tests verfälscht.

Die Unabhängigkeitsannahme wird in Einzelfalluntersuchungen häufig verletzt, und zwar dadurch, dass aufeinanderfolgende Werte in der Regel eine starke serielle Abhängigkeit aufweisen (d. h. Werte, die direkt aufeinander folgen, sind sich ähnlicher als Werte, die zeitlich weiter auseinander liegen). Die serielle Abhängigkeit entsteht u. a. dadurch, dass derselbe Fehler- oder Zufallseinfluss die AV über mehrere aufeinanderfolgende Messgelegenheiten hinweg beeinflusst. Die Zeitreihenanalyse (Box & Jenkins, 1976) ist ein Verfahren, das diesen Umstand berücksichtigt, aber sehr viele Messungen und weitergehende statistische Kenntnisse voraussetzt (eine einführende Darstellung findet sich z. B. bei Bortz & Döring, 2006). Sind ihre Voraussetzungen erfüllt, stellt die Zeitreihenanalyse eine sinnvolle Methode dar, Interventionseffekte in Einzelfalluntersuchungen auszuwerten und inferenzstatistisch abzusichern.

Zusammenfassend lässt sich feststellen, dass die Auswertung von Einzelfallstudien bisher meistens über die grafische Darstellung mit visueller Inspektion erfolgt. Dieses Vorgehen ist dann ausreichend und zufriedenstellend, wenn die Verläufe eindeutig sind (also wenn die Baseline stabil und trendfrei und der Interventionseffekt groß ist). Statistische Methoden zur Auswertung von Einzelfallstudien haben oft sehr strenge Voraussetzungen (z. B. benötigt die Zeitreihenanalyse sehr viele Datenpunkte), existieren in vielen Varianten oder haben wiederum eigene methodische Probleme. Deshalb haben sie sich zur Auswertung von Einzelfallplänen bisher noch nicht durchgesetzt, obwohl eine Erhöhung der Objektivität und Reliabilität der Datenauswertung wünschenswert wäre und obwohl sie die Akzeptanz von Einzelfalldesigns als wissenschaftlicher Methode steigern könnten. Generell gilt aber, dass sich die statistische Auswertung an die Untersuchungsdesigns anpassen muss und nicht umgekehrt. Vorrang hat immer ein Design, das zum vorliegenden Problembereich passt, das die Intervention optimal ergänzt und das eine möglichst hohe interne Validität aufweist: „No selected statistic is able to improve the design of the experiment“ (Barlow et al., 2009, S. 302).

## Abschließende Bemerkungen

Das Ziel des vorliegenden Beitrages bestand darin, in die Logik experimenteller Einzelfallforschung einzuführen und deren Anwendungsmöglichkeiten bei der Überprüfung von Interventionseffekten anhand von Beispielen aus dem (sonder-) pädagogischen und psychologischen Kontext zu verdeutlichen. Experimentelle Einzelfallforschung ist gekennzeichnet durch zwei grundlegende Charakteristika:

- wiederholtes Messen mindestens einer AV bei mindestens einer Versuchsperson unter

- Baseline- und Interventionsbedingungen (Variation der UV)

Die im Beitrag dargestellten Versuchspläne, die je nach Fragestellung und Untersuchungskontext flexibel einsetzbar sind, tragen auf unterschiedliche Weise dazu bei, interne Validität herzustellen. Interne Validität bildet die zentrale Voraussetzung für einen eindeutigen Beleg von Interventionseffekten. Experimentelle Einzelfallforschung kann eingesetzt werden, um

- die Wirksamkeit von Maßnahmen in Einzelfällen zu überprüfen,
- einen Evidenzbeleg für die Wirksamkeit eines Treatments über den Einzelfall hinaus zu erbringen oder
- die differentielle Analyse von Interventionseffekten zu ermöglichen.

Die Beispiele aus unserem Beitrag beschäftigten sich hauptsächlich mit der Überprüfung der Wirksamkeit von Behandlungen in Einzelfällen. „Wirkt eine Intervention zur Reduktion problematischen Verhaltens bei Schüler X?“ oder „Führt ein therapeutisches Vorgehen zu einer Reduktion der Angstsymptomatik bei Patient Y?“ sind klassische Fragestellungen, bei denen experimentelle Einzelfallforschung die Methode der Wahl ist.

Der Abschnitt über Replikationen von Einzelfalluntersuchungen zeigt, dass experimentelle Einzelfallforschung darüber hinaus auch das Potenzial hat, Evidenzbelege für die Wirksamkeit einer Intervention über den Einzelfall hinaus zu erbringen. Im Zusammenhang mit der Diskussion um evidenzbasierte Praxis (siehe z. B. Kazdin, 2008) ist dieses Potenzial von nicht zu unterschätzender Bedeutung. Viele möglicherweise sinnvolle Behandlungsansätze werden es u. U. nie schaffen, in einer randomisierten Kontrollgruppenuntersuchung evaluiert zu werden (z. B. weil für ein bestimmtes Störungsbild nicht genügend Patienten gleichzeitig zur Verfügung stehen). Hier liefert die systematische Replikation

von Einzelfällen einen alternativen Evidenznachweis.

Die differentielle Analyse von Treatmenteffekten ist wichtig, wenn die Interventionswirkung sehr heterogen ausfällt. Dies ist z. B. dann der Fall, wenn eine Subgruppe von einer Maßnahme profitiert, eine andere jedoch nicht. Hier können Einzelfallanalysen mit dazu beitragen, Variablen zu identifizieren, die für die Wirksamkeit einer Intervention förderlich bzw. hinderlich sind. Wenn experimentelle Einzelfalluntersuchungen durchgeführt werden, sollte die Untersucherin bzw. der Untersucher

- die dahinter stehende Logik verstanden haben,
- sicherstellen, dass die vorgesehenen Untersuchungsinstrumente Behandlungseffekte angemessen abbilden können,
- einen adäquaten Versuchsplan auswählen und
- vorab Überlegungen anstellen, wie die Untersuchung ausgewertet werden soll.

Dabei ist es wichtig zu beachten, dass – wie bei allen anderen Typen von Untersuchungen auch – die Qualität einer Studie nicht durch Verwendung einer bestimmten Methode per se sichergestellt wird. Stattdessen ist zu überprüfen, inwieweit der Zweck des experimentellen Vorgehens erfüllt ist. Dieser besteht darin, eindeutige kausale Schlüsse zu ermöglichen. Dazu sind immer die Qualität der Datenerhebung, der Versuchsplan und die Datenauswertung im Kontext der jeweiligen Untersuchungsgegebenheiten gemeinsam zu bewerten.

## Literatur

- Barlow, D.H., Nock, M.K. & Hersen, M. (2009). *Single case experimental designs: Strategies for studying behavior change*. Boston: Pearson.
- Bortz, J. & Döring, N. (2006). *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler*. Berlin: Springer.
- Box, G.E.P. & Jenkins, G.M. (1976). *Time series analysis: Forecasting and control*. San Francisco: Holden-Day.
- Degner, M. & Nußbeck, S. (2011). Wirksamkeit Strukturierter Arbeitssysteme zur Förderung der Selbstständigkeit von Kindern mit Autismus. *Empirische Sonderpädagogik*, 3, 51–74.
- Hillenbrand, C. & Pütz, K. (2008). *KlasseKinderSpiel: Spielerisch Verhaltensregeln lernen*. Hamburg: Edition Körber Stiftung.
- Hussy, W. & Jain, A. (2002). *Experimentelle Hypothesenprüfung in der Psychologie*. Göttingen: Hogrefe.
- Julius, H., Schlosser, R. & Goetze, H. (2000). *Kontrollierte Einzelfallstudien*. Göttingen: Hogrefe.
- Kazdin, A.E. (2008). Evidence-based treatment and practice: New opportunities to bridge clinical research and practice, enhance the knowledge base, and improve patient care. *American Psychologist*, 63, 146–159.
- Kazdin, A.E. (2011). *Single-case research designs: Methods for clinical and applied settings*. New York: Oxford University Press.
- Kern, H.J. (1997). Einzelfallforschung: Versuchsplan-Kombinationen für die (Sonder-) Pädagogik. *Vierteljahresschrift für Heilpädagogik und ihre Nachbargebiete*, 66, 325–336.
- Parker, R.I., Brossart, D.F., Vannest, K.J., Long, J.R., De-Alba, R.G., Baugh, F.G. et al. (2005). Effect Sizes in Single Case Research: How Large is Large? *School Psychology Review*, 34, 116–132.
- Parker, R.I., Vannest, K.J. & Davis, J.L. (2011). Effect size in single-case research: A review of nine nonoverlap techniques. *Behavior Modification*, 35, 303–322.
- Petermann, F. (1996). *Einzelfallanalyse*. München: Oldenbourg Verlag.
- Price, A.T., Martella, R.C., Marchand-Martella, N.E. & Cleanthous, C.C. (2002). A comparison of immediate feedback delivered via an FM headset versus delayed feedback on the inappropriate verbalizations of a student with ADHD. *Education & Treatment of Children*, 25, 159–171.
- Ruiz-Olivares, R., Pino, M.J. & Herruzo, J. (2010). Reduction of disruptive behaviors using an intervention based on the good behavior game and the say-do-report correspondence. *Psychology in the Schools*, 47, 1046–1058.

- Scruggs, T.E., Mastropieri, M.A., Cook, S.B. & Escobar, C. (1986). Early intervention for children with conduct disorders: A quantitative synthesis of single-subject research. *Behavioral Disorders, 11*, 260–271.
- Sidman, M. (1960). *Tactics of scientific research*. Oxford, UK: Basic Books.
- Singh, N.N., Lancioni, G.E., Winton, A.S.W., Singh, A.N.A., Singh, J. & Singh, A.D.A. (2011). Effects of a mindfulness-based smoking cessation program for an adult with mild intellectual disability. *Research in Developmental Disabilities, 32*, 1180–1185.

## **Anschrift der Autoren**

*DR. ANITA JAIN*  
*Universität zu Köln*  
*Department Psychologie*  
*Herbert-Lewin-Straße 2*  
*50931 Köln*  
*a.jain@uni-koeln.de*