

Grundlagen der Diagnostik

Lerneinheit 3

Verhaltensbeobachtung



We are happy to share our materials openly:

The content of these [Open Educational Resources](#) by [Lehrstuhl für Psychologische Methodenlehre und Diagnostik, Ludwig-Maximilians-Universität München](#) is licensed under [CC BY-SA 4.0](#). The CC Attribution-ShareAlike 4.0 International license means that you can reuse or transform the content of our materials for any purpose as long as you cite our original materials and share your derivatives under the same license.

1. Das konjunktive Entscheidungsmodell eignet sich besonders, wenn Anforderungen nicht kompensierbar sind und jede Teilanforderung erfüllt werden muss.
2. Die vollständig-sequentielle Strategie erlaubt terminale Entscheidungen nach einem einzigen Verfahren.
3. Das Gütemerkmal der Objektivität ist insbesondere bei der Durchführung von Tests entscheidend.
4. Das konjunktive Entscheidungsmodell wird eingesetzt, wenn Teilanforderungen nicht kompensierbar sind.
5. Eine hohe Korrelation mit einem anderen Test ist immer ein Beweis für Validität.

1. Ziele und Planung

2. Beobachtungs- und Beurteilungsfehler

3. Spezielle Anwendungen

Lernziele 

1. Ziele & Planung

Ziele einer Verhaltensbeobachtung

- „Mit einem Verfahren zur Verhaltensbeobachtung und Verhaltensbeurteilung versucht man, durch die Anwendung wissenschaftlich fundierter Prinzipien die Aufnahme, Integration und Verwertung von Verhaltensinformationen in Eignungsbeurteilungen so zu gestalten, dass **Fehlereinflüsse reduziert werden und letztlich die Wahrscheinlichkeit für eine hohe Gültigkeit der getroffenen Aussagen steigt.**“ Höft & Kersting, 2017

Anwendungsszenarien

Beispiele:

- Gruppendiskussionen
- Rollenspiele
- Präsentationsübungen
- Mutter-Kind-Interaktionen

...

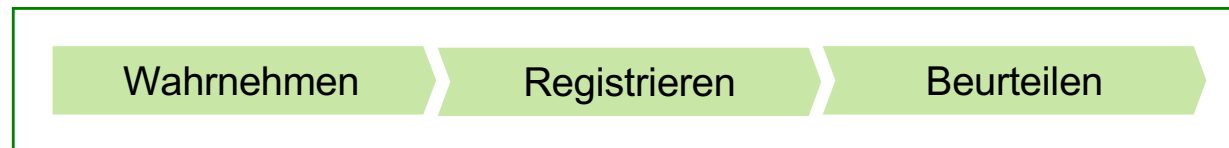
Kontexte:

- Bewerberauswahl
- Potenzialanalyse
- Forschung
- Klinische Diagnosen

...

Ziele einer Verhaltensbeobachtung

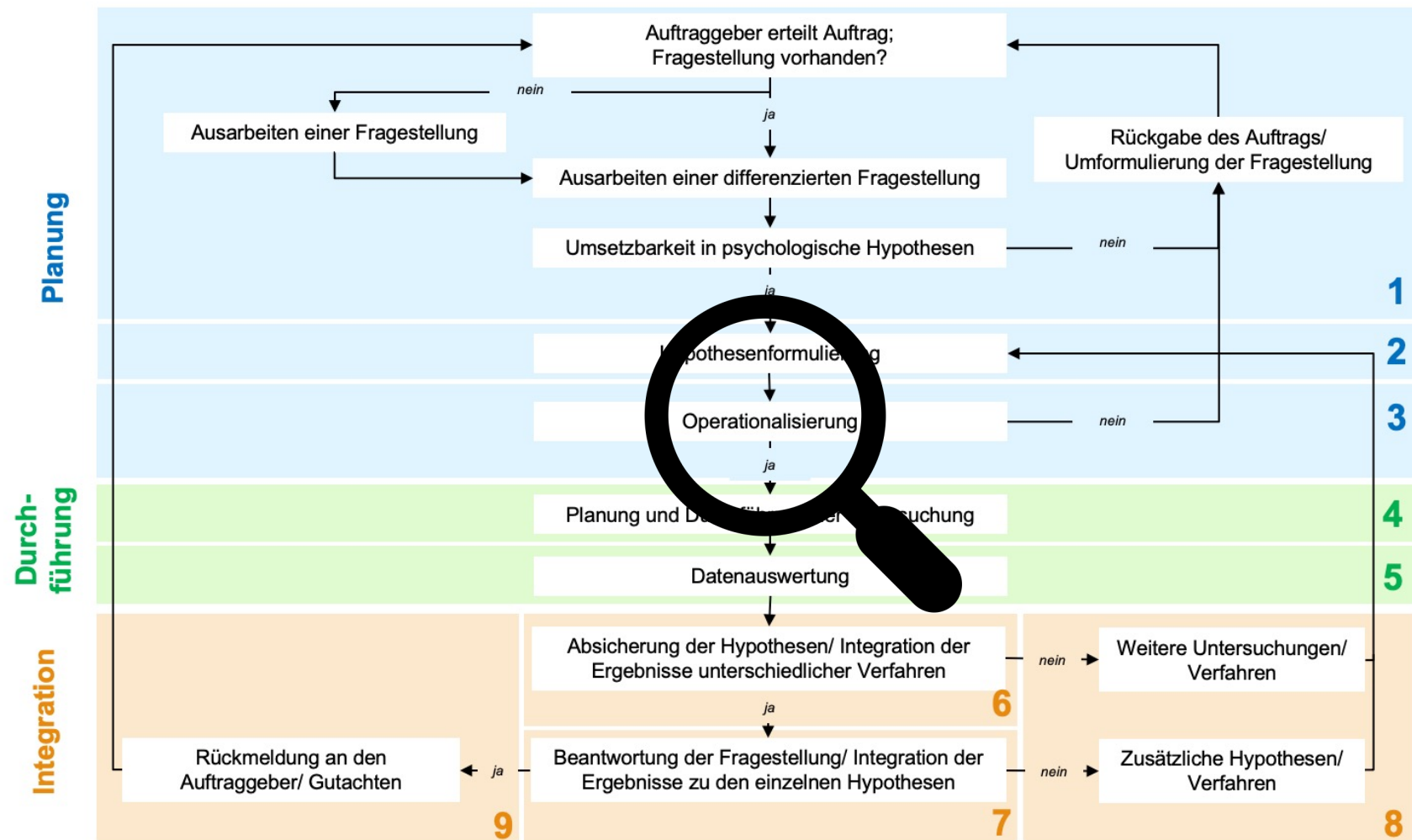
- „**Verhaltensbeobachtung liefert im Idealfall Fakten, die frei von Bewertungen sind.**“ Schmidt-Atzert & Amelang, 2012, S.309
- Diese Zielsetzung impliziert eine wichtige Unterscheidung zwischen **Beobachtung vs. Beurteilung**
- Wahrnehmung (bzw. Beobachtung), Registrierung (bzw. Protokollierung) und Beurteilung sollten im Prozess der Verhaltensbeobachtung getrennt vorgenommen werden:



Schritte der Verhaltensbeobachtung

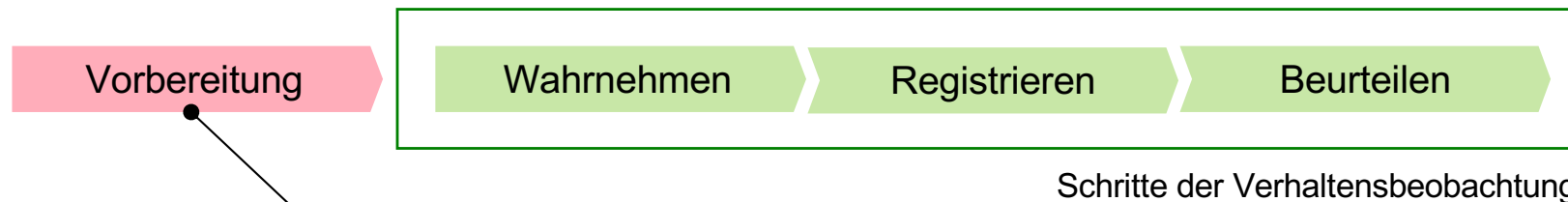
- Eine gute Wahrnehmung setzt voraus, dass genau definiert wird, was und wie etwas beobachtet (und später beurteilt) werden soll
→ Operationalisierung, Beobachtungsplan

1. Verhaltensbeobachtung: Ziele und Planung



Systematik der Beobachtung

Wie?	Unsystematisch-frei (unstrukturiert)	vs.	Systematisch-gebunden (strukturiert) → Beobachtungsplan!
Wer?	Selbstbeobachtung	vs.	Fremdbeobachtung
			Teilnehmend vs. Nicht-teilnehmend (aktiv, passiv)
Wo?	Feld	vs.	Labor
Wann?	Indirekt (versetzt)	vs.	Direkt (gleichzeitig)
Womit?	Technisch (vermittelt)	vs.	Unvermittelt
Wissenschaftlich?	Verdeckt	vs.	Offen



Phasen bei der **Vorbereitung** einer Verhaltensbeobachtung:

1. Operationalisierung der Zielmerkmale
2. Art der Registrierung / Protokollierung festlegen
3. Beurteilungsregeln festlegen

1. Operationalisierung der Zielmerkmale

Die Operationalisierung der Zielmerkmale bei der Verhaltensbeobachtung erfordert die Festlegung von **Indikatoren**:

- Indikatoren sind Hinweisreize, die direkt beobachtbar und klar voneinander abgrenzbar sind und für das Vorhandensein des Merkmals sprechen
- Unterscheide
 - **Offenes Verhalten**
 - z.B. verbale Äußerungen, motorische Aktivitäten, Gestik und Mimik
 - **Inneres Erleben**
 - z.B. gedankliche Prozesse, Stimmungen, körperliche Befindlichkeiten, Motivation
 - kann in der Fremdbeobachtung nur indirekt über offenes Verhalten erfasst werden
 - Alternative: Selbstbeobachtung über Introspektion, z.B. über Experience Sampling
- Beispiele
 - Wiederholen des Gesagten in einem Gespräch als Indikator für aktives Zuhören
 - Schubsen, Schlagen als Indikator für Aggressivität
 - Erröten bei Aufregung als Verhaltensindikator inneren Erlebens

1. Operationalisierung der Zielmerkmale

2. Art der Reg./Prot. festlegen

3. Beurteilungsregeln festlegen

Wie kann man Indikatoren festlegen? → Critical Incident Technique (CIT)

- Definition:
 - Methode der kritischen Ereignisse
 - Zentrale Technik, um **verhaltensnahe Indikatoren** für **erfolgskritische Situationen** zu generieren
- Ursprünglich wurde CIT als **induktiver Bottom-Up-Ansatz** verwendet:
 - Merkmale / Eigenschaften sind **noch nicht vorhanden**
 - Ableitung der Merkmale und Indikatoren aus konkreten Situationen als Grundlage für eine Anforderungsanalyse / Fragestellung
 - Dazu werden Workshops/Interviews mit Expertinnen durchgeführt, z.B. „Versuchen Sie sich bitte an eine konkrete Situation im Kontext XY zu erinnern, in der **Ihr Verhalten** erfolgreich / förderlich / hinderlich war!“
 - Verhaltensbeschreibung lässt Rückschlüsse auf das Zielmerkmal zu

1. Operationalisierung der Zielmerkmale

2. Art der Reg./Prot. festlegen

3. Beurteilungsregeln festlegen

Wie kann man Indikatoren festlegen? → Critical Incident Technique (CIT)

- Da im diagnostischen Prozess in der Phase der Operationalisierung das Zielmerkmal schon festgelegt ist, wird CIT in diesem Kontext als **deduktiver Top-Down-Ansatz** verwendet:
 - Merkmale / Eigenschaften sind **schon vorhanden**,
z.B. durch Fragestellung, Kompetenzmodell des Unternehmens
 - Ableitung der Indikatoren aus konkreten Situationen
 - Dazu werden auch Workshops/Interviews mit Expertinnen durchgeführt,
z.B. „Versuchen Sie sich bitte an eine konkrete Situation zu erinnern,
in der **diese Kompetenz** wichtig war!“

1. Operationalisierung der Zielmerkmale

2. Art der Reg./Prot. festlegen

3. Beurteilungsregeln festlegen

Wie kann man Indikatoren festlegen? → Critical Incident Technique (CIT)

- Beispiele für Nachfragen im Rahmen der CIT, um das genannte Verhalten/Kompetenzen einordnen zu können:
 - Könnten Sie diese Situation genauer beschreiben?
 - Wie kam es zu dieser Situation?
 - Was waren die Umstände oder Hintergrundbedingungen, die zu der Situation führten?
 - Wer war an der Situation beteiligt?
 - Welche Konsequenzen hatte das genannte Verhalten genau?

1. Verhaltensbeobachtung: Ziele und Planung

1. Operationalisierung der Zielmerkmale

2. Art der Reg./Prot. festlegen

3. Beurteilungsregeln festlegen

Beispiele eines Ergebnisses einer CIT aus der Praxis:

Kompetenz 1: Gastorientierung: gute Umgangsformen, spricht und erscheint vertrauenswürdig, sicheres Auftreten, Serviceorientierung, Freude daran, Freude zu bereiten				
Situation	Verhalten	Negatives Verhalten	Folge	Zitat
Mitarbeiter arbeitet im Front Cooking/ Buffet Bereich	das Essen vor dem Gast zubereiten, ihn nett und freundlich empfangen, Essen "verkaufen", Empfehlungen aussprechen, Essen anpreisen/Vorzüge betonen z. B. "frisches Wokgemüse mit besonderen Palmenherzen"		nicht nur Gast beeindruckt sondern evtl. zusätzlich Umsatzsteigerung da Essen höheren Preis hat	
Großveranstaltung z. B. Abiball mit 900 bis 1000 Gästen, Führung plant falsch mit zu wenigen Mitarbeitern	vorher Feedback an Führung "Vielleicht passt das nicht ganz, bitte nochmal überdenken", sich immer wieder fragen "Kann ich leisten, was von mir erwartet wird?"			
Gäste bestellen a la carte abends, Küchenchef ist nicht da	Als Stellvertreter mal zum Gast gehen, fragen ob es geschmeckt hat, Reklamationen entgegennehmen, sich entschuldigen, Kompensationen anbieten, sich für Lob bedanken können			
Entgegennehmen einer Reklamation	wissen in welchem Rahmen man handeln darf, was man geben darf, auch mal selbstständig entscheiden, wenn es über den vorgesehenen Rahmen hinausgeht sich bedenkenlos an den Vorgesetzten wenden, sich entschuldigen wenn Reklamation berechtigt, relevante und attraktive Kompensation anbieten (nicht nur eine Tasse Kaffee), Feedback vorher einholen, um welchen Kunden es sich handelt (Firma, Vollpension, selbst zahlend), abwägen was wichtiger ist: Verlust von 16,50 Euro für ein Schnitzel oder eventuelle finanzielle Folgeschäden		Kunde ist Sprachrohr nach außen, Unzufriedenheit kann bis zu 10.000 Euro kosten	
Essenproduktion (z. B. Kaffeepause)	Speisen wetterabhängig zusammenstellen z. B. bei Hitze keine Cremesuppe sondern kalte Melonensuppe, Einkauf darauf abstimmen, keine Speisen anbieten die schlecht werden können, Speisen in Pausen immer wieder erneuern/auffrischen, kurzfristig noch ein Highlight obendrauf für die Gäste setzen, z. B. bei Hitze als Überraschung Eis anbieten		Gästen ist geholfen, z. B. damit sie sich bei Hitze angenehmer fühlen, nicht müde werden, nicht schwitzen	

1. Operationalisierung der Zielmerkmale

2. Art der Reg./Prot. festlegen

3. Beurteilungsregeln festlegen

Wie differenziert sollten Indikatoren sein?

- Indikatoren können auf unterschiedlichem **Differenzierungslevel** erhoben werden,
 - z.B. für das Zielmerkmal Aggressivität: Indikatoren „Schlagen“, „Treten“, „Haare ausreißen“, „Beißen“, „Kratzen“ oder: „Körperlich verletzen“
- Differenzierungslevel ist abhängig von:
 - Zweck der Untersuchung
 - Kapazität und Erfahrung der Beobachterinnen
 - Beobachtungsdauer
 - Beobachtungssituation

1. Operationalisierung der Zielmerkmale

2. Art der Reg./Prot. festlegen

3. Beurteilungsregeln festlegen

Sind die Indikatoren geeignet? → Überprüfung der Güte: Validität

- **Zentrale Frage:**
 - Kann man das Verhalten (Indikatoren) wirklich dem Zielmerkmal, also einer bestimmten Kompetenz/Fertigkeit/Eigenschaft zuordnen?
- Bei der Beurteilung der Güte von Indikatoren sollte man also bedenken:
 - **Inhaltsvalidität:** Ist mit den gewählten Indikatoren das Merkmal in allen relevanten Facetten beschrieben?
 - **Situationsabhängigkeit:**
 - Ist theoretisch erwartbar, dass die Indikatoren in der Beobachtungssituation überhaupt vorkommen?
 - Nicht in jeder Situation werden vorhandene Verhaltensgewohnheiten gezeigt, z.B. Pflichtbewusstsein nur bei Wahlfreiheit beobachtbar, Ängstlichkeit nur bei Bedrohung

1. Operationalisierung der Zielmerkmale

2. Art der Reg./Prot. festlegen

3. Beurteilungsregeln festlegen

Sind die Indikatoren geeignet? → Überprüfung der Güte: **Validität**

- Statistische Analysen zur Überprüfung der Validität:
 - Interkorrelation von Skalen → „klassische“ Konstruktvalidität
 - Faktorenanalysen
 - Analysen zur Kriteriumsvalidität
 - Gruppenvergleiche
 - Veränderungen im Auftreten der Indikatoren nach einer Intervention
→ Stichwort: kausale Validität!
- Was kann man tun, um die Validität der Indikatoren zu verbessern?:
 - Indikatoren anpassen, indem Expertinnen einbezogen werden
 - Beobachtungs- und Beurteilungsfehler reduzieren (z.B. Halo-Effekt)

1. Operationalisierung der Zielmerkmale

2. Art der Reg./Prot. festlegen

3. Beurteilungsregeln festlegen

Sind die Indikatoren geeignet? → Überprüfung der Güte: **Objektivität**

- **Zentrale Fragen:**

- Lassen sich die Indikatoren zuverlässig von allen Beobachterinnen identifizieren?
- Verstehen alle Beobachterinnen unter den Indikatoren dasselbe?
- Sind die Indikatoren verhaltensnah, konkret bzw. genau definiert?
- Haben die Indikatoren eine klar erkennbare zeitliche Erstreckung (Beginn, Dauer, Ende)?

1. Operationalisierung der Zielmerkmale

2. Art der Reg./Prot. festlegen

3. Beurteilungsregeln festlegen

Sind die Indikatoren geeignet? → Überprüfung der Güte: **Objektivität**

- Statistische Analysen zur Überprüfung der Objektivität:
 - Berechnung von Beobachter- bzw. Beurteilerübereinstimmung (→ LE4-6)
- Was kann man tun, um die Objektivität der Indikatoren zu verbessern?:
 - Indikatoren anpassen in Hinblick auf Überlegungen zur Objektivität
 - Beobachterschulung
 - Beobachtungs- und Beurteilungsfehler reduzieren (z.B. Beobachterdrift)

1. Operationalisierung der Zielmerkmale

2. Art der Reg./Prot. festlegen

3. Beurteilungsregeln festlegen

Sind die Indikatoren geeignet? → Überprüfung der Güte: Reliabilität

- **Zentrale Fragen:**

- Ist die Liste der Indikatoren von einer einzelnen Beobachterin während des Beobachtungszeitraums insgesamt überblickbar?
- Ansonsten ähnliche Aspekte wie bei Objektivität

1. Operationalisierung der Zielmerkmale

2. Art der Reg./Prot. festlegen

3. Beurteilungsregeln festlegen

Sind die Indikatoren geeignet? → Überprüfung der Güte: **Reliabilität**

- Statistische Analysen zur Überprüfung der Reliabilität:
 - Retestreliabilität bei Annahme von zeitlicher / situativer Stabilität
 - Ggf. Konsistenzanalysen
 - Bestimmte Maße der Beurteilerübereinstimmung (ICC) können auch als Reliabilitätsschätzung verwendet werden
- Was kann man tun, um die Reliabilität zu verbessern?
 - Aggregation **über verschiedene Situationen** wegen der Situationsabhängigkeit des Verhaltens!
 - z.B. Aggressivität im Schulkontext oder Präsentationsfähigkeiten einer Lehrkraft → Aggregation über mehrere Unterrichtsstunden
 - Aggregation **über verschiedene Indikatoren** eines Merkmals in einer Situation (z.B. Summenscore)
 - Beobachtungs- und Beurteilungsfehler reduzieren (z.B. Strenge)

2. Art der Registrierung / Protokollierung

= Die Indikatoren sind ausgewählt – Wie halte ich sie fest?

- Festlegung eines **Indikator-, „Maßes“**:

	Häufigkeit/Dauer	Intensität
Was passiert bei einer starken Ausprägung einer Eigenschaft?	Häufigkeit / Dauer des Indikators steigt	Intensität des Indikators steigt
Beispiel: Kundenorientierung	Häufigeres Lächeln, Längerer Blickkontakt	Sprachliche Anpassung (Dialekt, Lautstärke)
Wie erfolgt die Beobachtung?	Häufig Zeichensysteme (z.B. Strichlisten) oder Zeitmessung	Häufig freie Mitschrift während der Beobachtung und anschließende Gewichtung von Indikatoren

- Festlegung eines **Sampling-Plans**:

	Event-Sampling	Time-Sampling
Vorgehen	Wie oft oder wie lange insgesamt?	Unterteilung in exakt gleich lange Beobachtungsintervalle: Verhalten im entsprechenden Zeitintervall beobachtet?
Häufigkeit	Person hat in der Beobachtungszeit 7 Fragen gestellt, 3 andere unterbrochen, 1 Lösungsvorschlag unterbreitet	Person hat in allen 4 Zeitintervallen Fragen gestellt
Dauer	Person hat insgesamt 4 Minuten lang Fragen gestellt	Person hat in allen 4 Zeitintervallen Fragen von jeweils 1 Minute gestellt



Vorteil:

Möglichkeit der Pause für
Beobachterinnen, zeitabhängige
Beobachtung

1. Verhaltensbeobachtung: Ziele und Planung

1. Operationalisierung der Zielmerkmale

2. Art der Reg./Prot. festlegen

3. Beurteilungsregeln festlegen

Beispiel für ein Time-Sampling:

	Zeitintervall (1 bis n)						Σ
Indikator / Merkmalsbereich	1	2	3	4	...	n	
Greift Vorschlag anderer auf							
Revidiert eigenen Vorschlag							
Zeigt Zustimmung							
Σ Kooperation							
Verteidigt eigenen Vorschlag							
Kritisiert andere							
Unterbricht andere							
Σ Durchsetzung							

1. Operationalisierung der Zielmerkmale

2. Art der Reg./Prot. festlegen

3. Beurteilungsregeln festlegen

- Festlegung von **Hilfsmitteln**:
 - Videoaufzeichnung (kontrollierbar und wiederholbar vs. Ökonomie und Reaktivität)
 - Tabellensysteme (siehe Time Sampling)
 - Protokollbögen
 - Freie Mitschrift
 - Stoppuhr / Zeitgeber

3. Beurteilungsregeln festlegen

Beurteilungsmodelle müssen **vor** der Verhaltensbeobachtung festgelegt werden; und zwar für **alle (!)** verfügbaren Informationen:

Schritt 1:

- Zusammenführen vieler einzelner Indikatoren zu einer Beurteilung der Merkmalsausprägung

Schritt 2:

- Zusammenführen vieler Zielmerkmale zu einer Entscheidung

1. Verhaltensbeobachtung: Ziele und Planung

1. Operationalisierung der Zielmerkmale

2. Art der Reg./Prot. festlegen

3. Beurteilungsregeln festlegen

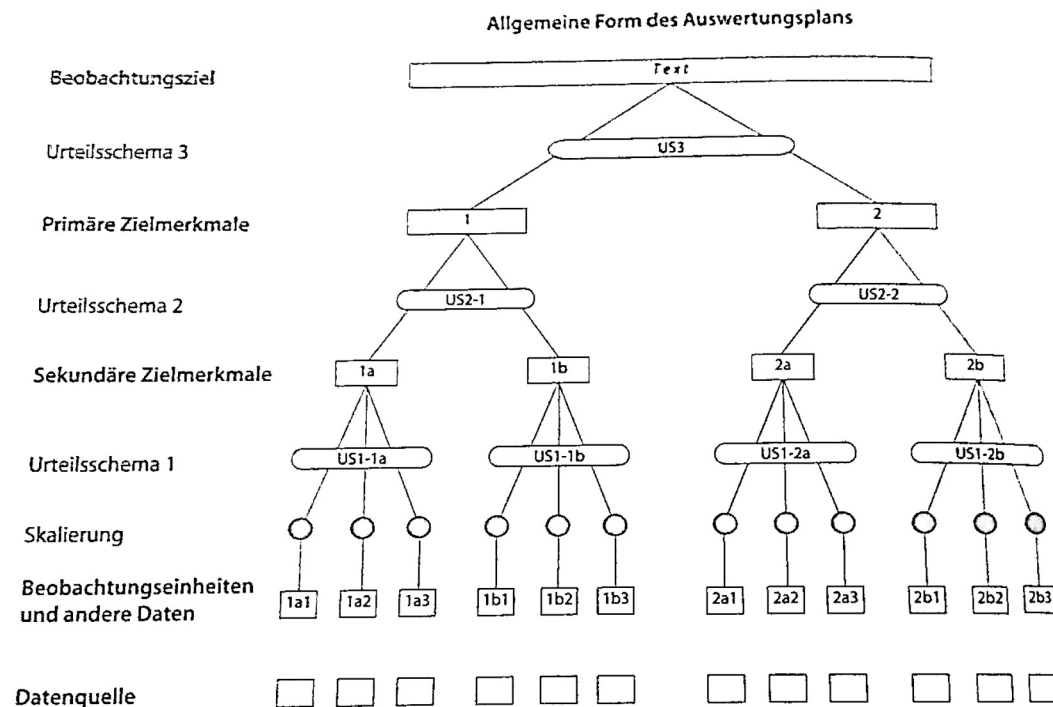


Abbildung 1-15: Allgemeine Form des Auswertungsplans, hier am Beispiel von zwei primären und vier sekundären Zielmerkmalen sowie jeweils drei Beobachtungseinheiten oder anderen Daten. Am Fuß der Hierarchie wird in der Zeile «Datenquelle» eingetragen, woher die Daten stammen. Urteilsschemata (US) werden mit Angabe ihrer hierarchischen Stufe (1, 2 oder 3) und ihres Zielmerkmals bezeichnet.

1. Operationalisierung der Zielmerkmale

2. Art der Reg./Prot. festlegen

3. Beurteilungsregeln festlegen

Schritt 1: Zusammenführen vieler einzelner Indikatoren zu einer Beurteilung der Merkmalsausprägung

Integration von **Häufigkeit** vs. **Bedeutsamkeit** des Auftretens des Verhaltens (→ Skalierung):

- Viele Indikatoren, die für das Merkmal sprechen und wenige Indikatoren, die gegen das Merkmal sprechen → hohe Ausprägung
- Bedeutsame Indikatoren, die für das Merkmal sprechen und nicht bedeutsame Indikatoren, die gegen das Merkmal sprechen → hohe Ausprägung

Schritt 1: Zusammenführen vieler einzelner Indikatoren zu einer Beurteilung der Merkmalsausprägung

- Wie kombiniere ich die Indikatoren?
 - Gibt es Indikatoren, die eine höhere Gewichtung haben?
 - Idee der Entscheidungsmodelle (→ LE2) kann herangezogen werden:
 - Konjunktiv (auf jedem Indikator eine Mindestanzahl)?
 - Disjunktiv (auf mindestens einem Indikator eine Mindestanzahl)?
 - Kompensatorisch (in der ggf. gewichteten Gesamtsumme eine Mindestanzahl)?

- Beispiel:

Indikator / Merkmalsbereich	Zeitintervall (1 bis n)						Σ
	1	2	3	4	...	n	
Greift Vorschlag anderer auf							
Revidiert eigenen Vorschlag							
Zeigt Zustimmung							
Σ Kooperation							

1. Operationalisierung der Zielmerkmale

2. Art der Reg./Prot. festlegen

3. Beurteilungsregeln festlegen

Schritt 1: Zusammenführen vieler einzelner Indikatoren zu einer Beurteilung der Merkmalsausprägung

Häufig werden die Indikatoren (v.a. bei freier Mitschrift) in verhaltensverankerte Skalen übersetzt:

Aggressivität				
0	1	2	3	4
gar nicht	wenig	etwas	deutlich	stark
0 = zeigt kein verbales oder körperlich aggressives Verhalten 1 = schimpft gelegentlich zurück, wird aber nicht initiativ 2 = schimpft, wenn etwas nicht klappt 3 = schreit andere an oder wirft Sachen um 4 = brüllt bei geringsten Anlässen oder wirft mit Gegenständen oder rempelt andere an				

Schritt 1: Zusammenführen vieler einzelner Indikatoren zu einer Beurteilung der Merkmalsausprägung

Beispiel einer verhaltensverankerten
Beurteilungsskala aus der Praxis: wurde als
Basis für Mitarbeitergespräche benutzt

1. Kompetenz: *Gastorientierung*

Der Mitarbeiter zeigt gute Umgangsformen, ist vertrauenswürdig, hat ein sicheres Auftreten, zeigt Serviceorientierung, hat Freude daran Freude zu bereiten.

	Verhalten
Anwenden	- Gast freundlich begrüßen - Mit dem Gast Blickkontakt aufnehmen auch, wenn man gerade beschäftigt ist (z.B. telefoniert oder anderen Gast aufnimmt) - Positive Ausstrahlung zeigen, z.B. lächeln - Gästewünsche erfragen und im Rahmen der eigenen Möglichkeiten berücksichtigen - Relevante Informationen an betroffene Kollegen weiterleiten und ggf. Rücksprache halten, ob Kundenwünsche erfüllbar - Nach Zufriedenheit vom Gast fragen und dieses auch weiterleiten - Angebote im Haus und der Umgebung kennen - Auf saubere Uniform / Kleidung achten - Direkt und positiv auf Reklamationen bzw. Beschwerden reagieren und Verständnis zeigen - USP im Haus / Marke kennen - Bei Bedarf Hilfe / Unterstützung / Zusammenarbeit abteilungsübergreifend anfragen
Anpassen	- Positiv auf den Gast eingehen und Verständnis für die Belange des Gasts zeigen - Stammgäste wiedererkennen - Handlungskompetenz / Beschwerdestandards bei Reklamationen kennen und nutzen - Sich angemessen im Rahmen der eigenen Ressourcen Zeit für den Gast nehmen - Aufmerksam für Gästebemerkungen / Gästewünsche sein, die mögliche Verbesserungspotenziale aufzeigen und diese Information weiterleiten - Sich für Lob bedanken können
Gestalten	- Marktlage und Handlungsspielraum kennen und wissen, was angeboten werden kann - Optimal auf das Gastbedürfnis abgestimmt Zeit für den Gast nehmen - Gäste mit Namen ansprechen - Vorhandene Informationen über Stammgäste nutzen - Auch kleine Fehler erkennen und reagieren - Behebung weitergeleiteter Mängel / Beschwerden nach verfolgen

Schritt 2: Zusammenführen vieler Zielmerkmale zu einer Entscheidung

- Siehe diagnostischer Prozess: Integration der Ergebnisse in einem Entscheidungsmodell zur Urteilsfindung
- Beispiel: Wie werden die Ergebnisse von Kooperation und Durchsetzung zu einem Urteil über die Person kombiniert?

Indikator / Merkmalsbereich	Zeitintervall (1 bis n)						Σ
	1	2	3	4	...	n	
Greift Vorschlag anderer auf							
Revidiert eigenen Vorschlag							
Zeigt Zustimmung							
Σ Kooperation							
Verteidigt eigenen Vorschlag							
Kritisiert andere							
Unterbricht andere							
Σ Durchsetzung							

2. Beobachtungs- und Beurteilungsfehler

Allgemeines

- Beobachtungs- und Beurteilungsfehler sind (experimentell) gut erforscht, trotzdem ist **wirksame** Fehlervermeidung sehr schwierig
- Statistische Kontrolle ist meist praktisch nicht umsetzbar
- Abhilfe in gewissem Umfang leisten Aufklärung und sinnvolle Gegenmaßnahmen
- Fehler sind häufig Ausdruck unbewusster oder bewusster Entscheidungen (z.B. Unsicherheit → zentrale Tendenz)
- **Es bleibt das Problem der Reaktivität:** Allein das Wissen, dass beobachtet wird, verändert bereits die Situation, ganz unabhängig davon, wer beobachtet

Allgemeines

- **Vorteile technischer Methoden:** Fehler und subjektive Verzerrungen bei der Beobachtung minimieren
- **Einhaltung gesetzlicher Bestimmungen:**
 - Einverständnis des Beobachteten zur Untersuchung und zur Aufzeichnung von Daten: Unwissentliche Beobachtungen kommen in der Diagnostik nicht in Frage!
 - Schweigepflicht
 - Datenschutz

Beispiele für Beobachtungs- und Beurteilungsfehler

Übersicht:

- Verhalten, das Reaktivität fördert
- Beobachterdrift
- Haloeffekt
- Logische Fehler
- Primacy Effekt & Recency Effekt
- Milde-Effekt & Besondere Strenge
- Zentrale Tendenz
- Extremurteile

Beispiele für Beobachtungs- und Beurteilungsfehler

Verhalten, das Reaktivität fördert

- Hervorhebung der Präsenz
- Auskosten der „Macht“
- Geflüster, Gespräche
- nonverbale „Kommentierung“ des Geschehens (Lächeln, Kopfschütteln, Hochziehen der Augenbrauen etc.)
- Kleidung

→ Angemessen ist neutrales, zurückhaltendes Verhalten:

Prinzip Zimmerpflanze

Beispiele für Beobachtungs- und Beurteilungsfehler

Beobachterdrift

- Ohne tatsächliche Verhaltensänderung werden zunehmend mehr oder immer weniger relevante Ereignisse entdeckt
- Ermüdung und nachlassende Aufmerksamkeit (Abnahme)
- Übungseffekt beim Erkennen der Indikatoren (Zunahme)

→ **Angemessen lange Beobachtungsphasen wählen (max. 20 Minuten)**

Beispiele für Beobachtungs- und Beurteilungsfehler

Haloeffekt

- Eine hervorstechende Eigenschaft „überstrahlt“ andere Eigenschaften
 - Beispiel: Charme führt dazu, dass auch fachliche Kompetenz sehr positiv beurteilt wird
 - Haloeffekt erklärt möglicherweise die Tatsache, dass Ratings stark unterschiedlicher Eigenschaften oft mittelstark zusammenhängen
- **Begrenzung der zeitgleich beurteilten Merkmale (max. 4)**
- **Außerdem: Können die Merkmale in einer Übung überhaupt differenziert und konkret beobachtet werden?**

Beispiele für Beobachtungs- und Beurteilungsfehler

Logische Fehler

- Zur Beurteilung werden Hinweise herangezogen, die nicht passen, z.B. Brille → Intelligenz
- Falsche Schlussfolgerungen: „Wenn er das eine gut kann, dann kann er das andere sicher auch gut“
- Implizite Persönlichkeitstheorien: „Höfliche Menschen sind auch teamfähig“, „Dominante Kinder sind auch aggressiv“

→ **Standardisierte und verhaltensbasierte Beurteilungsinstrumente / Richtlinien**

Beispiele für Beobachtungs- und Beurteilungsfehler

Primacy Effekt

- „Auf den ersten Eindruck kommt es an.“
- Informationen, die zu Beginn aufgenommen werden...
 - ...helfen, die Wahrnehmung zu organisieren
 - ...leiten die Hypothesen über die Person: Fokus auf bestätigende Hinweise, obwohl eine Falsifizierungsstrategie klüger wäre
 - ...werden besser gemerkt, sind bei Urteilsbildung dann sehr gut verfügbar und werden daher besonders berücksichtigt

→ **Kontinuierliche Verhaltensaufzeichnung wenn möglich, sonst Protokollierung**

Beispiele für Beobachtungs- und Beurteilungsfehler

Recency Effekt

- Zuletzt gewonnene Informationen werden übergewichtet
- Geringerer Verfallsprozess im Gedächtnis, dadurch bessere Verfügbarkeit
- Plausible Erklärung: Wenn zuvor ein unsicherer Eindruck entstanden ist, stützt man sich auf die letzte Information
- Primacy und Recency Effekt schließen sich nicht aus, sie können auch zusammen vorkommen

→ **Kontinuierliche Verhaltensaufzeichnung wenn möglich, sonst Protokollierung**

Beispiele für Beobachtungs- und Beurteilungsfehler

Milde-Effekt & Besondere Strenge

- Verzerrung zum Positiven oder Negativen
- Beispiel: Kenntnis der Berufsanforderungen auf Seiten der Beurteilenden führt zu Sympathie oder Mitleid

→ **Kritische Beobachterpaarungen vermeiden**

→ **Standardisierte und verhaltensbasierte Beurteilungsinstrumente / Richtlinien**

Beispiele für Beobachtungs- und Beurteilungsfehler

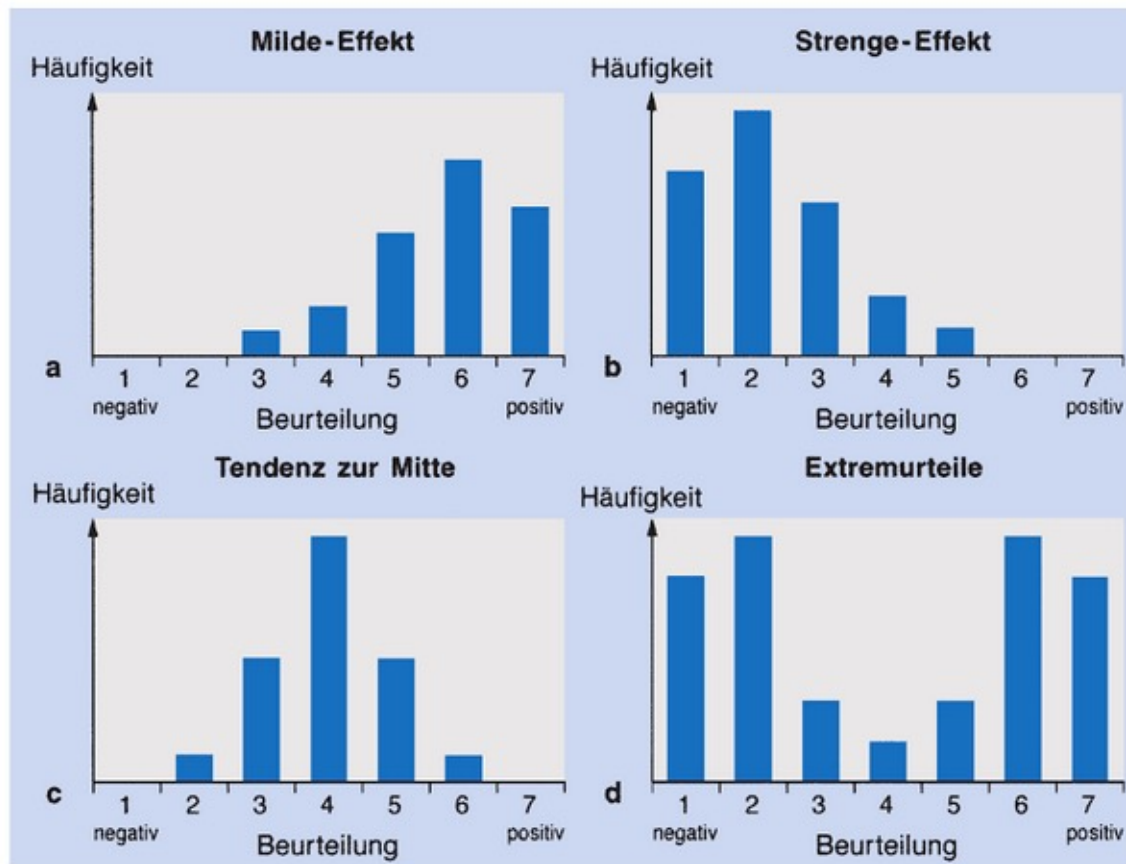
Zentrale Tendenz

- Mittlere Ratings werden bevorzugt gewählt, Vermeidung von Extremurteilen
- Häufig Ausdruck von Unsicherheit oder Überforderung: Flucht zur Mitte, wenn Beurteiler nicht weiß, welches Urteil angemessen ist
- Förderung durch Verwendung bipolarer Skalen wegen Indifferenzbereich in der Mitte
- Gegenteil: Extremurteile (seltener)
- Vgl. Phänomen der Antwortstile bei standardisierten Fragebögen

→ **Verwendung unipolarer Skalen ohne Mittelkategorie**

→ **Antwortkategorie „nicht beobachtbar“ (auch Nachteile!)**

Beispiele für Beobachtungs- und Beurteilungsfehler



■ **Abb. 3.29a-d.** Illustration einiger Beurteilungsfehler bei der Einschätzung von Eigenschaftsausprägungen oder Verhaltenstendenzen

Vorsicht: Aus den Verteilungen der Urteile kann nicht eindeutig geschlossen werden, ob Beobachtungs- und Beurteilungsfehler vorliegen, da man die wahre Verteilung der Merkmale nicht kennt!

Generelle Maßnahmen

- Ausreichendes Beobachtertraining
(etwa 2 Tage, z.B. mit Videos)
- Zusammensetzung der Beobachter:
 - Neutrale, erfahrene, externe Beobachterinnen wählen
 - Prinzip der Doppelbeobachtung
 - Mittelung der Bewertungen (→ höhere Reliabilität) oder gemeinsames Urteil finden (→ Beobachterkonferenz)

Generelle Maßnahmen

- Workload der Beobachterinnen beachten
 - max. 20 Minuten
 - max. 4 Merkmale gleichzeitig
 - max. 1-2 Personen gleichzeitig
- Standardisierter Beobachtungsplan mit verhaltensverankerten Beobachtungs- und Beurteilungsbögen

3. Spezielle Anwendung

Anwendungsfall 1: Assessment Center

Anwendungsfall 2: Ambulatory Assessment

Anwendungsfall 1: Assessment-Center

- 1920 in Deutschland für das Militär zur Auswahl von Offiziersanwärtern entwickelt
- „Ein Assessment-Center ist ein Beurteilungsverfahren, in dem durch die **Beobachtung der Leistung und des Verhaltens** der Teilnehmerinnen in Simulationen, Rollenspielen, Tests und Fallstudien Rückschlüsse auf Kompetenzen, Persönlichkeitseigenschaften und Potenziale gezogen werden.“ Paschen et al., 2005, S.16

Anwendungsfall 1: Assessment Center

Anwendungsfall 2: Ambulatory Assessment

Fünf Anwendungsprinzipien:

- **Anforderungsbezug:** Anforderungsprofil erstellen
- **Simulation:** Rollenspiele, Fallstudien, Gruppendiskussionen
- **Methodenvielfalt:** Mehrere Übungen pro Anforderungsmerkmal
- **Einsatz mehrerer Beobachterinnen:** Beobachterkonferenz
- **Transparenz:** Information an die Teilnehmerinnen über die Anforderungen

Anwendungsfall 1: Assessment Center

Anwendungsfall 2: Ambulatory Assessment

Validität?

- **Kriteriumsvalidität**
 - **mit viel Varianz** (zwischen $-.25$ und $.78$ bei Gaugler et al., 1987), aber auch weil **ACs sehr heterogen** in den Methoden, Anforderungen und Zielmerkmalen sind („by design“!)
 - **Durchschnitt von $r > .30$** bei verschiedenen Meta-Analysen mit unterschiedlichen Kriterien (Beförderung, Beurteilung durch Vorgesetzte) (vgl. Schmidt-Atzert & Amelang, 2012, S.467)
 - **Neuere Studien zeigen schlechtere Ergebnisse:** Negative Korrelation zwischen Validierungsevidenz und Alter der Studie ($r = -0.25$; Hardison & Sackett, 2007)
- **Inkrementelle Validität** von ACs über Intelligenz und Persönlichkeit (Meriac et al., 2008; Melchers & Annen, 2010)

Validität?

„Das AC ist zur Spielwiese der **Laiendiagnostik** geworden. Dazu kommt: ACs arbeiten weit unter ihrer methodischen Kapazität. Sie werden vornehmlich von Nicht-Psychologen durchgeführt. Die Augenscheingültigkeit der eingesetzten Teilverfahren hat sich durchgesetzt. [...] Der Anforderungsbezug steht ebenso in Frage wie die vorgebliche Organisationsspezifität. **Ergebnisse zur psychometrischen Qualität werden gering bewertet oder gar nicht zur Kenntnis genommen.**“ Schuler, 2007, S.28

Anwendungsfall 1: Assessment Center

Anwendungsfall 2: Ambulatory Assessment

Validität?

„Der hier vorgeschlagene Weg ist der, das AC als multiples eignungsdiagnostisches Verfahren ernst zu nehmen. **Das heißt erstens, seine Entwicklung, Durchführung und Prüfung an den gleichen Standards zu orientieren, die generell in der psychologischen Diagnostik gelten.** Zweitens bedeutet es, die Multiplizität beim Wort zu nehmen. Multipel heißt, es müssen verschiedene Teilverfahren kombiniert werden. Ob Teilverfahren verschieden sind, bemisst sich nicht nur nach dem Augenschein, aus psychometrischer Sicht bemisst es sich nach ihrer inkrementellen Validität, also nach ihrem Validitätszuwachs untereinander.“ Schuler, 2007, S.29

Anwendungsfall 2: Ambulatory Assessment

- „Verwendung spezieller feldtauglicher, heute meist elektronischer Geräte und computer-unterstützter Erhebungsmethoden, um Selbstberichtsdaten, **Verhaltensbeobachtungsdaten**, psychometrische Verhaltensmaße, physiologische Messwerte sowie situative und Setting-Bedingungen im Alltag der Untersuchten zu erfassen“ Fahrenberg, Myrtek, Pawlik, & Perrez, 2007, S.13
- U.a. Experience Sampling, Ecological Momentary Assessment, Mobile Sensing
- Zeit- oder Ereignisabhängige Messung möglich

Im Diagnostik-Kontext „Beobachtung“ von

- **Sensiblen Verhalten**
 - z.B. Sexualverhalten im Kontext von Paartherapien
- **Alltagsverhalten** (Feld)
 - z.B. Konsumverhalten bei Sucht / Zwängen (Alkohol, Zigaretten, Nahrungsaufnahme)
- **Physiologischen Zuständen**
 - z.B. Blutdruck
- **Kontext**
 - Situationen in denen Verhalten auftritt
- **Diagnostischen Verläufen**
 - z.B. Therapie-Erfolg
- **Psychologischen Dynamiken**
- Auch Präventionen denkbar, z.B. Veränderung der soziale Einbettung bei Patientinnen mit Depressionen erkennen

Anwendung bei Interventionen

“Wichers et al. (2011) recently introduced a promising hardware device (PsyMate) for AA. It is designed to be used in standard clinical care for **enhancing self-management by giving patients a visualization of their symptoms and their symptoms’ context dependency**. This feedback is thought to actively engage patients in the process of recovery and increase the patient’s insight into the nature of the emotional dysregulation. In addition, **clinicians using this device in their practice might get a better understanding of the nature of the individual’s depression and increased insight in how medication impacts daily life mood states in the individual, and this information may better guide decisions on medication type or dosage.**” Trull & Ebner-Priemer, 2014, S.165

Vorteile von Ambulatory Assessment

- Beobachtung im Feld (sogenannte „Ökologische Validität“)
- Interaktive, kontextbezogene Datenerfassung (Einbezug anderer Informationen)
- Prozessdiagnostik möglich
- Vermeidung von Retrospektionseffekten
- Zuverlässiges zeitliches Protokoll der Eingaben (einschließlich der Antwortlatenzen) durch automatische Alarm- und Erinnerungsfunktion
- Automatische Datensicherung (Fehlerkontrolle), Auslesen der Daten für statistische Weiterverarbeitung

Nachteile von Ambulatory Assessment

- Großer Aufwand
 - Reaktivität bleibt vorhanden, z.T. sogar als (vorteilhafte) Intervention
- Bisher größtenteils Anwendung im Forschungskontext

Leitfragen zur Nachbereitung

1. Welche Ziele hat die Verhaltensbeobachtung?
 - Welche Schritte sollte man bei der Verhaltensbeobachtung differenzieren und warum?
 - Welche Aspekte sollte man bei der Erstellung eines Untersuchungsplans beachten?
 - Was ist die CIT? Wie funktioniert sie?
2. Anhand welcher Kriterien kann man die Güte von Indikatoren überprüfen?
 - Pro Kriterium: Überlegungen, Statistische Analysen, Verbesserung?
3. Welche Entscheidungen muss man bei der Festlegung von Indikatoren treffen?
4. Welche Schritte muss man bei der Festlegung von Beurteilungsregeln beachten?
5. Welche Beobachtungs- und Beurteilungsfehler können auftreten?
6. Welche Beispiele für Verhaltensbeobachtung in der Anwendung gibt es und wie sehen diese aus?



Quellen

- ❖ Fahrenberg, J., Leonhart, R., & Foerster, F. (2002). *Alltagsnahe Psychologie: Datenerhebung im Feld mit hand-held PC und physiologischem Mess-System*. Huber.
- ❖ Fahrenberg, J., Myrtek, M., Pawlik, K., & Perrez, M. (2007). Ambulantes assessment-Verhalten im Alltagskontext erfassen. *Psychologische Rundschau*, 58(1), 12-23.
- ❖ Gaugler, B. B., Rosenthal, D. B., Thornton, G. C., & Bentson, C. (1987). Meta-analysis of assessment center validity. *Journal of applied psychology*, 72(3), 493.
- ❖ Hardison, C. M., & Sackett, P. R. (2007). Kriterienbezogene Validität des Assessment Centers: lebendig und wohlauf. *Assessment Center zur Potenzialanalyse*, 192-202.
- ❖ Höft, S., & Kersting, M. (2018). Anforderungsprofil, Verhaltensbeobachtung und Verhaltensbeurteilung. In *Personalauswahl kompetent gestalten* (pp. 27-63). Springer, Berlin, Heidelberg.
- ❖ Melchers, K. G., & Annen, H. (2010). Officer selection for the Swiss armed forces. *Swiss Journal of Psychology*.
- ❖ Meriac, J. P., Hoffman, B. J., Woehr, D. J., & Fleisher, M. S. (2008). Further evidence for the validity of assessment center dimensions: A meta-analysis of the incremental criterion-related validity of dimension ratings. *Journal of Applied Psychology*, 93(5), 1042.
- ❖ Schmidt-Atzert, L., & Amelang, M. (2012). *Psychologische Diagnostik*. Springer Science & Business Media.
- ❖ Schuler, H., & Personalpsychologie, S. F. (2007). Spielwiese für Laien? Weshalb das Assessment-Center seinem Ruf nicht mehr gerecht wird. *Wirtschaftspsychologie aktuell*, 2, 27-30.
- ❖ Stemmler, G., & Margraf-Stiksrud, J. (2015). *Lehrbuch Psychologische Diagnostik*. Verlag Hans Huber.
- ❖ Paschen, M., Beenen, A., Turck, D., & Stöwe, C. (2013). *Assessment Center professionell: Worauf es ankommt und wie Sie vorgehen*. Hogrefe Verlag.
- ❖ Trull, T. J., & Ebner-Priemer, U. (2014). The role of ambulatory assessment in psychological science. *Current Directions in Psychological Science*, 23(6), 466-470.
- ❖ Ziegler, M., & Bühner, M. (2012). *Grundlagen der psychologischen Diagnostik*. Springer-Verlag.