

Grundlagen der Diagnostik

Lerneinheit 2

Einführung II



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.

Wahr-oder-Falsch Aussagen zur letzten Sitzung

1. Psychologische Diagnostik dient der Entwicklung allgemeiner Gesetzmäßigkeiten.
2. Projektive Verfahren gehören zur Kategorie psychometrische Persönlichkeitstests.
3. Bestandteil des diagnostischen Prozesses ist die Formulierung einer Fragestellung.
4. Im Rahmen der Modifikation können sowohl Personen als auch Bedingungen gezielt verändert werden.
5. Die Schweigepflicht gilt nur gegenüber Klienten und Klientinnen, nicht gegenüber Kolleginnen und Kollegen.

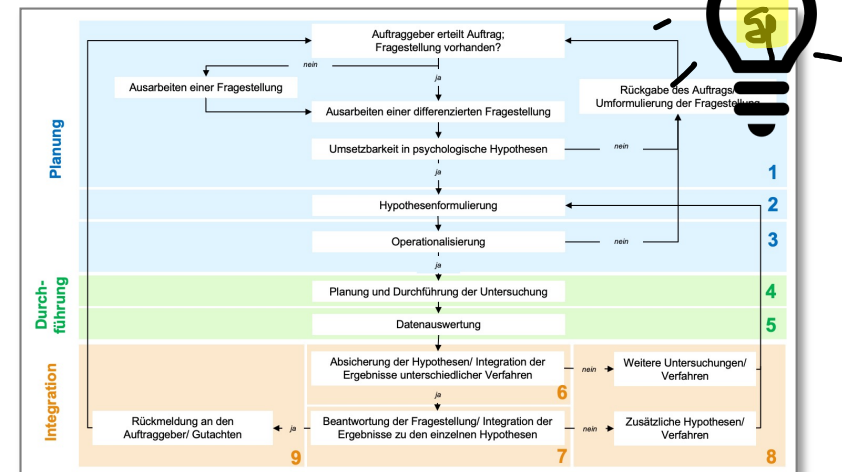
Einführung Teil II

Integrationsphase/Planungsphase

4. Arten diagnostischer Entscheidungen
5. Entscheidungsstrategien
6. Entscheidungsmodelle

Planungsphase

7. Gütekriterien diagnostischer Verfahren



Lernziele



4. Arten diagnostischer Entscheidungen

Entscheidungsfindung

Auf Basis diagnostischer Informationen wird i.d.R. eine **Entscheidung** über anstehende Fragen gefällt:

- Wird eine Bewerberin eingestellt?
- Welche Schülerinnen sollen welchen Unterrichtseinheiten zugeordnet werden?
- Ist eine Verhaltens- oder eine Gesprächstherapie indiziert?
- Kann eine Klientin aus einer geschlossenen in eine offene Abteilung überführt werden?

Arten diagnostischer Entscheidungen nach Cronbach und Gleser, 1965

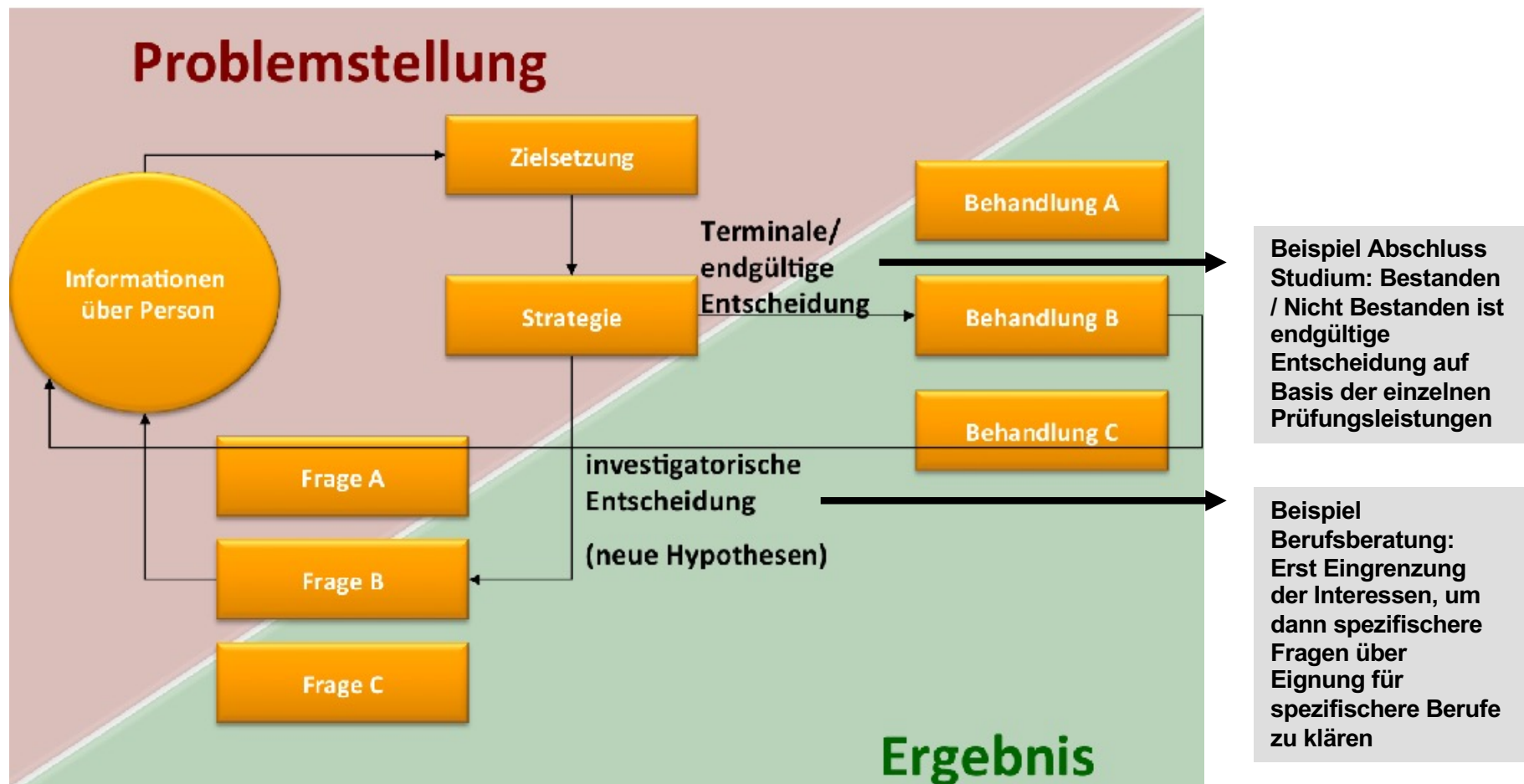
- **Nutzen** der Entscheidung: Institution vs. Individuum
- **Annahmequote**: festgelegt vs. variabel
- **Möglichkeit von Ablehnungen**: ja (Personenselektion) vs. nein (Platzierung oder Modifikation)
- **Informationsdimensionen**: univariat vs. multivariat
- **„Behandlungen“**: singular vs. multibel
(auch Untersuchungsschritte: einstufig vs. mehrstufig)
- **Entscheidungen**: terminal vs. investigatorisch → ...

... **Entscheidungen** nach Cronbach und Gleser, 1965

- Endgültig terminale Entscheidungen sind häufig auf Fälle beschränkt, in denen Personen aus institutionellen Systemen entlassen werden
- Persönliche (z.B. Studienwahl) und institutionelle (Einzel-) Entscheidungen (z.B. Vergabe Studienabschluss) sind
 - meist sinnvollerweise einzureihen in eine lange Kette von Entscheidungen (d.h. zunächst v.a. investigatorische Entscheidungen kombiniert mit mehrstufigem Vorgehen)
 - Dennoch: In der diagnostischen Praxis kommt aus Zeit- und Kostengründen das einstufige Vorgehen auch noch häufig vor

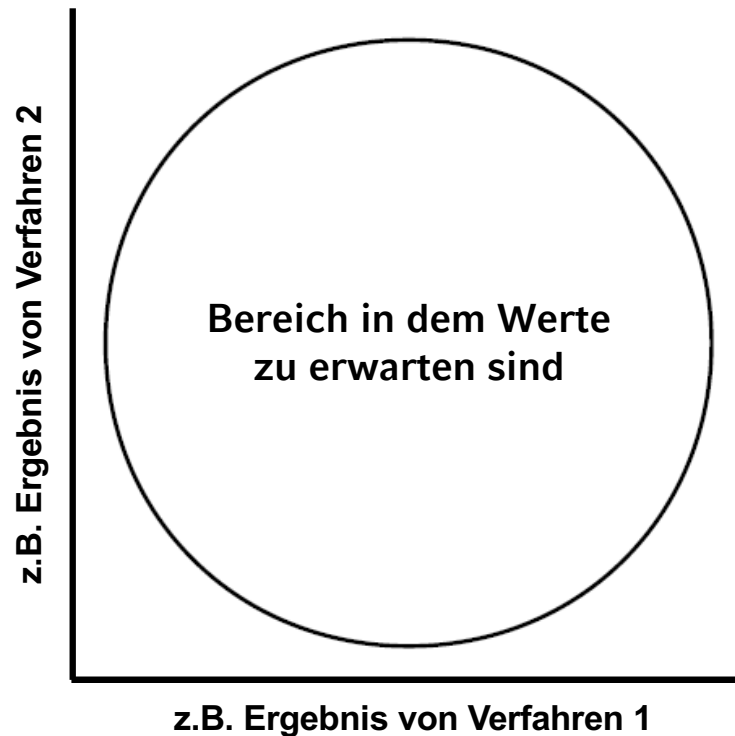
... Entscheidungen

→ Regelkreismodell des Entscheidungsprozesses (nach Tack, 1976)



5. Entscheidungsstrategien

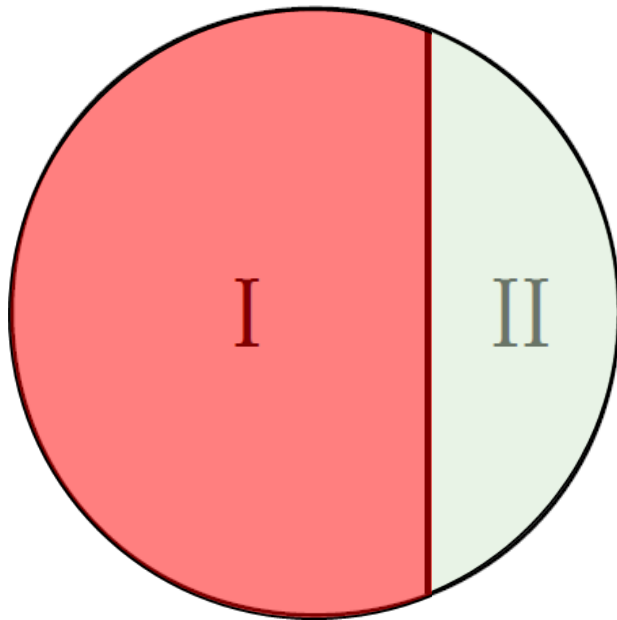
Strategien diagnostischer Entscheidungen



- **(1) Einstufige Strategien** basierend auf einer Datenerhebung (ggf. mehrerer Verfahren) **zu einem Zeitpunkt für alle Personen**, direkt gefolgt von einer Entscheidung für alle
- **(2) Mehrstufige Strategien** basieren auf Datenerhebungen **zu mehreren Zeitpunkten teilweise nur für Teilgruppen**, basierend auf Zwischenentscheidungen
- Der Untersuchungsplan muss je nach gewählter Strategie entsprechend angepasst werden!

(1) Einstufige Strategien

Single Screen (Einzelteststrategie)



Ergebnis von Verfahren 1

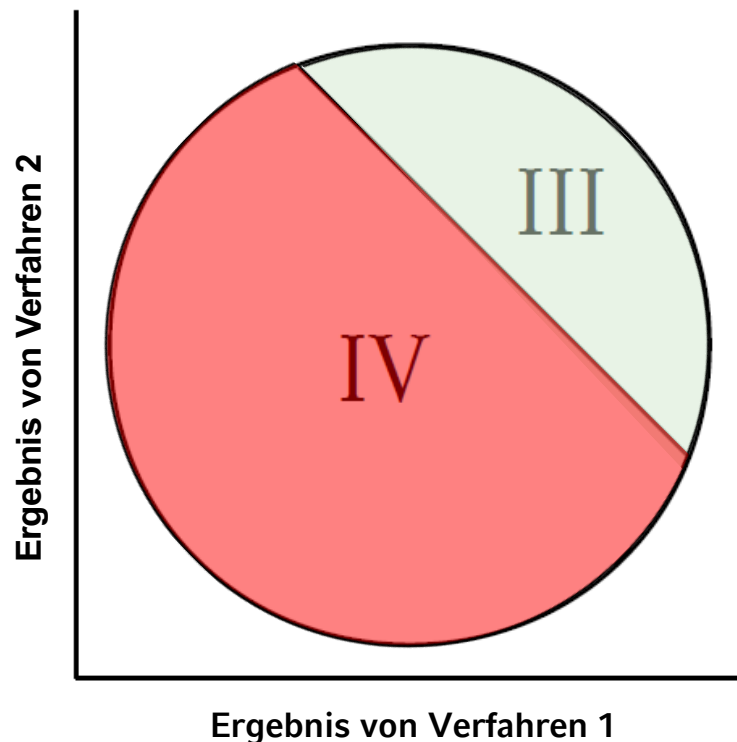
Ablehnungsbereich = I

Annahmebereich = II

- Entscheidung beruht auf dem Ergebnis von nur **einem Verfahren** von allen Personen
- Beispiel: schriftlicher Sprachtest für ein bestimmtes Sprachniveau

(1) Einstufige Strategien

Nichtsequentielle Untersuchungsatterie



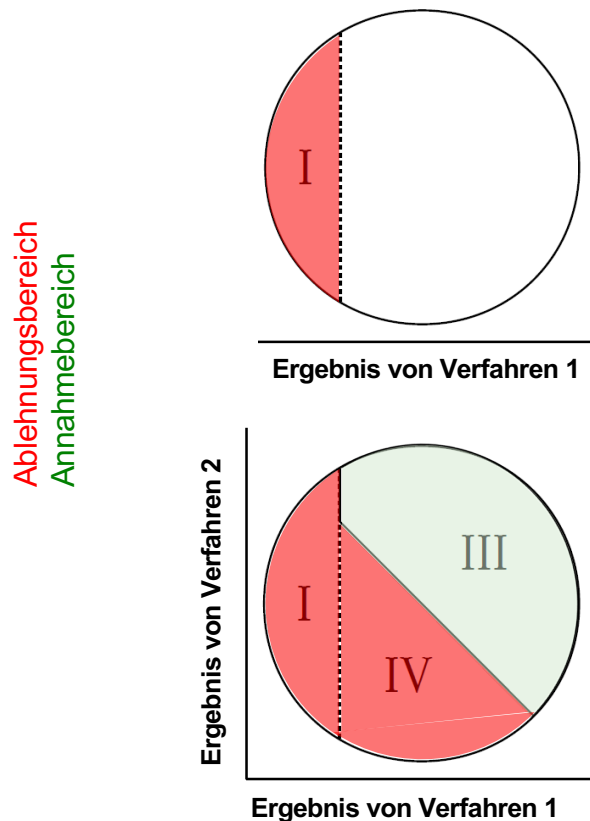
Ablehnungsbereich = IV

Annahmebereich = III

- **Eine Batterie von diagnostischen Verfahren** wird allen Personen vorgegeben
- Ausgewählt werden die Personen, die im zugrunde gelegten Entscheidungsmodell am besten abschneiden (z.B. gewichteter Summenwert)
- Beispiel: offener Assessment-Tag

(2) Mehrstufige Strategien

Vorauswahl- oder Pre-reject-Strategie



Schritt 1:

Nach erster Datenerhebung werden Probandinnen, die bestimmten Wert nicht erreichen, ausgeschlossen (**Bereich I**)

Schritt 2:

Die verbleibenden Probandinnen absolvieren weitere Verfahren, eine Entscheidung über Annahme (**Bereich III**) oder Ablehnung (**Bereich IV**) erfolgt aus der Kombination zwischen Erst- und Folgeverfahren

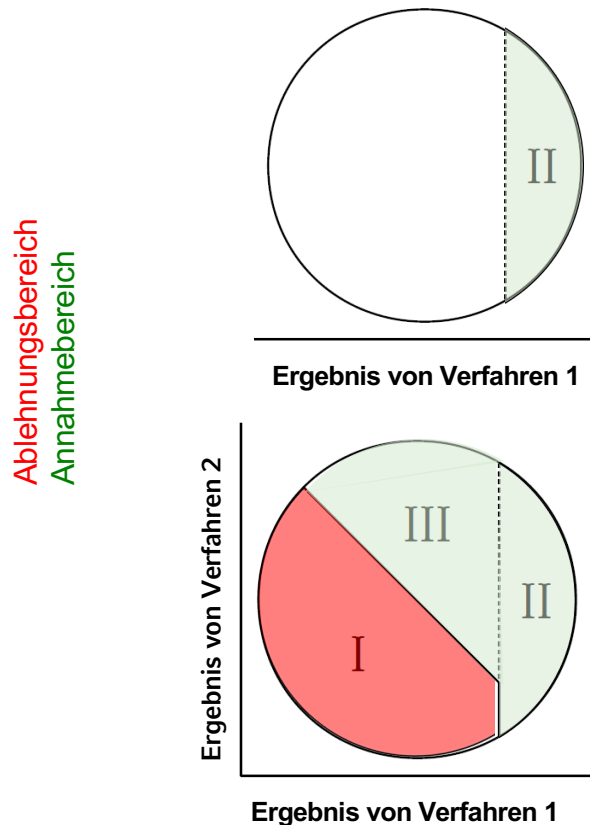
(2) Mehrstufige Strategien

Vorauswahl- oder Pre-reject-Strategie

Beispiel: Bei der „Studienstiftung des Deutschen Volkes“ werden nur Kandidatinnen mit herausragenden schulischen (oder universitären) Leistungen nominiert und anschließend sequentiell getestet

(2) Mehrstufige Strategien

Vorentscheidungs- oder Pre-accept-Strategie



Schritt 1:

Nach erster Datenerhebung werden Probandinnen, die bestimmten Wert überschreiten, bereits (terminal) akzeptiert (**Bereich II**)

Schritt 2:

Mit den verbleibenden Probandinnen wird analog zur Vorauswahlstrategie verfahren, eine Entscheidung über **Annahme (Bereich III)** oder **Ablehnung (Bereich I)** erfolgt aus der Kombination zwischen Erst- und Folgeverfahren

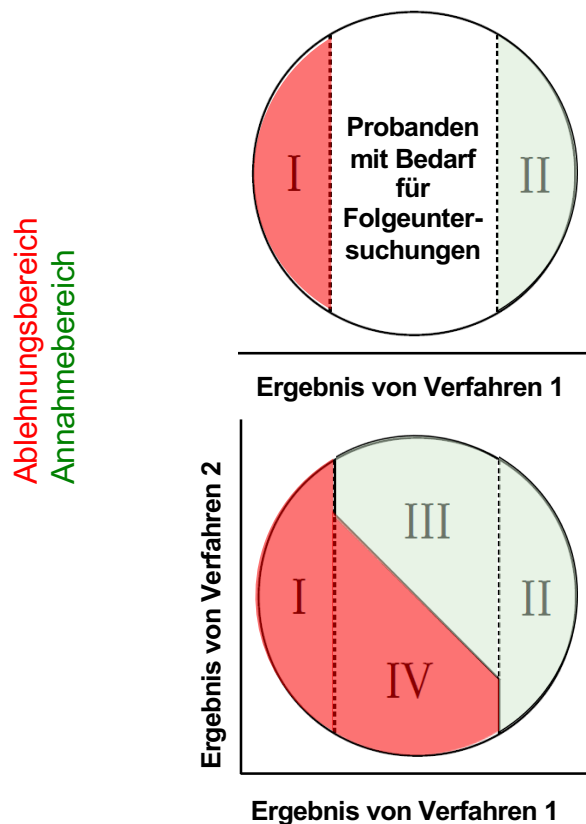
(2) Mehrstufige Strategien

Vorentscheidungs- oder Pre-accept-Strategie

Beispiel: Bei der Zulassung zum Medizinstudium wird ein festgelegtes Kontingent zentral über die Abiturnote (bzw. die Wartezeit) vergeben, die verbleibenden Plätze werden von den Hochschulen auf Basis (der Kombination) von Einzelnoten, einem fachspezifischen Studierfähigkeitstest (TMS) oder weiteren Kriterien vergeben

(2) Mehrstufige Strategien

Vollständig sequentielle Strategie



Kombination aus Vorauswahl- und Vorentscheidungsstrategie:

Schritt 1:

Nach Ergebnissen in einem diagnostischen Verfahren erfolgt eine Aufteilung der Probandinnen in drei Gruppen: (Terminale) **Ablehnung (Bereich I)** und **Annahme (Bereich II)** sowie **Bedarf für Folgeuntersuchungen**

Schritt 2:

Die Entscheidung über **Annahme (Bereich III)** oder **Ablehnung (Bereich IV)** dieser Probandinnen erfolgt erneut aus der Kombination zwischen Erst- und Folgevverfahren

(2) Mehrstufige Strategien

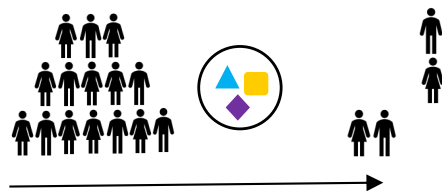
Vollständig sequentielle Strategie

Beispiel: Unternehmen, welches für einen bestimmten Job...

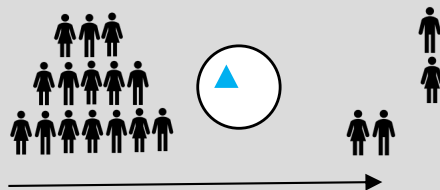
- Bewerberinnen ohne Abschluss in einem bestimmten Fach nicht einlädt
- Bewerberinnen mit einer Mindestnote in einem bestimmten Abschluss direkt einstellt
- mit den restlichen Bewerberinnen ein Assessment-Center durchführt

Zusammenfassung & Unterscheidung der Begriffe

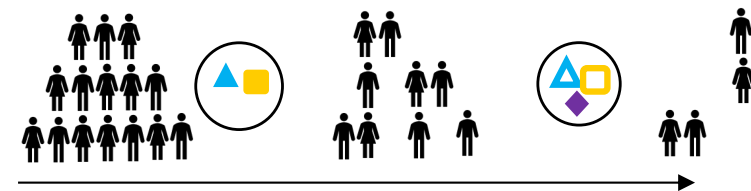
nichtsequentiell



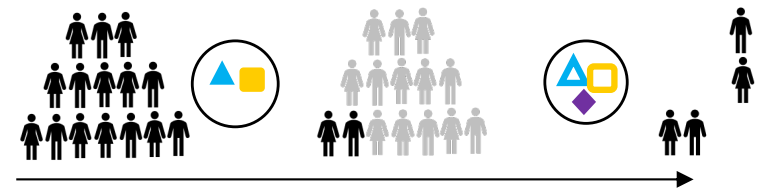
Achtung: wenn nur ein Verfahren eingesetzt wird,
würde man nicht von nichtsequentieller Strategie,
sondern von **Single Screen** sprechen:



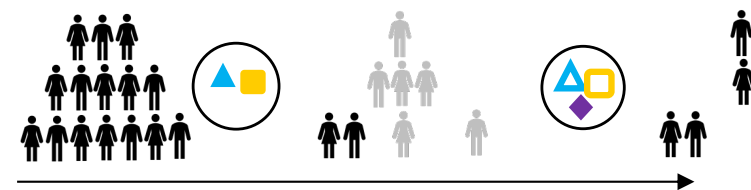
sequentiell



Vorauswahl-Strategie



Vorentscheidungs-Strategie

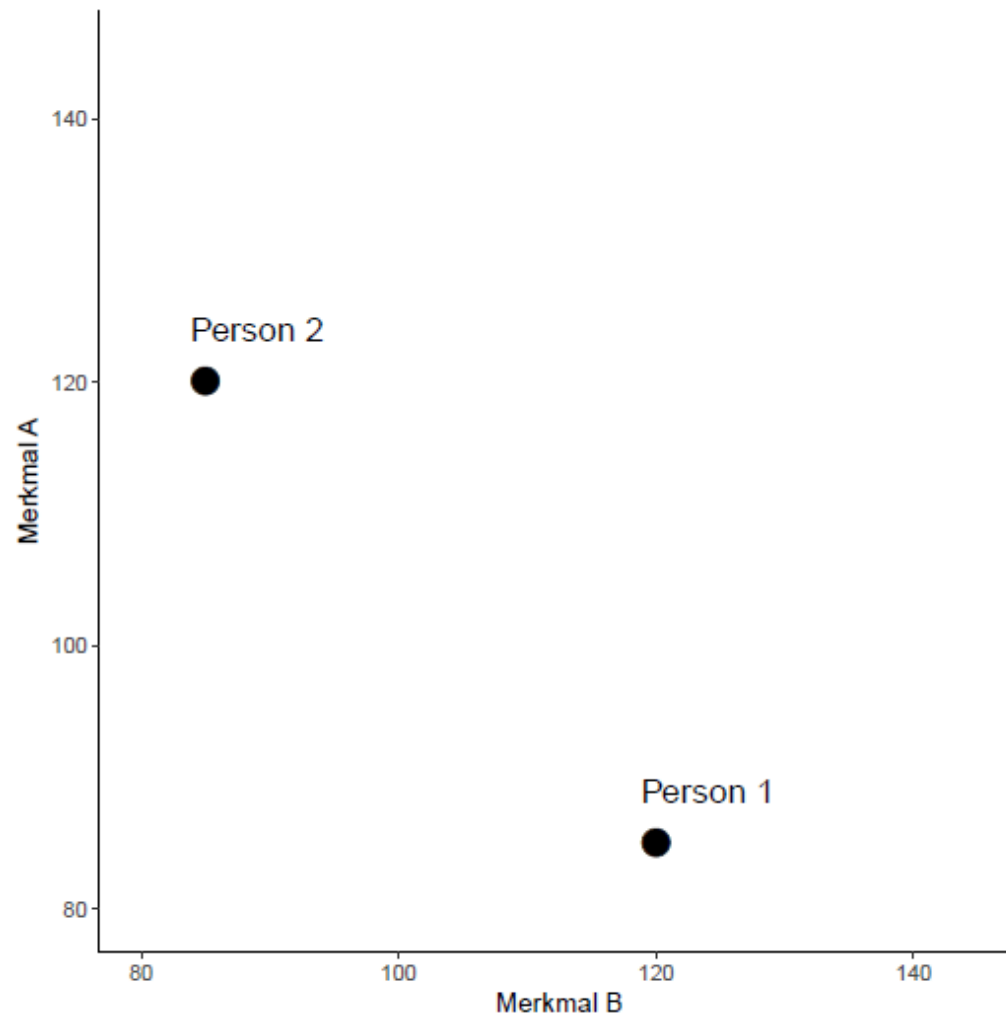


Vollständig-sequentielle Strategie

6. Entscheidungsmodelle

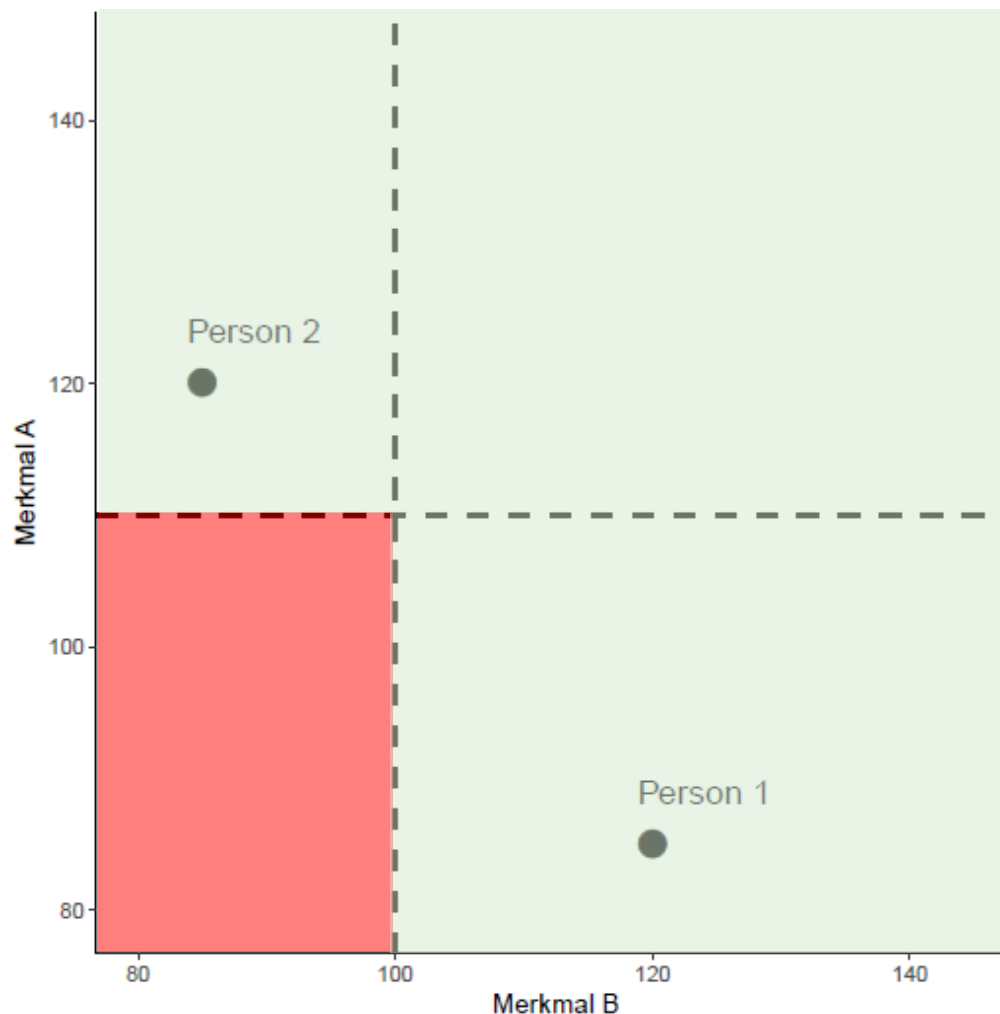
Entscheidungsmodelle = Art der Kombination

- Liegen die Ergebnisse mehrerer diagnostischer Verfahren vor („Nichtsequentielle Batterie“ und mehrstufige Verfahren), handelt es sich i.d.R. um mehrdimensionale, d.h. multivariate diagnostische Informationen
- Diese müssen **in geeigneter Weise kombiniert** werden, um eine Entscheidungsfindung zu ermöglichen
- Es lassen sich drei grundlegende Kombinationsmöglichkeiten bzw. Entscheidungsmodelle unterscheiden:
 - (1) **Disjunktives Modell („Oder“-Strategie)**
 - (2) **Konjunktives Modell („Und“-Strategie)**
 - (3) **Additives/Kompensatorisches Modell**



Zur Veranschaulichung werden hier nur 2 Merkmale dargestellt. Die Entscheidungsmodelle sind aber auch auf mehr als 2 Merkmale anwendbar.

(1) Disjunktives Modell – „Oder“-Strategie



- Mindestanforderungen in beiden Merkmalen, aber Probandin muss diese Anforderungen nur in einem Merkmal erfüllen
- Das heißt, für eine positive Bewertung genügt ein entsprechend hoher Punktwert in einem der beiden Merkmale

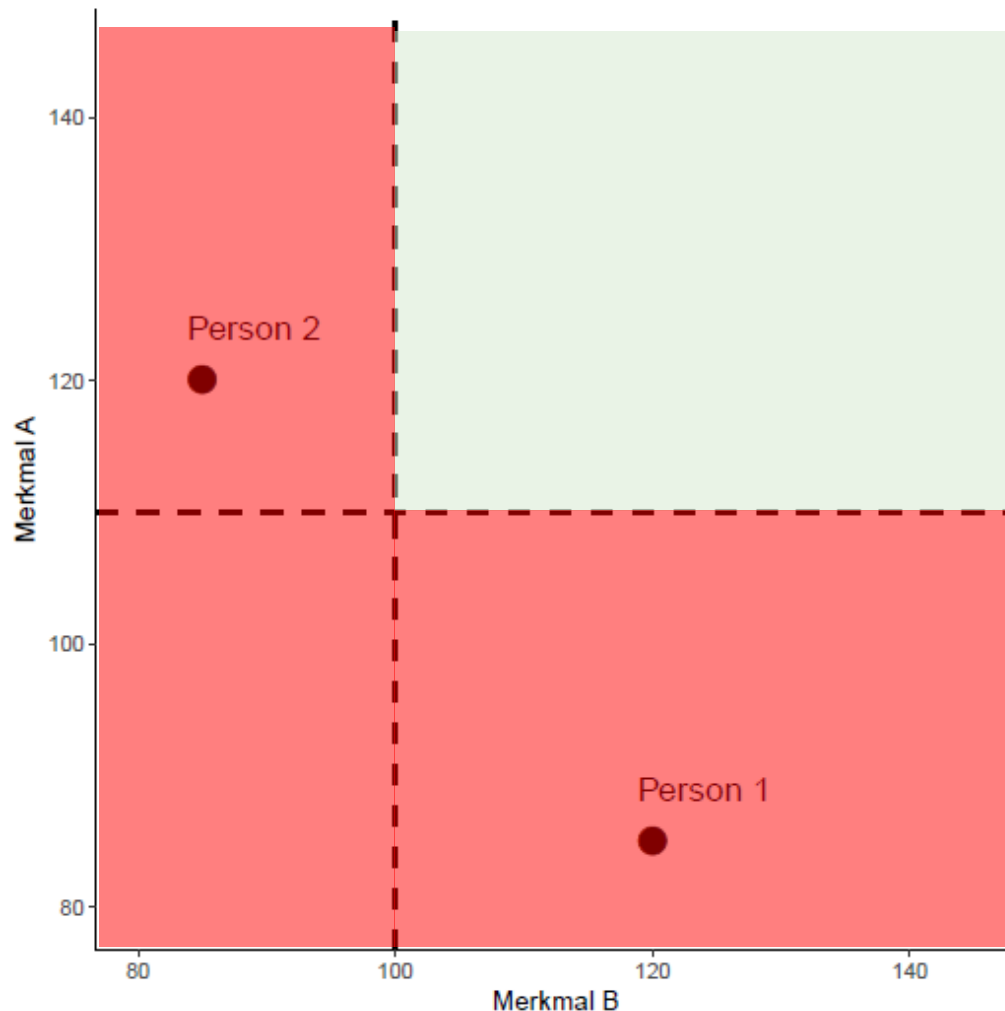
(1) Disjunktives Modell – „Oder“-Strategie

- Angemessen, wenn die geforderte Leistung entweder auf die eine oder andere Weise erbracht werden kann
- Beispiel 1 (→ LE1):

Hypothese	Verfahren	Entscheidungsmodell: Hypothesen	
Verfügt ein*e Bewerber*in...		Konjunktiv	Disjunktiv
... im Vergleich zu ähnlich akademisch ausgebildeten Personen mindestens durchschnittliche Englischkenntnisse?	Zeugnis / Arbeitsprobe		Sprachtestniveau mindestens B1 oder in Arbeitsprobe mindestens durchschnittlich im Vgl. zur Norm

- Beispiel 2: Einladung von Bewerberinnen für einen Job in der Personalabteilung, wenn ein abgeschlossenes Studium in Psychologie oder BWL vorgewiesen wird

(2) Konjunktives Modell – „Und“-Strategie



- Mindestanforderungen in beiden Merkmalen müssen erfüllt werden
- Das heißt, für eine positive Bewertung ist ein entsprechend hoher Punktwert in beiden Merkmalen notwendig

(2) Konjunktives Modell – „Und“-Strategie

- Angemessen, wenn die geforderte Leistung nicht selektiv erbracht oder kompensiert werden kann
- Beispiel 1 (→ LE1):

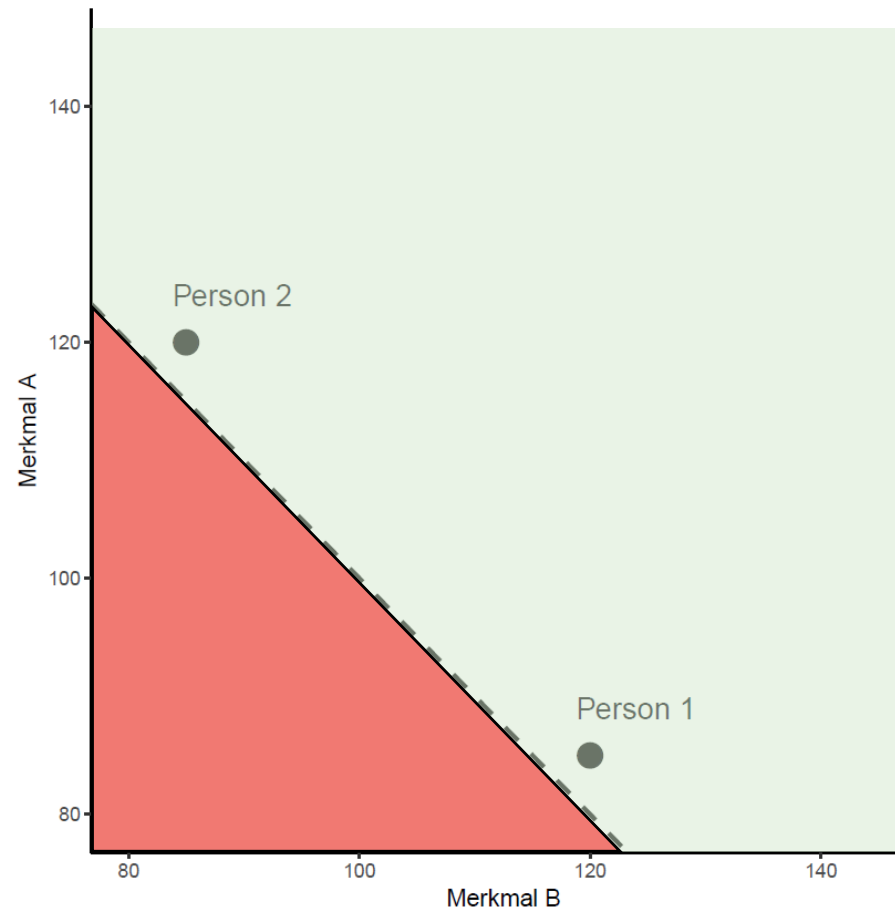
Hypothese	Verfahren	Entscheidungsmodell: Hypothesen		Entscheidungsmodell: Urteilsfindung	
Verfügt ein*e Bewerber*in...		Konjunktiv	Disjunktiv	Konjunktiv	Kompensatorisch
... im Vergleich zu anderen Psychologieabsolventen über überdurchschnittliches Wissen im Bereich Diagnostik?	Zeugnis / Interview	Modulnote mindestens 1.7 und im Interview 3 von 4 Fragen richtig beantwortet		X (kann nicht kompensiert werden)	

- Beispiel 2: mangelnde Sehfähigkeit kann bei angehenden Pilotinnen nicht ausgeglichen werden

(3) Additives/kompensatorisches Modell

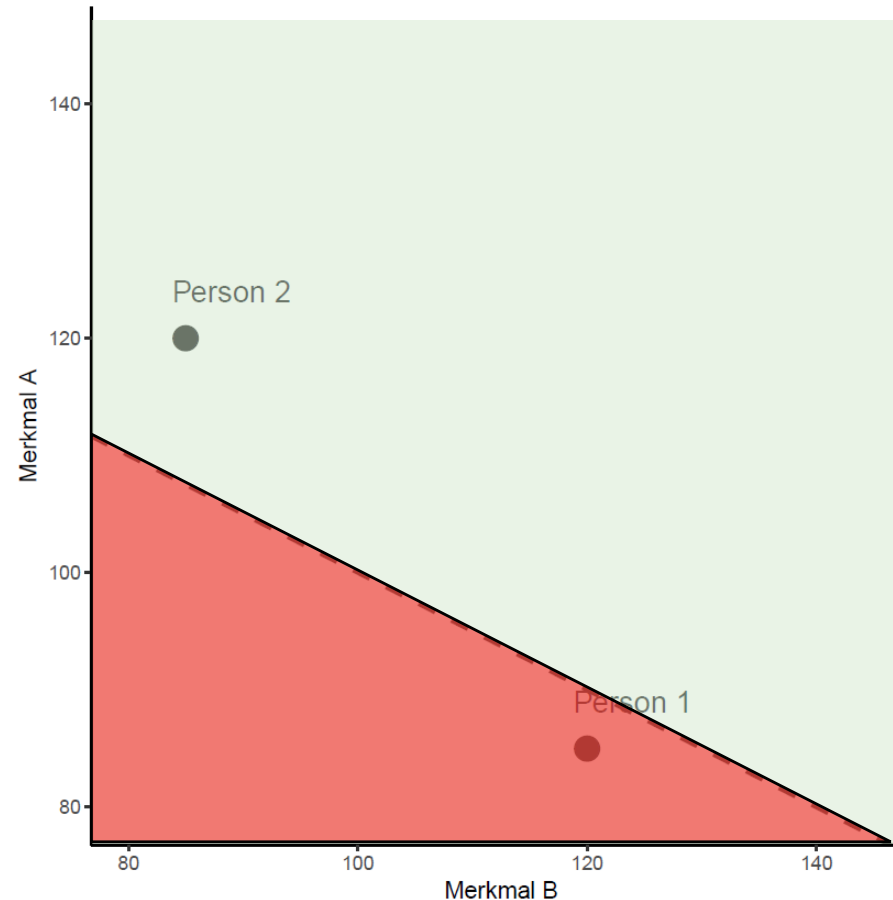
- Mittlerweile am häufigsten eingesetzte Entscheidungsstrategie: Annahme, dass sich die Merkmale/Merkmalssausprägungen (bis zu einem gewissen Grad) gegenseitig ausgleichen können
- Ermittlung eines Gesamtwerts auf Basis der (gewichteten) Merkmale, z.B. mithilfe einer linearen Regression → empirische Datengrundlage wünschenswert
- Der selbe Gesamtwert kann durch verschiedene Merkmalskonfigurationen in den Einzelerhebungen erreicht werden
- Beispielsweise können niedrige Leistungen in einem Test durch hohe Leistungen in anderen Tests ausgeglichen bzw. kompensiert werden

(3) Additives/kompensatorisches Modell



Einfaches Modell: $\text{Summenwert} > 200$
(Gleichgewichtung der Merkmale)

(3) Additives/kompensatorisches Modell



Lineares Regressionsmodell
(mit schwächerer Gewichtung von Merkmal B)

Entscheidungsmodelle = Art der Kombination

→ Konsequenzen für die diagnostische Praxis:

- Strategien vergleichen: Konjunktives Modell ist i.d.R. am strengsten (Bild des „Hürdenlaufs“: Wenn man zu viele (schwere) Anforderungen konjunktiv verknüpft ist am Ende nur noch ein sehr kleiner Pool an geeigneten Personen übrig)
- Vor der Entscheidung für eine Strategie die gegebenen Anforderungen beachten und Konsequenzen einer möglichen Fehlentscheidung prüfen (Fehler 1. und 2. Art)
- Mögliche Konsequenzen für die Fragestellung transparent machen
- Mischformen in Erwägung ziehen

7. Gütekriterien

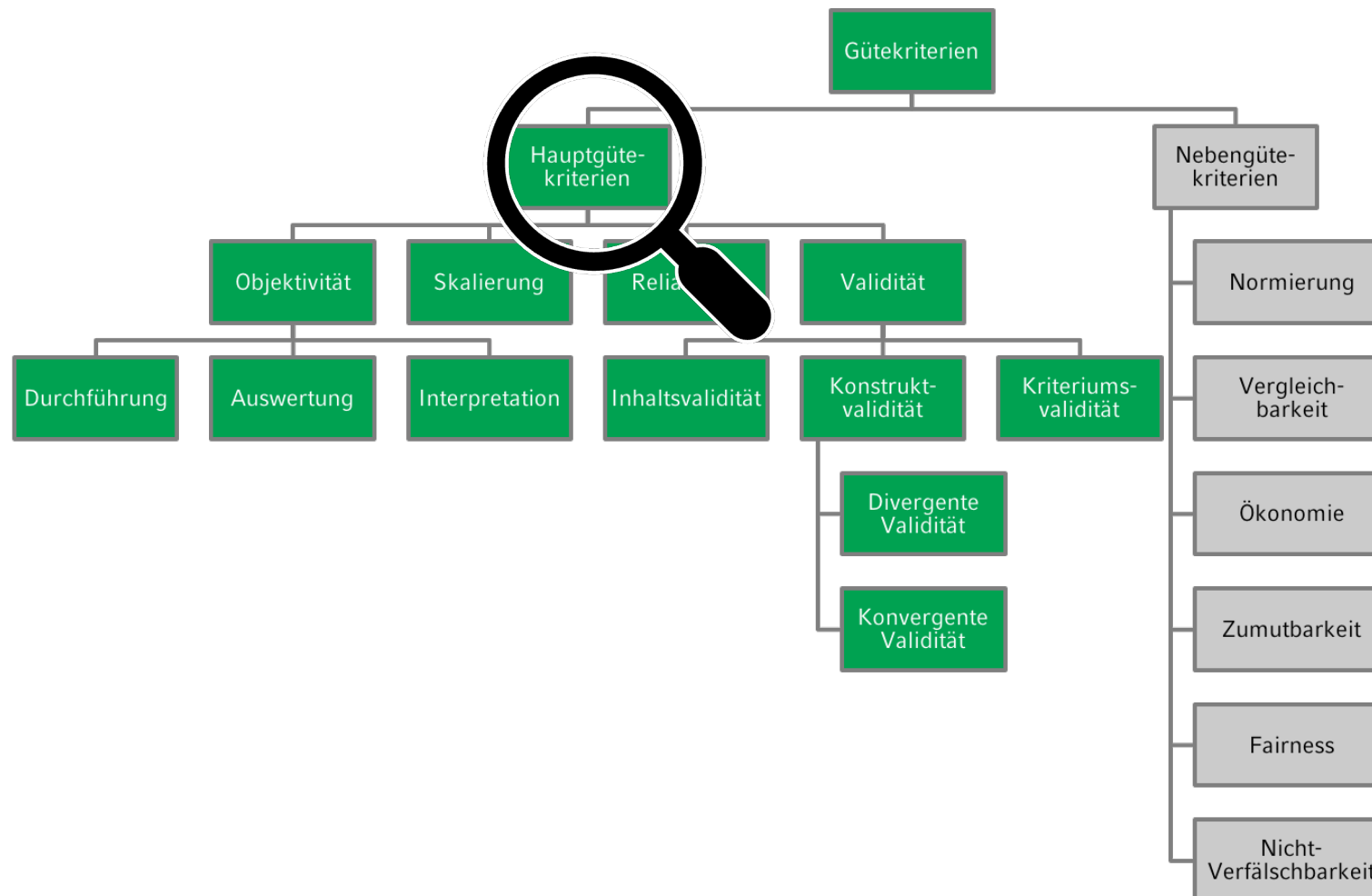
Haupt- und Nebengütekriterien

Wie lässt sich die Qualität diagnostischer Verfahren beurteilen?

- **Gütekriterien** als relevante Qualitätsmerkmale psychologischer Tests (und anderer diagnostischer Verfahren)
- Relevanz im Kontext der Testkonstruktion (vgl. Vorlesung Psychologische Testtheorie) sowie bei der Operationalisierung im Rahmen des diagnostischen Prozesses

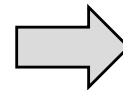
→ Zur Auffrischung/Vertiefung wird ein Blick in die Folien der VL Psychologische Testtheorie aus dem WiSe empfohlen

Haupt- und Nebengütekriterien



Hauptgütekriterien: Übersicht

- **Objektivität**
 - Durchführungsobjektivität
 - Auswertungsobjektivität
 - Interpretationsobjektivität
- **(Skalierung)**
- **Reliabilität**
- **Validität**
 - Inhaltsvalidität
 - Kriteriumsvalidität
 - Konstruktvalidität



Ein Testmanual muss Belege dazu vorweisen, inwiefern diese Gütekriterien erfüllt sind – und für empirische Untersuchungen zum Nachweis der Gütekriterien qualitativ hochwertige Stichproben heranziehen (ausreichend groß, ausreichend repräsentativ)!

In einem *vollständigen* Testmanual findet man also Informationen zu...

- Motivation, Zielsetzung / Messanspruch des Tests
- Theoretischem Hintergrund des Tests & Literatur
- Testmaterial und Durchführungsmöglichkeiten
- Hauptgütekriterien
 - Objektivität
 - Reliabilität
 - Validität
- Auswertung und Interpretationen
- Normen

Objektivität

Skalierung

Reliabilität

Validität

Definition Objektivität:

Ein diagnostisches Verfahren ist dann objektiv, wenn die
(1) Durchführung und **(2) Auswertung** des Verfahrens sowie die
(3) Interpretation der resultierenden Ergebnisse nicht variieren, auch
wenn unterschiedliche Personen das Verfahren zum Einsatz bringen.

→ Es gibt also drei Teilaspekte der Objektivität:

- (1) Durchführungsobjektivität
- (2) Auswertungsobjektivität
- (3) Interpretationsobjektivität

Objektivität

Skalierung

Reliabilität

Validität

(1) Durchführungsobjektivität

Standardisierung der Untersuchungsbedingung in Bezug auf z.B.:

- Instruktion
- Testsituation (z.B. Einzel- und Gruppentestungen, Abstände, Zeitbegrenzung)
- Material, Hilfsmittel
- Umgang mit Nachfragen, Störungen im Testablauf
- Anforderungen an Probandinnen & Testleiterinnen
- Interviewleitfaden bzw. Beobachtungsplan

→ Im Optimalfall: Testperson ist einzige Variationsquelle

Objektivität

Skalierung

Reliabilität

Validität

(2) Auswertungsobjektivität

Standardisierung der Auswertung, z.B.

- Exakte Auswertungsregeln (z.B. zur Bildung von Itemsummen/-mittelwerten, v.a. bei offenen Antworten)
- Computergestützte Auswertung
- Auswertungshilfen (z.B. Schablonen)
- Festgelegte Ausschlusskriterien
- Informationen zum Umgang mit fehlenden Werten
- Festlegung von Antwortmöglichkeiten, z.B. bei Interviews
- Verhaltensverankerte Ratingskalen, z.B. bei Verhaltensbeobachtungen
- Training von Beobachtern und Beurteilern
- Beobachter- und Beurteilerübereinstimmungsmaße (→ LE4-6)
- Berechnung von Konfidenzintervallen (→ LE9)

Objektivität

Skalierung

Reliabilität

Validität

(3) Interpretationsobjektivität

Standardisierung der Interpretation, z.B.

- Vorgaben zur Benennung und Beschreibung des erhobenen Merkmals, sowie der Bedeutung seiner Ausprägungen
- Informationen zu Normstichproben
- Hinweise auf Interpretation auf Basis von Konfidenzintervallen und Klassifikation in Kategorien (z.B. „durchschnittlich“)
- Fallbeispiele
- ...

Objektivität

Skalierung

Reliabilität

Validität

Definition Skalierung:

Ein diagnostisches Verfahren gilt als skalierbar, wenn die Zuordnung der Messwerte zu den Personen auf der Basis eines empirisch nachgewiesenen Testmodells geschieht.

→ siehe Vorlesung Psychologische Testtheorie!

Nebenbemerkung: Wird häufig auch als Teilaspekt der „Strukturellen Validität“ bezeichnet

Objektivität

Skalierung

Reliabilität

Validität

Definition **Reliabilität**:

Die Reliabilität beschreibt die Genauigkeit, mit der ein Verfahren ein Merkmal erfasst. Ein Test ist dann reliabel (zuverlässig), wenn er das Merkmal, das er misst, exakt, d.h. ohne Messfehler, misst („Messgenauigkeit“).

→ Die Reliabilität eines Items ist der Anteil der Varianz des systematischen Teils der Itemantwort an der gesamten Varianz der Itemantwort:

$$REL(X_i) = \frac{VAR(\tau_i)}{VAR(X_i)} = \frac{VAR(\tau_i)}{VAR(\tau_i) + VAR(\varepsilon_i)}$$

Nebenbemerkung: in der Diagnostik-VL interessieren wir uns weniger für die Reliabilität einzelner Items i , sondern für die Reliabilität eines Tests (d.h. des Itemmittelwertes oder des Itemsummenwertes, siehe auch Vorlesung Psychologische Testtheorie!)

Objektivität

Skalierung

Reliabilität

Validität

- Die Reliabilität ist eine wichtige Einflussgröße auf die Breite von Konfidenzintervallen (in der Einzelfalldiagnostik).
- Schätzung der Reliabilität:
 - Wann immer möglich wird eine Schätzung mittels Methoden der **internen/inneren Konsistenz** empfohlen (d.h. abhängig vom jeweils geltenden Testmodell durch Spearman-Brown, Cronbach's α oder McDonald's Omega)
 - Weitere Methoden:
 - **Retest-Reliabilität**
 - **Paralleltest-Reliabilität**
 - **Testhalbierungs-Reliabilität (Split-Half-Reliabilität)**
 - Wichtig: Welche Art der Reliabilitätsschätzung sinnvoll(er) ist, hängt auch von Eigenschaften des Konstrukts ab (z.B. Stabilität)!

Objektivität

Skalierung

Reliabilität

Validität

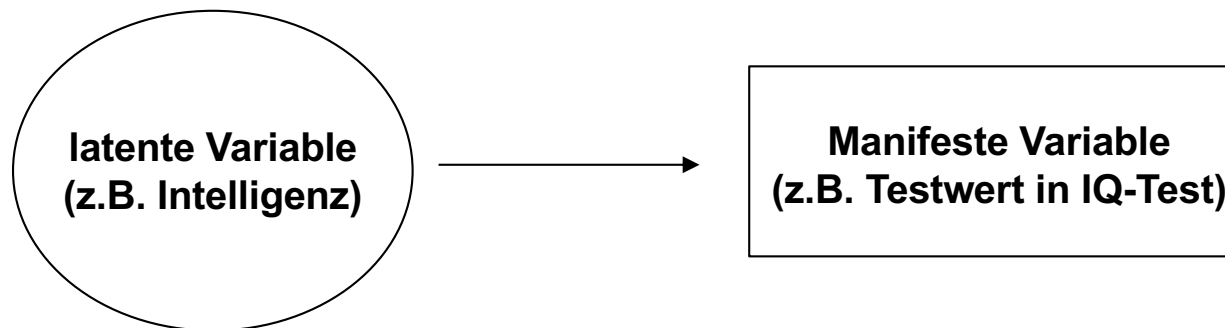
Definition **Validität**:

Die Validität gibt an, ob ein diagnostisches Verfahren auch wirklich das misst, was es zu messen beansprucht (und nichts anderes).

Relevante Frage:

Verursacht Variation im Merkmal eine Variation im Testwert?

Das (und nur das) ist Validität – es gibt nicht mehrere „Validitäten“.



→ Bei Validität geht es um Kausalität! (d.h. beobachtbare Unterschiede in den Testwerten gehen auf Unterschiede im zugrundeliegenden latenten Merkmal zurück)

Objektivität

Skalierung

Reliabilität

Validität

ABER: es gibt verschiedene Arten der Überprüfung der Validität (eines Tests/diagnostischen Verfahrens) / Validitätslabels:

(siehe auch VL11/ Testtheorie)

- **Inhaltsvalidität**
- **Konstruktvalidität**
- **Kriteriumsvalidität**

Arbeitsdefinition zur Überprüfung der Validität eines Tests (VL Testtheorie):

Testung von **Hypothesen** über den Zusammenhang der latenten Variable(n) des Tests **untereinander**, sowie mit **anderen** Variablen (z.B. manifeste Variablen wie Alter; latente Variable(n) anderer Tests) im **nomologischen Netzwerk**.

Objektivität

Skalierung

Reliabilität

Validität

Exkurs zur Unterscheidung von Begriffen

Klassischerweise 2 Teilaspekte im **Kontext der Empirie**:

(aka „Sind die auf Basis der Ergebnisse von Studien gezogenen Schlüsse gültig?“)

- **Interne Validität**
 - Variation in der AV kann eindeutig auf Variation in UV zurückgeführt werden
 - Ausschluss von Alternativerklärungen
- **Externe Validität**
 - Frage nach Generalisierbarkeit (über Personen und Situationen hinweg)
 - Ein Aspekt: ökologische Validität

Objektivität

Skalierung

Reliabilität

Validität

Inhaltsvalidität

- Ist gegeben, wenn die Bestandteile eines Verfahrens (z.B. die Items eines Tests) das zu messende Merkmal hinreichend und in seiner kompletten Breite repräsentativ erfassen
- Hinweise auf Vorliegen der Inhaltsvalidität:
 - Empirisch fundierte theoretische Modelle: Abgleich der Items mit Arbeitsdefinition
 - Expertenbefragung
 - Nachvollziehbare Beschreibung, wie die Aufgaben der finalen Testversion aus einer größeren Itemmenge ausgewählt wurden
 - Vergleich mit bestehenden Verfahren (Vorsicht: Zirkelschluss!)

Objektivität

Skalierung

Reliabilität

Validität

Konstruktvalidität

- Erwartungsgemäße empirische Belege für Zusammenhänge zwischen den latenten Variablen innerhalb eines/mehrerer Tests werden häufig unter dem Begriff der Konstruktvalidität zusammengefasst
- Unterscheidung:
 - Konvergente Validität
 - Diskriminante / divergente Validität
 - Faktorielle Validität

Objektivität

Skalierung

Reliabilität

Validität

Konstruktvalidität (so wie sie im diagnostischen Prozess häufig verstanden wird)

- **Konvergente Validität:**
 - Wie stark hängt ein Testergebnis (als Ausprägung eines Merkmals) mit weiteren Indikatoren (andere Methoden/Tests) **des gleichen Merkmals** (gleicher oder ähnlicher Messanspruch) zusammen?
 - Für den Nachweis der konvergenten Validität würde man entsprechend hohe Korrelationen erwarten
- **Diskriminante Validität** (auch divergente Validität):
 - Wie stark hängt ein Testergebnis (als Ausprägung eines Merkmals) mit Indikatoren **anderer Merkmale** zusammen?
 - Für den Nachweis der divergenten Validität würde man entsprechend tendenziell niedrigere Korrelationen erwarten
 - Besonders aussagekräftig bei oberflächlich ähnlichen Konstrukten

Objektivität

Skalierung

Reliabilität

Validität

Konstruktvalidität

- Mit der sogenannten „**Multitrait-Multimethod-Methode**“ können diese Erwartungen systematisiert und entsprechende empirische Belege verordnet werden

Tab. 2.21 Multitrait-Multimethod-Matrix

		Methode 1			Methode 2			Methode 3		
		Trait 1	Trait 2	Trait 3	Trait 1	Trait 2	Trait 3	Trait 1	Trait 2	Trait 3
Methode 1	Trait 1	(Rel.)								
	Trait 2		(Rel.)							
	Trait 3			(Rel.)						
Methode 2	Trait 1				(Rel.)					
	Trait 2					(Rel.)				
	Trait 3						(Rel.)			
Methode 3	Trait 1							(Rel.)		
	Trait 2								(Rel.)	
	Trait 3									(Rel.)

Anmerkungen.
 »Reliabilitätsdiagonale«: In der Hauptdiagonalen stehen die Reliabilitäten (Rel.) der Verfahren.
 Graue Felder = »Validitätsdiagonalen« (»monotrait-heteromethod«): Ein Merkmal wird mit verschiedenen Methoden gemessen.
 Blaue Felder = »Heterotrait-Monomethod-Dreiecke«: Verschiedene Merkmale werden mit der gleichen Methode erfasst.
 Alle weißen Felder unter der Reliabilitätsdiagonalen = »Heterotrait-Heteromethod-Dreiecke«: Korrelation zwischen verschiedenen Merkmalen, die zudem mit unterschiedlichen Methoden gemessen wurden.
 Die Felder über der Hauptdiagonalen bleiben leer.

Objektivität

Skalierung

Reliabilität

Validität

Konstruktvalidität

- **Faktorielle Validität:**
 - Erwartungsgemäße empirische Belege für Zusammenhänge zwischen mehreren latenten Variablen **eines Tests** werden in der Praxis häufig unter dem Begriff der faktoriellen Validität zusammengefasst
 - Teilweise wird hier auch die **Skalierung** verordnet: Erwartungsgemäße Faktorenstruktur des Tests

Objektivität

Skalierung

Reliabilität

Validität

Kriteriumsvalidität

- Ergebnis des diagnostischen Verfahrens soll in Zusammenhang stehen mit konkreten Leistungen oder Verhaltensweisen außerhalb der Testsituation (Kriterien) - diese müssen für den vorgesehenen Einsatzbereich des Verfahrens relevant sein
- Beispiele:
 - Depressionsfragebogen: Dauer des Aufenthaltes in einer psychiatrischen Klinik
 - Intelligenztest: Abiturnote drei Jahre nach Testdurchführung
 - Aufmerksamkeitstest (zur Messung von Fahreignung): Fehler in einer standardisierten Fahrprobe

Objektivität

Skalierung

Reliabilität

Validität

Kriteriumsvalidität

- Verschiedene Perspektiven:
 - **Vorhersagevalidität:** Korrelation mit zeitlich später erhobenen Kriterien (z. B. Intelligenzleistung vor Beginn der Lehre wird mit Abschlussnote der Ausbildung korreliert)
 - **Übereinstimmungsvalidität:** Korrelation mit zeitgleich erhobenen Kriterien (z. B. Konzentrationsleistung vor Klausur wird mit Klausurnote korreliert)
 - **Retrospektive Validität:** Korrelation mit zeitlich vorher ermittelten Kriterien (z. B. Intelligenzleistung während des Studiums und Abiturnoten)
 - **Inkrementelle Validität:** Beitrag eines diagnostischen Verfahrens zur Verbesserung der Vorhersage eines Kriteriums über andere diagnostische Verfahren hinaus

Objektivität

Skalierung

Reliabilität

Validität

Kriteriumsvalidität

- Es können auch **(Extrem-)gruppenvergleiche** herangezogen werden, die erwartungsgemäße Muster zeigen, d.h. Studien zu Mittelwertsunterschieden zwischen Gruppen, bei denen ein Unterschied erwartet wird
- Beispiel: Test, der Einstellung gegenüber der Kirche misst, sollte bei Kirchengängerinnen höher ausfallen als bei Nicht-Kirchengängerinnen

Objektivität

Skalierung

Reliabilität

Validität

Weiterer Begriff: Validierung

Validierung beschreibt den Versuch, durch verschiedene (unperfekte) Zugänge die Validität festzustellen

- Man nutzt in der Praxis Korrelationen, Work-Arounds und potentielle Zirkelschlüsse, um sich an die Frage der Validität anzunähern
- Korrelation bedeutet nicht zwingend, dass Kausalität vorliegt!
- Eine Korrelation kann indikativ für Validität sein, aber nie konstitutiv
- Oft steht „Validität“ drauf – fragen Sie sich: Ist sie auch drin?

Objektivität

Skalierung

Reliabilität

Validität

Validierungsprobleme

Zirkelschlüsse:

- Manual Test A: „Test A korreliert (erwartungsgemäß) mit Test B, also ist A valide!“
- Manual Test B: „Test B korreliert (erwartungsgemäß) mit Test A, also ist B valide!“
- Was wäre denn, wenn Test A und Test B beide etwas völlig anderes messen würden? (z.B. Intelligenz vs. Arbeitsgedächtnis)

Objektivität

Skalierung

Reliabilität

Validität

Validierungsprobleme

Post-hoc-Erklärungen:

- Konvergente Konstruktvalidität: „Eine hohe Korrelation – egal mit was – wird als Beleg für konvergente Konstruktvalidität interpretiert.“
 - Das ist Quatsch!
- Ob etwas konvergent oder divergent ist, ist eine theoretische Annahme!
- Wie hoch eine divergente Korrelation sein müsste ist eine theoretische Annahme, die a priori aufgestellt werden muss (Präregistrierung!)

Objektivität

Skalierung

Reliabilität

Validität

Validierungsprobleme

Korrelationen (und deren Signifikanz) sind von zahlreichen Einflüssen abhängig:

- Reliabilität der Tests, die korreliert werden
(→ Minderungskorrekturen möglich, siehe VL11 Testtheorie!)
- Validität des Kriteriums („Kriteriumskontamination und –defizienz“)
- Symmetrie bzw. Asymmetrie der Indikatoren (z.B. Generalitätsebene, Messmethode, z.B. Methodenvarianz)
- Stichprobengröße und andere Stichprobenmerkmale (z.B. Varianzeinschränkung, Repräsentativität)

Zusammenhang der Gütekriterien

Falls Verfahren nicht objektiv...

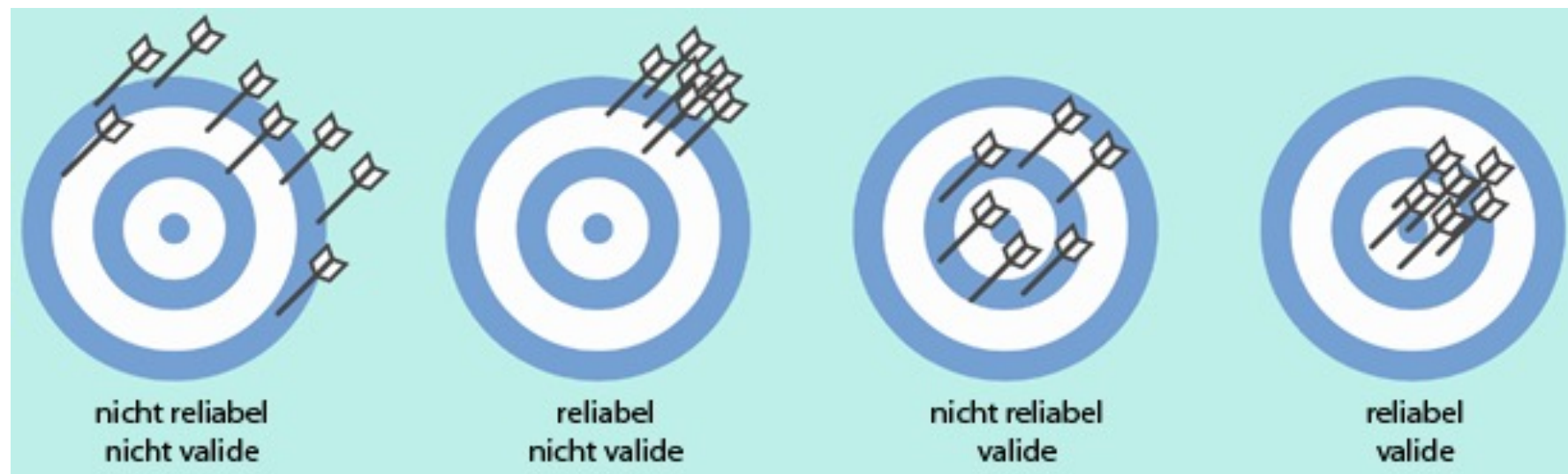
...kann keine hohe
Reliabilität erreicht werden.

Es können keine guten
Vorhersagen getroffen werden.

Validität ist nicht gegeben.

**Keine Schwarz-Weiß-
Entscheidungen**
→ daher muss jedes
Gütekriterium
eigenständig beurteilt
werden!

Zusammenhang der Gütekriterien

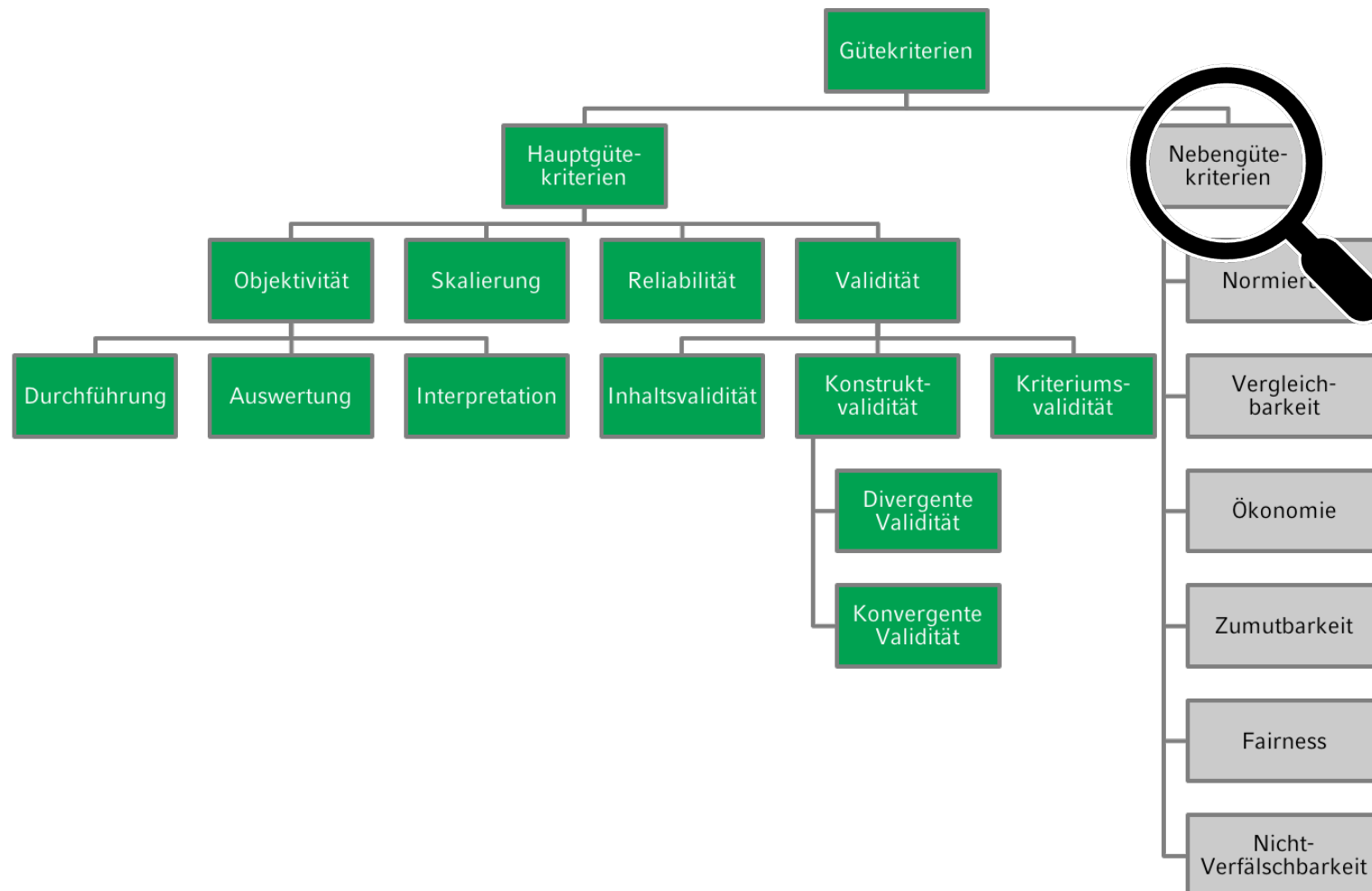


(aus Deckert, Apfelbacher & Schmitt, 2015)

Außerdem ist zu beachten bzgl. den Items eines Tests:
Bei gewisser Heterogenität sind teilweise zu hohe Reliabilitätsschätzungen
suspekt („modern talking items?“)

→ Widerspruch Inhaltsvalidität und interne Konsistenz möglich

Haupt- und Nebengütekriterien



Normierung

- Die Normierung liefert ein Bezugssystem, um die individuellen Testwerte vergleichend einordnen zu können.
- Normen müssen **aktuell, repräsentativ** und **ausreichend groß** sein.
- Normen sollten für potentiell interessante Vergleichsgruppen vorhanden sein, wenn davon auszugehen ist (z.B. aufgrund empirischer Untersuchungen), dass entsprechende Charakteristika Einfluss auf die Normwerte haben.

Vergleichbarkeit

- Ein Verfahren ist vergleichbar, wenn eine oder mehrere Parallelförmigkeiten bzw. andere Verfahren mit dem gleichen Gültigkeitsbereich verfügbar sind. Vor allem für Prozessdiagnostik relevant (um Übungseffekte zu vermeiden).

Ökonomie

- Ein Verfahren gilt als ökonomisch, wenn es kurz, einfach in der Handhabung, für eine Gruppenuntersuchung geeignet, wenig materialintensiv und schnell auswertbar ist.
- Relevant ist dabei das Verhältnis von Kosten und Nutzen im Vergleich mit anderen Verfahren.

Zumutbarkeit

- Ein Verfahren ist zumutbar, wenn die untersuchten Personen in zeitlicher, psychischer und körperlicher Hinsicht geschont werden

Fairness

- Ein Verfahren gilt als fair, wenn einzelne (für die Untersuchung relevante) Gruppen durch die erhaltenen Ergebnisse eines Verfahrens nicht diskriminiert werden, d.h. nicht systematisch benachteiligt werden aufgrund einer test-irrelevanten Eigenschaft

Nicht-Verfälschbarkeit

- Ein Verfahren ist nicht verfälschbar, wenn eine Person ihre Testleistung nicht willentlich beeinflussen kann, so dass sie daraus ungerechtfertigt einen Vorteil erlangt

Nutzen*

- *taucht in Übersicht nicht auf, wird trotzdem teilweise als Nebengütekriterium genannt
- Ein Verfahren ist nützlich, wenn es ein praktisches Bedürfnis für die Erfassung des Konstrukts gibt (und es nicht andere Verfahren schon besser machen)

Informationen zu psychologischen Tests

- Leibniz-Zentrum für Psychologische Information und Dokumentation (ZPID): <https://www.zpid.de>
- Diagnostik- und Testkuratorium (DTK) der Föderation Deutscher Psychologenvereinigungen (BDP, DGPs):
<https://www.zpid.de/index.php?wahl=Testkuratorium>
- PSYINDEX-Datenbank (mit PSYINDEX Tests):
<https://www.zpid.de/psychologie/psychologietests.php>

Leitfragen zur Nachbereitung

- Welche Kriterien diagnostischer Entscheidungen gibt es?
 - Was versteht man unter dem Begriff „Behandlungen“ und welche relevanten Unterscheidungen gibt es?
- Welche Entscheidungsmodelle gibt es?
- Gütekriterien, Gütekriterien, Gütekriterien!
 - Hauptgütekriterien (Teilaspekte der jeweiligen Gütekriterien?)
 - Zusammenhang der Gütekriterien
 - Nebengütekriterien



Quellen

- ❖ Bühner, M. (2011). *Einführung in die Test-und Fragebogenkonstruktion*. Pearson Deutschland GmbH.
- ❖ Borsboom, D., Mellenbergh, G. J., & Van Heerden, J. (2004). The concept of validity. *Psychological review*, 111(4), 1061.
- ❖ Cronbach, L. & Gleser, G. (1965). Psychological tests and personnel decisions. *Urbana: Univ. Illinois Press*.
- ❖ Deckert, S., Apfelbacher, C., & Schmitt, J. (2015). Messung der Erkrankungsschwere in der Dermatologie. *Der Hautarzt*, 66(9), 699-711.
- ❖ Schmidt-Atzert, L., & Amelang, M. (2012). *Psychologische Diagnostik*. Springer Science & Business Media.
- ❖ Sedlmeier, P., & Renkewitz, F. (2018). *Forschungsmethoden und Statistik für Psychologen und Sozialwissenschaftler*. 3., aktualisierte Auflage.
- ❖ Tack, W. H. (1976). Diagnostik als Entscheidungshilfe. *Diagnose der Diagnostik: Beiträge zur Diskussion der psychologischen Diagnostik in der Verhaltensmodifikation*, 103-130.
- ❖ Ziegler, M., & Bühner, M. (2012). *Grundlagen der psychologischen Diagnostik*. Springer-Verlag.