

## 1. Definition und Einführung

### Definition „Psychologische Diagnostik“

- Amelang, Schmidt-Atzert
  - Teildisziplin d. Psych.: bezogen auf Beschreibung, Klassifikation, Erklärung + Vorhersage menschl. Verhaltens/Erlebens
  - Gezielte Erhebung von Informationen über Verhalten/Erleben + relevante Bedingungen
  - Erhobene Infos dienen zur Beantwortung d. Fragestellung (durch Interpretation)
  - Leitung d. diagnostischen Handelns v. psych. Wissen + wissenschaftl. Standard d. Methoden
- Jäger & Petermann
  - System v. Regeln, Anleitungen + Algorithmen => Instrumente
  - Erfassung psych. Relevanter Charakteristika v. Merkmalsträgern
  - Integration d. Daten zu einem Diagnostik-Urteil
- wichtige Aspekte (gr. „diagnoskein“ = „durch und durch erkennen“)
  - breites Spektrum an Merkmalsträgern, Gruppen, Institutionen etc. (keine Einzelpersonen)
  - durch bestimmte Aufträge (!) initiiertes Handlungs- & Entscheidungsprozess (z.B. klinisch)
  - Anwendung in fast gesamter Psych., Grundlagen aus Pers., Kognition, Motivation, Sozial, Entwicklung, Pädagogik, Verkehr+Sprachpsychologie (25-44% Diagnostik i.d. Psychologie)
  - Methoden: Exploration, Interview, Befragung, Fragebogen, Testverfahren, Verhaltensbeobachtung
- kann landläufig hoch eingeschätzte Korrelationen überprüfen und evtl. entkräften (z.B. Aspirin/weniger Herzinfarkt)

### Arten v. Diagnostik

- Selektionsdiagnostik
    - Auswahl v. geeigneten Personen/Bedingungen mit dem Ziel d. Optimierung eines Kriteriums (z.B. berufl. Leistung)
    - Personenselektion (vorgegebene Bedingungen), z.B. danach auswählen, ob sie Bedingungen erfüllen (Qualifikation z.B. in Aufnahmeprüfung)
    - Bedingungsselektion, z.B. f. Absolventen besten Job suchen => geeignete Bedingungen?
  - Modifikationsdiagnostik
    - Veränderung d. Erlebens/Verhaltens oder der Bedingungen
    - Verhaltensmodifikation: Veränderung d. Person (z.B. Auswahl klinischer Interventionen)
    - Bedingungsmodifikation: Ziel = Optimierung d. Bedingungen (z.B. Arbeitsplatzgestaltung)
- ⇒ häufig auch Mischstrategien: z.B. Personen auswählen und dann individuell schulen/„modifizieren“

## 2. Entwicklungslinien d. Diagnostizierens

### Fähigkeitsdiagnostik

- Beginn d. Entwicklung: praktische Eignungsfragen, Ziel = Verhaltensvorhersage für Bewährungssit.
  - z.B. China öfftl. Amt: schriftl. + praktischer Teil, mehrstufiges Auswahlverfahren, bereits Notwendigkeit f. obj. Bewertung erkannt (mehrere Beobachter) => implizite Anforderungsanalyse, sequentielle Strategie, multimethodal + objektiv = modern!
  - z.B. Antike, Buch d. Richter: Auswahl v. 300 Soldaten aus 32000 durch „Selbstbefragung“, wer blöd ist, und Aufgabenerfüllung (oder Nachtwächter im antiken Griechenland: Mut, Selbstdisziplin, Unbestechlichkeit geprüft)
- 1. Periode: Galton
  - 1822-1911: anthropometrisches Laboratorium zur Erforschung interindividueller Unterschiede (bez. auf Fähigkeiten)
  - Starke Beeinflussung v. Evolutionstheorie: Vererbung kogn. Fähigkeiten? I = allg. kogn. Fähigkeit (n-Vert. angenommen) => Schärfe/Unterscheidungsfähigkeit d. Sinne als Index
  - Messungen: Reaktionszeiten, Pfeife zur Prüfung d. Hörschwelle

- James McKeen Cattell (Schüler v. Wilhelm Wundt)
  - Erweiterung auf das Mentale (≠ „physical test“), interind. Differenzen als systematisch zu untersuchender Gegenstand (davor: „Störfaktor“)
  - Forderung d. Ergebnisvergleichbarkeit durch Standardisierung d. Testsituation
  - Testbatterien (physical z.B. Kraft + mental, z.B. Größenbeurteilungen)
- Wissler: Korrelationsstudie (Kritik an Cattell)
  - Nur geringe Interkorrelationen und Korrel. mit ext. Kriterium f. Intelligenz wie Schulnoten d. „mental tests“, physical untereinander gut, aber nicht mit mental
  - Aber:
    - geringe Zuverlässigkeit d. mental tests
    - wenige Messdurchgänge (Aggregationsprinzip!)
    - schwache Außenkriterien
    - Varianzeinschränkung (homogene SP)
- 2. Periode: Binet
  - Kritik d. I-Messung über Sensorik => lieber komplexere geistige Prozesse (praktischer Auftrag = obj. Verfahren zur Diagnose lernschwache Schüler)
  - 1905: Binet & Simon: erster Intelligenztest, objektiveres Verfahren
    - Entwicklung altersspezifischer Aufgaben, je früher gelöst, desto klüger
    - I-Bestimmung über Vergleich IA mit LA (Problem: steigendes LA, I aber nicht)
  - Stern (1912):  $IQ = 100 * IA/LA$
  - Wechsler: Normierung d. IQ als Abweichung v. Mitte:  $IQ = 100 + 15 * z$

### **Persönlichkeitsdiagnostik**

- maximum vs. typical performance nach Cronbach
- Beginn
  - Modifikation d. Interview-Methode, da f. viele Personen unpraktikabel: schriftlich + in Gruppe
  - Woodworth (1918): „Personal Data Sheet“, 1. Persönlichkeitsfragebogen zur Truppenauswahl f. USA im 1. Weltkrieg (Selektionsaufgabe)
  - Sammlung psychiatrischer Standardfragen, Nachfolger = MMPI

### **Projektive Verfahren**

- Wortassoziationstest nach Jung (1910)
    - Standard-Wortliste vorgelesen: mit 1. Wort im Sinn antworten
    - Inhalt, Zeit, em. Reaktion soll Weg zu „Komplexen“ einer Person aufzeigen
  - Rorschach-Test (1921)
    - 10 Tintenkleecks-Tafeln (Jung-Schüler), Erweiterung v. verbal auf figural
    - Auswertung v. Gesamtdeutungszahl, Details, Form, Farbe, Bewegung, Deutungsinhalt
- ⇒ Begriff „projektiv“ erst seit 1938

### **Alltag**

- Diagnostischer Urteilsprozess: Datengewinnung -> Einzelbewertung -> Kombination/Integration -> Entscheidung
  - Wie kommt man zum Urteil? Güte? Einflussfaktoren auf Urteilsgüte?
- Kriterien zur Überprüfung v. Personenbeurteilungen
  - Konsensus: Wie sehr stimmen unabh. Beobachter im Urteil überein?
  - Mögl. Probleme: gleiche Auffassung? Fehlerhafte Urteile, gemeinsame Fehlertendenzen (z.B. Stereotype, Halo-Effekt)
- Selbst-Fremdurteil-Übereinstimmung
  - Variante d. Konsensus-Kriteriums
  - z.T. ähnliche Probleme wie beim Konsensus + unterschl. Datenbasis (Zielperson kennt sich in mehr Situationen)
- Genauigkeit
  - Übereinstimmung zw. Urteil und Kriterium (Kriteriumsvalidität)

- Probleme: shared meaning (Test und Kriterium überschneiden sich inhaltlich), Kriteriumsvalidität
- Einflussfaktoren
  - Art der Information: lebensecht/artifiziell? Menge?
  - Art d. Merkmals: global/situationsspezifisch? Beobachtbarkeit (z.B. „guter Student“ einfacher als „guter Partner“)? Bewertungsbezogenheit/Valenz (Gefahr soz. Erwünschtheit)?
  - Eigenschaften d. Zielperson
    - Extraversion (wenn hoch, pos. HE auf Beobachtbarkeit)
    - Selbstüberwachung (je nach Verzerrung pos./neg. HE)
    - Emotionsunterdrückung (neg. HE)
    - Selbstwertgefühl (hohe „Lesbarkeit“, pos. HE)
  - Art d. Beziehung zw. Beurteiler und Zielperson: Institutionell, vertraut (ggf. Verfälschung, aber bessere Kenntnisse), zero acquaintance (Nullbekanntschaft)? (z.B. Lehrer-Bewertungsstudie: bereits nach 15s valide Einschätzung mögl., CEO-Studie v. Fotos Erfolg beurteilen lassen)
  - Merkmale d. Beurteilers: Alter, Geschlecht, IQ, emotionale Stabilität? (IQ hilft, Frauen nicht besser als Männer, außer bei nonverbal Simuliertem)

### 3. Prozess d. diagnostischen Urteilsbildung

#### Diagnostischer Prozess

- Def. „Diagnostische Urteile“
    - Aussagen auf Basis vorliegender/eigens erhobener psych. Daten (Person, Gruppe, Sache)
    - z.B. „XY hat Depression“ => Urteile als Grundlage f. Empfehlungen und Entscheidungen
  - Diagnostischer Prozess (mehrstufig + rückgekoppelt)
    - Fragestellung/Anliegen/Auftrag => Diagnostisches Gespräch => Hypothesenbildung
    - Auswahl v. Tests/Herstellung v. Untersuchungssituationen
    - Datensammlung => Datenbewertung + -kombination (bez. auf Hypothesen)
    - Urteil (Entscheidung, Diagnose, Prognose...)
    - Überprüfung
  - Zentrale Fragen
    - Urteils-Entstehung (paramorphe/deskriptive Diagnostik-Modelle)
    - Validität?
    - Optimierung/Qualitätsverbesserung? (präskriptive Modelle)
- ⇒ früher nur unter Validitätsaspekt beleuchtet, heute: auch auf Zukunft/Vergangenheit bezogen + Vgl. mit Kriterium notwendig (z.B. Richtigkeit klin. Diagnosen)

#### Klinische & Statistische Methode

- Paul Meehl (1954): Scheinbarer Antagonismus in der Diagnosefindung
    - Buch als Meilenstein d. Kontroverse
    - Klinische Vorhersage (damals Standardmodell)
      - Basis = menschl. Beurteilung, Besonderheit d. konkreten Falls
      - Erfahrungs- und intuitionsgesteuert ohne konkrete Regeln (nicht nur in Klinik)
    - Statistische Vorhersage
      - Emp. gesicherte Zusammenhänge zw. Prädiktor- und Kriteriumsvariable
      - Explizite Regeln, keine Involvierung menschl. Beurteilung
  - Statistische Vorhersagemodelle
    - Einfache/multiple Regression
    - Bei dichotomem Kriterium: logistische Regression
- ⇒ Bestimmung d. Gewichte anhand vorliegender Daten (Kreuzvalidierung an anderer SP notwendig)

- ⇒ Additiv + kompensatorisch (aber nicht-kompensatorisch auch mögl.)
- Beispiel-Studie Sarbin (1942): Vorhersagegenauigkeit v. akademischem Erfolg
  - Professionelle Studienberater (mit allen mögl. Daten, Interviews etc) vs.
  - Vorhersagegleichung (Eignungstest + Schulabschlussnote, haben auch die Berater)
  - Kriterium = Noten im ersten Studienabschnitt
  - Ergebnisse: klinisch: ♂.35 bzw. ♀.69, statistisch: ♂.45 bzw. ♀.70
- ⇒ Einschätzung durch beide gut, HE: ♀ leichter vorherzusagen, Statistik bei ♀ leicht besser
  - Kritik
    - Was macht man bei „Gleichstand“?
    - Ist Statistik wirklich ökonomischer? (Modell-Erstellung etc. z.B. Anforderungsanalyse oder Regression + Kreuzvalidierung)
    - Wie valide ist das Kriterium (Noten im 1. Abschnitt?) => Aussagekraft?
    - Gab es evtl. besonders schlechte Rater/individuelle Rating-Unterschiede?

### Datenerhebung und -kombination

- Datenerhebung
  - Psychometrische Daten (aus Verfahren mit einheitl. Klassifikation + Verrechnung): frei von Ermessensentscheidungen d. beurteilenden Person
  - Nicht-psychometrische Daten (keine Standardisierung + Auswertung): keine Aussage über Zuverlässigkeit + Objektivität d. Messung aus (können es dennoch sein)
- Datenkombination
  - Formell (mechanisch, explizit): durch vorgegebene Regeln kombiniert (stat. Vorhersage)
  - Informell (beurteilend, intuitiv)
  - Kombinationsmethode kein Zusammenhang mit Vorhersage-Effektivität
- ⇒ 4 Basale Möglichkeiten d. Diagnosefindung: Kreuzung (in-)formell <-> (nicht-)psychometrisch (oder Vorliegen beider Datenarten)
- ⇒ Konfundierung (Verfahren bevorzugen jeweils die „dazugehörigen“ Datenart)
  - Validitätsunterschiede aufgrund von Datenart, -kombination oder beidem
  - Prüfung durch jeweils (in-)formelle Kombination d. gleichen Daten für 1 gl. Kriterium
- Goldberg-Studie (1965): Vergleich d. Vorhersagegenauigkeit
  - Kriterium = klin. Diagnose Psychose/Neurose, Daten = MMPI (5 von 11 Skalen)
  - Vorhersage durch 1) klin. Psychologen 2) statistische Verrechnung im Goldberg-Index
  - Ergebnis: Validität statistisch .44, klinisch von .14-.39, im Mittel aber .28
- Metaanalysen
  - Meehl (1965): 51 Studien, 33 statistische Prognose besser, 18 unentschieden
  - Grove et al. (2000): 136 Studien, 63 stat. besser, 65 gleich, 8 klinisch besser
- ⇒ formelle Datenkombination besser als informelle, zusätzliche nicht-psychometrische Daten verbessern die Urteilsqualität
- Diskussion
  - Mögl. Ursachen f. Überlegenheit d. formellen Datenkombination
    - Menschliche Verzerrung (Bias, Fehlertendenzen)
    - Inkonsistenzen bei d. Verwendung diagnostischer Entscheidungsregeln
    - Wenig Möglichkeiten, über Rückmeldungen d. Diagnoserichtigkeit zu verbessern
  - klinische Argumente: eher Betrachtung v. Einzelfällen, Berücksichtigung d. Einzigartigkeit, zusätzl. Information außerhalb d. stat. Analyse
  - statistische Argumente: seltene Einzelfälle kein stat. Problem, Einbettung v. Zusatzinformation in das Modell

### Paramorphe Modelle d. Diagnostizierens

- Erforschung d. Urteilsgenese („im Kopf“ des Diagnostikers, Richtigkeit/Validität egal)
  - Paramorphe Modelle = formelle Repräsentationen d. diagn. Entscheidungsverhaltens
  - Zusammenhang zw. zur Verfügung stehenden Infos + Entscheidung immer ähnlich
  - Ziel: mögl. gute Darstellung d. inneren Modells im außen

- Erforschung
  - Diagnostiker = Proband
  - Verfahren = laut denken lassen, direkt fragen, Zusammenhang Urteil <-> Info untersuchen
  - Zunächst untersucht: einfache lineare Regressionen, später auch kurvilineare Zusammenhänge und/oder interaktive Komponenten (z.B. wenn A, dann nicht B) oder Gewichtungen
- Empirische Befunde
  - Urteilsbildung oft durch Haupteffektmodelle gut repräsentiert
  - Vgl. Selbstbericht <-> Repräsentation
    - Diagnose für komplexer gehalten als tatsächlich
    - Andere Einschätzung d. Gewichte
    - Entscheidungsregeln nicht vollständig zu explizieren
    - Bericht v. Lehrbuchwissen, wenn nach Diagnosefindung berichtet

### Konfigurationsmodelle (Muster v. Werten)

- Selbstverständnis v. Diagnostikern
  - Ausprägungsmuster wichtig (z.B. mehrere glz. auftretende Variablen, holistische Sicht)
  - Nicht immer Kompensationsmöglichkeiten (Unterschied zur math. Regression)
  - Sequentieller Entscheidungsprozess (nur „weiter“, wenn best. Bedingung erfüllt)
- Konfigurationsmodelle
  - Entscheidungen als Sequenz v. Wenn-Dann-Regeln
  - Person muss bestimmtes Wert-Muster („Konfiguration“) aufweisen (Kompensation mögl., aber nicht zwingend da)
  - z.B. i. Auswahlverfahren Konzentrations-Score >0,5, Extraversion >-1, G >1, wenn 1 nicht erfüllt, Ablehnung und Annahme nur, wenn alles erfüllt
    - Oder mit Kompensation z.B. Ausgleich eines zu niedrigen Konzentrationscores durch hohe Vigilanz-Werte
- Hypothesenagglutinerung (Wottawa et al.)
  - Elaborierte Form d. Rekonstruktion d. diagn. Urteilsbildung (schrittweise Rekonstruktion)
  - Logisch verknüpfte „agglutinierte“ Hypothesen (Wenn-Dann) führen zur Diagnose
  - Ziel: Auffinden d. impliziten Entscheidungsregeln durch Befragung -> Begründung -> vorläufiger Regelformulierung und deren anschließender Überprüfung + Modifikation
  - Sobald Modell formuliert ist: Anwendung auf neuen Fall -> Entscheidung d. Diagnostikers und d. Modells vergleichen -> bei Treffer wieder auf neuen Fall anwenden, bei Fehler wieder Begründung erfragen und ggf. Modell modifizieren
  - Bsp.: Metallfaharbeiter, Kombination versch. Leistungstests, teilw. Kompensation mögl.
- ⇒ Goldberg-Paradoxon
  - Berechnung d. „idealen“ inneren Modells aus empirisch erfragten/beobachteten Diagnosen auf die Daten angewandt (z.B. bei Goldberg MMPI-Scores) ist besser als der Diagnostiker
  - Begründung: Mensch wendet innere Modelle nicht konsequent an

### Optimierung d. diagnostischen Urteilsbildung

- Aufstellen expliziter Modelle
  - Transparenz => eher reliabel
  - Optimierbarkeit durch explizite Regeln (wo läuft was schief?)
  - Lernen und Erfahrungsaustausch (bessere Verbalisierbarkeit und Kommunikation)
  - Entlastung (auch finanziell durch einfachere Modelle, nicht alle Variablen)
  - Konsistenz (2 kommen zum gl. Urteil => Fairness, aber nicht unbedingt ↑ Validität)
  - Validitätssteigerung (u.a. auch durch Optimierbarkeit)
- Nutzung d. klinischen Inferenz (Schlussfolgern)
  - Wert informeller Daten (z.B. Interviews)

- Möglicherweise aber suboptimale Aufbereitung/Kombination
- ⇒ „Königsweg“ d. Diagnosefindung
  - Klinische Tätigkeit: Datensammlung => klinische Schlussfolgerung und Kombination => Q-Sort oder andere Form systematischer Beschreibung
  - Formelle Prozeduren: Klassifikation + Vorhersage
  - Übergang v. klinischem zu formellem Bereich: z.B. durch systematische Verhaltensbeobachtung (einigermaßen gute Systematisierung d. Datenwusts)

#### 4. Entscheidungstheoretische Modelle

##### Prädiktive Validität

- Grundlage = präskriptive Entscheidungstheorie (Ziel = mögl. optimale rationale Entscheidungen) ≠ deskriptive Beschreibung d. faktischen Entscheidungsverhaltens von Personen/Institutionen
- Diagnostische Relevanz d. Entscheidungstheorie v.a. in Personalauswahl
  - Güte von Selektionsentscheidungen?
  - Finanzieller Nutzen + Vergleich v. diagnostischen Verfahrenen?
- Definition: Korrelation zw. zukünftigem Kriterium (z.B. Berufserfolg) und Prädiktor (z.B. Eignungstest)
  - Traditionelle Sicht: Güte d. Diagnostik hauptsächlich. abh. v. prädiktiver Validität
  - Aber: Kosten-Nutzen-Abwägung nötig!

##### Entscheidungsstrategien

- Information über eine Person -> Entscheidung -> ja/nein -> jeweils geeignet oder nicht (4 Mögl.)
- Variablenkombinationen
  - Single cutoff: 1 Variable muss best. Wert haben, sonst Ablehnung
  - Multiple Regression cutoff: 1 Wert x, der durch Regression entsteht, braucht best. Wert
  - Multiple cutoff: mehrere Variablen relevant
- Sequenzen
  - Nicht-sequentiell: terminale Entscheidung basierend auf 1 einzigen Testung
  - Sequentiell bzw. multiple-stage: mehrere Tests, ein Teil d. (mittelguten) Probanden „weiter“, erst dann Endentscheidung (nicht alle Probanden machen alle Tests)
- Bausteine d. vollständigen Strategie
  - Pre-reject-Strategie: Zurückweisung aller Probanden mit zu ↓ Score (Rest weitere Tests)
  - Pre-accept: sofortige terminale Annahme sehr guter VPn nach 1. Test
  - Mitte wird vorläufig angenommen + nach einem weiteren Test B angenommen/abgelehnt

##### Güte diagnostischer Entscheidungen

- Berechnung
  - 4-Feldertafel: Ablehnung/Annahme + Geeignet/Ungeeignet => falsche neg./pos. Entscheidung oder valide neg./pos. Entscheidung
  - Rel. Häufigkeit korrekter Entscheidungen =  $(VP + VN)/\text{Gesamt}$  (aber: nicht berechenbar, da Negativentscheidungen nicht verfolgbar)
- ⇒ deshalb: Erfolgsquote:  $VP/(VP+FP)$  Anteil d. Geeigneten an den Akzeptierten abh. von
  - prädiktiver Validität d. Auswahlverfahrens
  - Basisquote: Anteil d. tatsächlich Geeigneten unter den Bewerbern (nicht tatsächlich bekannt, eher geschätzt oder durch Annahme von ALLEN Bewerbern oder Ziehen einer Zufallsstichprobe beobachtbar)
  - Selektionsquote: Anteil d. Akzeptierten unter den Bewerbern
- Auswirkungen auf die Erfolgsrate
  - ↑ prädiktive Validität: ↑ Erfolgsrate (Ellipse quasi „dünner“, mehr VP + VN)
  - ↑ Basisquote (weniger Eignung erfordert -> mehr Geeignete vorhanden): ↑ Erfolgsrate, da gleich viele N, aber prozentual mehr FN + mehr VP

- ↓ Selektionsquote (wenig Leute genommen, hohe Werte im Test gebraucht): ↑ Erfolgsrate, insg. mehr Ablehnungen, aber besseres Verhältnis von VP + FP
- ⇒ Erfolgsquote als Gütemaß diskutabel, da FN nicht abgebildet
- Taylor-Russel-Modell: Erfolgsquote als Funktion von Validität, Basis- und Selektionsquote
  - Je valider das Verfahren, desto höhere EQ (wenn Validität = 0 (Zufallsauswahl): EQ = BQ)
  - Je ↓ die SQ, desto bedeutender der Validitätszuwachs (mögl. viele Bewerber testen!)
  - Bei sehr ↑ SQ und BQ bringt höhere Validität eher wenig (eh nicht genug bzw. mehr als genug Geeignete vorhanden) => Interaktion zw. d. Faktoren
  - Höchster Zuwachs bei BQ 50 & SQ 10
- ⇒ Aber: keine Berücksichtigung versch. Facetten
  - „Binnendifferenzen“ bei den (Geeigneten), nur dichotomisiert
  - Ökonomische Folgen von Personalentscheidungen (finanzieller Aspekt)
  - Finanzieller Aspekt d. Durchführung und Konzeption d. diagn. Verfahren

### Entscheidungstheoretische Nutzenerwägungen

- Ziele
  - Profit der diagnostischen Prozedur prüfen
  - Vergleich mehrerer Verfahren hinsichtlich ihres relativen Nutzens
- ⇒ zentrale Variable: ökonomischer Nutzen (Zuwachs ggüber Zufall/bereits eingesetztem Verfahren)
  - bestimmt durch prädiktive Validität d. Verfahrens, Selektionsrate (wie gut die Ausgewählten abschneiden) + zusätzlich erwirtschafteter Gewinn (i. Vgl. zu Durchschnitt)
- Formeln
  - $\Delta U$  (durchschnittl. Zuwachs an Nutzen) je Eingestellter =  $r_{xy} * Z_x * SD_y$
  - $r$  = prädiktive Validität,  $Z$  = durchschnittl. standardisierter Wert im Eignungstest ( $\approx$  SQ),  $SD$  = Standardabweichung d. Kriteriums y in Geldwert \$\$\$
  - z.B.: Validität = .50, Bademeister werden eingestellt, wenn sie 1 SD besser sind als das Mittel ( $Z = 1$ ) und erwirtschaften 2000\$ mehr =  $SD_y \Rightarrow$  Nutzen = 1000\$
- Berechnung d. Formelteile
  - $r_{xy}$ : Korrektur für Varianzeinschränkung (da nur Gute untersucht, Testwerte ↓ Range) + Korrektur für Unreliabilität d. Messung d. Kriteriums (einfache Minderungskorrektur)
  - $Z_x$ : abhängig von Selektionsquote (weniger eingestellt => Z höher)
    - z.B. SQ = 0,8:  $Z_x = 0,35$  und bei SQ = 0,2  $Z_x = 1,40$  (ablesbar an n-Vert)
  - $SD_y$  (wie sehr unterscheiden sich Mitarbeiter in ihrer berufl. Leistung in \$)
    - Schwierig zu berechnen („Achillesferse“)
    - z.B. durch obj. Daten (Verkaufszahlen)
    - subj. Daten: Experteneinschätzung d. Geldäquivalente am 15., 50. + 85. Prozentrang => Differenten v. 50-15 und 85-50 Schätzung für  $SD_y$
    - proportionale Regeln: 40% d. mittleren Gehalts (konservative Schätzung)
- ⇒ bis hierhin Bruttonutzen, noch zusätzl. Kosten d. Tests erforderlich!
  - $\Delta U$  je Eingestellter =  $(r_{xy} * Z_x * SD_y) * N_s * T - C * N$ 
    - $T$ : durchschnittl. Verweildauer im Betrieb,  $N_s$ : Anzahl akzeptierte Bewerber
    - $C$ : Kosten d. Testung,  $N$ : Anzahl d. durchgeführten Testungen
- ⇒ wenn SD groß (wichtige Position), lohnt sich valider Test
- Sensitivität + Spezifität
  - Sensitivität: Anteil pos. Diagnostizierter an faktisch Positiven (wie leicht die tatsächlich Geeigneten/Erkrankten usw. gefunden?)  $VP/(VP+FN)$
  - Spezifität: Anteil neg. Diagnostizierter an faktisch Negativen (wie gut werden die Ungeeigneten identifiziert?)  $VN/(VN+FP)$
  - Bei klin. Fragestellung kann nur eins maximiert werden
- Studie Hunter et al. (1979) zum Nutzen 1 I-Tests (PAT) für Programmierer f. öfftl. Dienst (Wirtschaft oder Regierung)
  - Messfehler- und Varianzeinschränkungs- korrigierte Validität = .76 (sehr hoch?!)

- SDy nach PR variiert („bessere“ bei 50-85 wertvoller als zw. 15-50)
- Ergebnisse in Abh. d. SQ und der Validität d. bisherigen Auswahlverfahrens und in Mio. \$ angegeben
  - Am höchsten bei SQ = 5% und vorheriger Validität .00 97,2Mio. \$ Gewinn
  - bei gleicher SQ und vorher Zufallsauswahl (Val = .50) nur 33,3 Mio.
- Kritik
  - Zu hohe Validitätsschätzung, Probleme bei SD-Schätzung
  - Berücksichtigung ökonomischer Variablen (Steuern, Zinsen, Inflation)?
  - F. Gesamtwirtschaft BQ nicht vorhanden!
- ⇒ Aber: Aufzeigen d. großen potentiellen Nutzens guter Diagnostik bei Personalauswahl
  
- Berücksichtigung d. zentralen ökonom. Variablen (Boudreau, 1983)
  - Steuern, Zinsen + Inflation (zukünftige Gewinne weniger wert als gegenwärtige)
  - Erweiterung d. Brodgen-Formel um variable = regelmäßige/feste = Erstellung Kosten d. Testung, TAX = Steuern und Zinsrate/Jahr
  - Anwendung durch Holling, 1998 für AC in einer Versicherungsgesellschaft
    - SQ = .50 => Z = .80, Einstellung über 3 Jahre => Berechnung d. Nutzens getrennt für jedes der 3 Jahre und anschließende Summierung
    - Inkrementelle (zusätzl.) Validität d. AC = .11, 40% Steuern
- ⇒ Erst Gewinn berechnet, Kosten abgezogen => Nutzen d. AC immer noch knapp 400.000DM
- ⇒ Vorteile: zusätzliche Rentabilitätsanalysen (bei z.B. höheren Kosten), Berechnung für einzelne Bausteine mögl., aber: Ergebnisse nicht auf andere Arbeitsfelder übertragbar (weil z.B. SD schwieriger zu berechnen)
  
- Bewertung
  - Ökonomische Argumentation in Geldeinheiten mögl. (Nutzen d. Diagnostik verdeutlicht, Kommunikation mit BWL-ern mögl.)
  - Vgl. unterschdl. Auswahlprozeduren mögl. (teurer+guter vs. billiger+weniger guter Test, Nutzen eines zusätzlichen teuren Tests, z.B. AC)
  - Ungenauigkeit d. Nutzenberechnungen (SD schwierig, aber auch bei anderen Entscheidungen unter Unsicherheit=)
  - Lineares Modell? (auch nichtlineare Zusammenhänge zw. Mitarbeiterleistung und Gewinn/Verlust für das Unternehmen mögl.)
- ⇒ Aber: Evidenz für relative Robustheit d. Modells in vielen Kontexten
- ⇒ Zusammenfassung insg.: Validität allein -> Taylor-Russel-Tafeln -> BCG-Modell -> Boudreau (zusätzl. Ökonom. Variablen)

### **Antwortabhängige Tests**

- Beschreibung
  - Verfahren, bei denen frühere Antworten d. VPn über d. weiteren Verlauf entscheidet
  - Sinn: Ökonomie (nicht alle VPn mit allen Tests) => Prereject + Preaccept günstiger
  - Vor-Strategien nicht von Testatterie abh., sondern von einzelnen Items d. einen Tests
  - ≈ Bedingungsselektion (für VPn passende Bedingungen/Items suchen), auch i. Explorationsgespräch (aber eher subjektiv)
- 2 allgemeine Aspekte
  - Quantitativer Aspekt: sequentielle Variation d. Itemanzahl bis zur Entscheidung („Kann ich abrechnen oder muss ich noch ein Item geben?“)
  - Qualitativer Aspekt: adaptive Anpassung d. Itemschwierigkeit an Fähigkeiten d. VPn („Welches Item gebe ich als nächstes?“)

- Sequentielles Testen (quantitativ)
  - Frage: Ist Entscheidung mit vorliegenden Antworten mögl. oder noch mehr Items? (z.B. Diagnose Depression, Prüfungsanforderungen erfüllt, Hochbegabung)
  - Auswahl nach wenigen Items möglich (i. Vgl. zu mittelmäßig Geeigneten mit gl. Itemzahl)
  - Variable = Zahl d. Items (so viele, bis Entscheidung bei allen VPn gleich sicher)
    - Vorherige Festlegung d. Sicherheitsniveaus/Risikos d. Entscheidung => Höhe d. mögl. Risikos kontextabhängig
    - Statistische Prüfung d. Entscheidungssicherheit (Binomialverteilung f. Gruppenzugehörigkeit)
    - 2 Fehler: Institutionenrisiko (falsch positiv/α-Fehler), Personenrisiko (falsch negativ, beta-Fehler)
  - Gleichwertige Items, Anzahl d. beantworteten Items relevant (Ø Binnendifferenzierung)
- ⇒ Ziel = Verkürzung d. Testdauer (Abbruch bei ausreichender Information)
  - + Klassifikations-/Selektionsentscheidung (keine Messung individueller Unterschiede)
- Adaptives Testen
  - Ziel: präzise Messung von Eigenschaftsausprägungen durch variable Anpassung d. Schwierigkeit d. Items
  - Frage nach dem bestgeeignetsten Item für die Fähigkeiten d. VPn (nach bisherigen Infos)
  - Kriterium: Informationszuwachs die zu messende Eigenschaft betreffend
  - Statistisch z.B. pyramidal angeordnete Items (wenn 1 Antwort falsch, in „darunter liegender“ Ebene eher zum leichteren Item, sonst zum schwierigeren) mit Ziel, d. perfekten Schwierigkeit zu finden
- ⇒ Vorteile
  - Ersparnis von Kosten für sonst erhobene irrelevante Information  
=> Ökonomie (Tempo, Genauigkeit)
  - Verhinderung v. Motivationsschwächung d. VPn (weniger Raten, mehr Anstrengung) durch gleichmäßig positive und negative Rückmeldung
  - Hoher Grad d. Individualisierung
    - Branched testing: nach festgelegten Verzweigungsregeln VPn leichtere oder schwierigere Aufgabe stellen)
    - Tailored testing: Errechnung des theoretischen Fähigkeitsniveaus d. VPn nach jedem Item und darauf basierende Item-Auswahl
- ⇒ Nachteile
  - Problematischer VPn-Vergleich (da unterschdl. viele und schwere Items)
  - Problematische Reliabilitätsmessung: kein fester Item-Satz, keine Referenzpopulation (weil unterschdl. Bearbeitungsreihenfolge)
  - Aber: Lösbarkeit d. Probleme außerhalb d. klass. Testtheorie (probabilistisch, Item Response Theory)
- Beispiele: FAKT-II (Frankfurter Adaptiver Konzentrationstest), AID (IQ-Test), AMT (Adaptiver Matrizentest)

## 5 – Handlungstheoretische Modelle (Kaminski)

### Allgemein

- Formuliert für Arbeitsfluss in d. modifikatorischen Praxis (≠ Selektion wie Entscheidungstheorie) und individuelle Entscheidungen (≠ institutionelle in der ET)
- Bessere Durchschaubarkeit v. Arbeitsprozessen für den diagnostischen Praktiker
- Diagnostisches Handeln als Kreislaufprozess
  - Steuerung durch kognitive Mikroprozesse

- Sequentielle Rückmeldeschleifen

## Modell

- Diagnostische Schleife: Eingangsdatenbeschaffung -> Hypothesen -> Prüfphase (Hypothesen ausreichend?) -> wenn nein, WH d. diagnostischen Schleife
- Praktische Schleife: Planung -> praktische Phase -> Prüfung: mit Hypothesen konkordant? -> wenn nein, WH d. diagnostischen Schleife
- Wenn ja: Prüfung, ob Zielkriterium erreicht -> wenn nein, WH d. praktischen Schleife, wenn ja Ende <3

## Diagnostischer Prozess

1. Fragestellung/Anliegen/Auftrag (z.B. Kind kommt in Schule nicht mit)
2. Diagnostisches Gespräch
3. Hypothesenbildung (z.B. emotionale Ursachen, Intellekt, organische Krankheit als Grund)
4. Auswahl v. Tests und Untersuchungssituationen (z.B. Familienbeobachtung, IQ-Test, medizinische Untersuchung für die jeweilige Hypothese)
5. Datensammlung, -erhebung (z.B. Änderungsidee Schulwechsel + Beratungsgespräche)
6. Datenbewertung und -kombination im Hinblick auf Hypothesen
7. Diagnostisches Urteil (Entscheidung, Diagnose, Modifikation...)
8. Überprüfung (z.B. d. emotionalen Zustands des Kindes nach Schulwechsel)

## Implikationen

- Grundsätzliche Ansicht von Diagnostik als hypothetische Interpretation (provisorisch erstellt und dann beibehalten/verworfen)
- Datenbeschaffung immer geplant und auf spezifische Hypothesen bezogen
- Genaue Explikation v. Prüf- und Entscheidungsprozessen
- Keine Einleitung d. praktischen Phasen ohne Formulierung konkreter Hypothesen
- Steuerung d. Diagnostik von erreichbaren Zielkriterien (Rückkopplung bis Ziel erreicht)
- Fehler durch ungenügende Hypothesenbildung: überflüssige Datenanhäufung, Kapazitätsprobleme, Unangemessene Informationsselektion und -akzentuierung, Überbetonung d. Diagnosestimmigkeit auf Kosten ihrer praktischen Relevanz

## Wissensspeicher

- Def.: „Gesamtheit d. Wissens des Psychologen zur Bearbeitung eines Falles, ggf. aktualisierungsbefürtigt und durch Speicher mit spezifischen abrufbaren Inhalten repräsentiert“
- Speicher 1: Änderungswissen
  - bezogen auf Veränderung d. unerwünschten Ausgangszustandes Z1 zu Zielzustand Z2
  - versch. Hochwertiges Wissen: wissenschaftlich begründetes Änderungswissen > berufspraktische Erfahrung > Alltagserfahrung
  - relevant für Hypothesenbildung und -bearbeitung
- Speicher 2: Kompetenzwissen
  - Kenntnis von Fachmenschen für bestimmte Daten und Änderungen (auch vor Diagnostik, wenn Hypothesen außerhalb psych. Zuständigkeitsbereich)
  - Ersatz für Speicher 1, wenn dieser inkompetent zur Planung d. Praktischen Phase
- Speicher 3: Bedingungswissen
  - Wissen über mögl. Bedingungshintergründe und deren zukünftige Auswirkungen
  - Ebenfalls zur Hypothesenbildung nötig
- Speicher 4: Gewissen
  - Individuelle Minimalziele d. Psychologen hinsichtlich von unbedingt zu erreichenden Zielzuständen und dringend zu vermeidenden Ausgangszuständen
  - Bei Bestimmung von Zielzustand in der Hypothesenbildung

## Bewertung

- Prozessmodell, Hinweis auf Notwendigkeit von grundwissenschaftlichen Wissens bei bestimmten Diagnosevorgängen
- Steuerung d. Arbeitsphasen durch präzise Hypothesen, Evaluation v. Zwischenzuständen
- Keine Theorie, sondern Rekonstruktion d. diagnostischen Prozesses
- ⇒ Eher beschreibend als die Entscheidungstheorie

## 6 – Methoden d. Datenerhebung

### Datenarten

- L-Daten (life-record)
    - Def.: ohne Beeinflussung, d. Betroffenen registrierbare Daten aus d. alltägl. Leben
    - z.B.: Verhaltensbeobachtung, Dokumentenanalyse, Lebenslauf, Fremdbeurteilungen
  - Q-Daten (questionnaire)
    - Def.: Selbstauskünfte in Befragungssituationen
    - z.B.: Interview, Persönlichkeits-, Interessen- oder Einstellungsfragebogen
  - T-Daten (test)
    - Def.: Ergebnisse vermeintlich objektiver Tests + Experimente (keine Augenscheinvalidität in Bezug auf Messintention) => keine Selbstauskunft
    - z.B.: physiologische Messungen, Experimentaldaten (z.B. Reaktionszeiten), Fähigkeitstests
- ⇒ Auch gleichzeitige Erhebung mögl. (z.B. am PC: Q durch Itemantworten, T mittels Reaktionszeiten und L durch Verhaltensbeobachtung)

### L-Daten

- Fremdbeurteilungen
  - Verhalten in Testsituationen, eher anforderungsfreie Verfahren (z.B. Rollenspiel)
  - Systematische Beobachtung wünschenswert, Bsp.: Peers, Q-Sort (über andere)
- Q-Sort
  - Testmaterial = 50-100 das Zielobjekt beschreibende Items auf Karten, Zielobjekt = man selbst (Q-Daten in diesem Fall) oder andere Person => Ipsativer Vergleich
  - Regeln d. Sortiervorgangs: Einordnung einer vorgegebenen Anzahl v. Karten in Kategorien (ca. 7-9 von am wenigsten – am meisten zutreffend) => Ziel = Quasinormalverteilung
- Anwendungsmöglichkeiten
  - Interpretation v. Ähnlichkeitswerten (z.B. Vgl. Selbstbild/Fremdbild/Idealbild, evtl. prä/post-Vergleiche, z.B. bei Therapie) => Korrelationskoeffizient aus Rangdifferenz
  - Vorhersage externer Kriterien durch Q-Set mit nicht/mittel/sehr erwünschten Items und Auswahl/Einschätzung v. Personen mittels dieser Sortierung

### T-Daten

- Probleme bei Verwendung v. Selbstauskünften
  - Durchschaubare Messintention => Manipulation d. Antwortverhaltens/Antworttendenzen
  - Selbstauskunft nur über innerlich Zugängliches (Introspektion notwendige Voraussetzung)
  - Verbaler Ausdruck nötig: evtl. Konfundierung Iteminhalt <-> persönl. Interpretation
- ⇒ Undurchschaubarkeit im objektiven Persönlichkeitstest (contra soz. Erwünschtheit)
- Def. nach Schmidt (1975)
  - Unmittelbare Verhaltens erfassung ohne Augenscheinvalidität und Eigenbeurteilung (in Aufgabenwahl und Auswertungsmethode)
  - Dennoch Anforderung an Standard-Gütekriterien
- Verfahrensüberblick
  - Fragebogen: z.B. Antwortzeiten oder Tendenz zu Extrema (spez. andere Auswertung)

- Leistungstests zur Erfassung v. Persönlichkeit, z.B. Rod-and-frame-Test zur Feldabhängigkeit (Linie vor Hintergrund vertikal ausrichten) oder Emotional-Stroop-Test
- Physiologie, Projektive Tests, Experimentelle Verfahren

## Projektive Verfahren

- Begriffsklärung
  - Art d. Beziehung zw. Testmerkmal und Persönlichkeit ist projektiv (Testverhalten = Projektion + diagnostisch interpretierbar) => offenes Antwortformat
  - Mehrdeutige Vorlage disambiguieren <3, z.B. Bildinterpretation od. Geschichte erzählen
- Projektionshypothese
  - Hervorrufen d. eig. Innenwelt durch Stimulus
  - Nach Freud: Projektion n. außen als Abwehrmechanismus für Ich-bedrohliche Reize
- Einteilung d. Projektiven Verfahren: Assoziationen, Formdeutungen (Rorschach), Thematische Apperzeptionen, Spielerische Verfahren, Zeichnerische Verfahren, Farbwahl, Bilderwahl (7)
- Rorschach-Test
  - Auswertung in 4 Hauptkategorien mit vielen Unterkategorien: Lokation (Ganzheit oder nur Ausschnitt interpretiert?), Determinanten (Form/Farbe/Schatten/Bewegung ausschlaggebend?), Inhalt (formale Auswertung) + Originalität
  - Auswertung nach Exner (1974): Comprehensive System CS, detaillierte Auswertungshinweise, Normen + Kategoriencluster mit klin. Entsprechungen
  - Problematisch: alle Gütekriterien
- TAT
  - Aufgabe: über mehrdeutiges Bild Geschichte erzählen (Hintergrund, Prognose, Emotionen d. Figuren d. Bildes)
  - 20 Tafeln, wobei eine ganz leer ist (am mehrdeutigsten)
  - Eher impressionistische Auswertung (interessierende Themen?), aber auch Quantifizierungsversuche (z.B. Aggression, Leistungsmotiv)
  - Bemänglung in Übersichtsarbeiten: nur vereinzelte Kriteriums- und wenig inkrementelle Validität, fehlende Replikationen, Objektivität/Reliabilität (nicht f. IQ kontrolliert)
  - Interpretationsschwierigkeit: Testverhalten = Verdrängtes, Bewusste Traits, States oder Rationalisierung/Rechtfertigung eigener Verhaltensweisen?
- ⇒ Gründe
  - Fehlen eines verbindlichen Auswertungssystems
  - Variation in Anzahl und Inhalt d. Vorlagen zwischen d. Untersuchungen
  - Mangelnde theoretische Fundierung, v.a. Unschärfe d. Projektionsbegriffs
- ⇒ Fazit: einsetzbar bei Teilbereichen d. Persönlichkeit, bei Kindern, Schwierigkeiten in Auswertung und Psychometrie + bei theoretischer Basis d. Projektion

## Interview

- Def.: Methoden zur Erhebung diagnostisch relevanter Daten mittels eines Gesprächs
  - z.B.: Anamnese (Vorgeschichte), Exploration, Einstellungs- und Auswahlgespräche
  - unterschl. Standardisierungsgrade in Durchführung + Auswertung (ganz, halb, nicht?)
- Phasen
  - Eröffnung (Vorstellung, Rollenklärung, Ablaufschilderung, Rechte, Einverständnis zu Aufnahmen etc., Überleitung zur nächsten Phase)
  - Informationserhebung (Einleitung, Thema nennen + erfragen, abschl. Zusammenfassung)
  - Abschlussphase (Zusammenfassung, Vergessenes + weiteres Vorgehen klären, Erkundigung nach Fragen d. Interviewten, Beruhigung durch em. Neutrales Thema, Verabschiedung)
- ⇒ Sicherheit f. alle Beteiligten + höhere Transparenz (in Ablauf + Umständen)
- Techniken d. Gesprächsführung
  - Aktives Zuhören (Aufmerksamkeit vermitteln, Paraphrasierungen, Zusammenfassungen)
  - Erkennen v. Widerstand (Verspätung, Gähnen, verbales Vermeidungsverhalten)
- SKID (Strukturierte klin. Interviews für DSM-IV)

- Übertragung d. DSM-Kriterien in standardisierte Fragen
- Eingangsfragen + Verzweigungsregeln, Verringerung d. Übersehens v. Komorbiditäten
- Strukturierte Interviews zur Eignungsdiagnostik, z.B. MMI (Schuler)
  - Leitlinie zum Aufbau d. Erstgesprächs, keine inhaltliche Ausgestaltung
  - Multimodal durch versch. Interview-Methoden
  - Aufbau: Gesprächsbeginn (unbewertet) -> Selbstvorstellung -> Fragen zur Berufswahl + fachl. Prüfung -> freier Gesprächsteil -> Biografie -> realistische Tätigkeitsinformation -> situative Fragen-> Abschluss (zuvor detaillierte Anforderungsanalyse)
  - Verhaltensverankerte Skalen
  - ↑ Interrater-Reliabilität als  $\emptyset$  standardisiert (.68 vs. .84) + inkrementelle Val. zu IQ-Tests (z.B. .51 wird zu .63)
  - Metaanalytische Validitäten (i. Bezug auf Berufserfolg): strukturiert .44, unstrukturiert .33 situativ .43, verhaltensbeschreibend .51

### Verhaltensbeobachtung

- Unterscheidung systematisch/unsystematisch bzw. (un-)kontrolliert etc.
- Klassifikation systematische Verhaltensbeobachtung
  - Naturalistisch „im Feld“ vs. nichtnaturalistisch „im Labor“
  - Aktiv/passiv oder nicht teilnehmend (Labor immer nicht-teilnehmend)
  - (un-) vermittelt (Feld immer unvermittelt = direkt zugeschaut)
- Arten d. Datenregistrierung
  - Isomorphe (= ganz genaue, 1:1) oder reduktiv
  - Reduktiv: Klassifikation in Zeichensysteme, Kategorien oder durch Ratings/Einschätzung
- Verhaltensklassifikation
  - Kategorien: Erfassung d. ganzen Verhaltensstroms, Unterteilung in unabhängige Klassen
    - Bsp.: Bales-Interaktionsanalyse: z.B. „zeigt Solidarität, Spannungslösung, Zustimmung“ für positives soziales Verhalten
    - Erfassung allen Verhaltens, MUSS durch 1 d. Kategorien abgedeckt werden
  - Zeichen: Liste interessierender Verhaltensweisen, z.B. nonverbal „unruhige Sitzhaltung“
    - $\neq$  Kategorien, da versch./gar keine Kodierung gleichzeitig möglich (aber 1 reicht)
    - z.B. Aggressions-Beobachtung: Verhaltensanker über Zeit betrachtet: X machen
  - Segmentierung d. Verhaltensstroms (f. Zeichensysteme)
    - Time-Sampling (Zeit-Teil-Methode): an Verhalten angepasstes Zeitraster über Verhaltensstrom legen, Kodierung pro Zeitintervall (Verhalten ja/nein pro Zeit)
    - Event-Sampling (Ereignis-Teil-Methode): Ereignisse rastern Verhaltensstrom, z.B. Anfang/Ende von „Weinen“ notieren => genauere Zeitschätzung (Video-Analyse)
  - Rating- und Einschätzverfahren: Beurteilung! (nicht Beobachtung)
    - Meist 5-7-stufige Ratingskalen
    - Größere Datenreduktion, weniger verhaltensnah/abstrahierter, stärkerer Einfluss v. Beurteilungsprozessen + Beobachtungsfehlern
- Berechnung d. Interrater-Reliabilität
  - Kategorien + Zeichensystemen: Berechnung v. Cohens Kappa (mhmm)
  - Ratingverfahren: Intraklassenkorrelationen (Rechnung m. abs. Werten, nicht Rangfolge)
  - Steigerung d. Objektivität durch Beobachtertrainings!

### Leistungsmotivation

- Objektiver Leistungsmotivationstest OLMT
  - Aufgabe: unter Zeitdruck 100 Bildschirmfelder durch Tastendrucke abfahren
  - 10s/Durchgang (n=30) => AVs: Schnelligkeit + Anzahl d. zurückgelegten Felder
  - Auswertung
    - Aufgabenbezogene Anstrengung: Anz. d. zurückgelegten Felder
    - Motivation durch Ziele/eig. Anspruchsniveau: Anzahl erste 10 Felder/zweite 10 Felder bzw. im Vergleich mit eigener Zielsetzung
    - Motivation durch Konkurrenz: bei fiktivem Gegner Leistungen vgl.

- Validität
  - .30 (mit Leistungstests, Abinoten + Raven-Matrizen)
  - -.24 (mit Vordiploms-Note Psychologie, inkrementelle Val. zur Abinote)
- Verfälschbarkeit? Normen? Evtl. Konfundierung Leistung <-> Leistungsmotivation  
Bewertung: entwicklungsfähiger Versuch!

## Q-Daten

- Def.: Selbstauskünfte in Befragungssituationen (Erhebung: Interview, Fragebogen)
- Gründe für häufige Verwendung v. Fragebögen
  - Oft einzigste Mögl., hohe Augenscheinvalidität -> Motivation, Ökonomie, Standardisierung, hohe Objektivität, Testtheorie-fähig, gute Reliabilitäten (aber weniger als Leistungstests)
- Arten v. Persönlichkeitstests
  - Testbatterien für ganze Persönlichkeit, z.B. FPI-R, NEO-FFI, NEO-PI-R, MMPI, BIP
  - Einzelne Dimensionen: STAXI, Rosenberg Self-Esteem-Scale, FKK (Kompetenz + Kontrollüberzeugung)
  - State-Tests: z.B. STAI (State-Trait-Angst-Inventar)
  - Interessens-, Motivations- und Einstellungstests

## 7 – Diagnose d. Studieneignung

### Rahmenbedingungen

- Rechtliche Grundlagen: §12 „Recht auf freie Berufs-, Arbeitsplatz und Ausbildungsstättenwahl“
  - NC gerechtfertigt, wenn Kapazitäten d. Bildungseinrichtung nachweislich erschöpft + „sachgerechte“ Bewerberauswahl erfolgt
  - Auswahlkriterien: Abinote, gewichtete Einzelnoten, Studierfähigkeitstest, Berufsausbildung, Gespräch mit der Hochschule oder Kombination
- ⇒ Länderspezifische Hochschulgesetze (Lokale Unterschiede in Auswahlstrategie)
- Organisatorische Grundlagen
  - Bewerberzahlen: 12.500 Bewerber auf 2100 Plätze (41 Hochschulen)  
=> 5 Bewerbungen/Person
  - Freie Wahl d. Auswahlverfahrens, Mögl. d. Unterstützung durch die DPG

### Prädiktoren Studienerfolg

- Allg. Anforderungen an Diagnostische Verfahren: Gütekriterien, Anforderungsanalyse, Ökonomie, Nutzen, Fairness
- Messung d. Kriteriums: evtl. Noten, Abbruchquote, Zufriedenheit, Inhalte/Erreichen d. Lernziele
- Befragung zu erforderlichen Eigenschaften f. Psychologie-Studienplatz
  - Soz. Kompetenz, Interesse am Menschen/d. Psychologie
  - math. Fähigkeiten, naturwissenschaftl. Interesse, logisches Denken
  - Sprachkenntnisse (Englisch)
- Metaanalysen
  - Baron-Bold (1988) Schulnoten f. versch. Studienfächer
    - Abi-Gesamtnote: .46 (Schwankung f. einzelne Studienfächer .38-.56)
    - Einzelnoten: .07-.34
  - Steyer (2005) Schulnoten f. Psychologiestudenten (Abinote -> Vor-/Diplomsnote)
    - Gesamtnote .39, f. Diplom .28, Einzelnoten niedrigere Korrel (zeitl. entfernter)
  - Schmidt-Atzert (2005):
    - Abinote – Vordiplom: .37 (unkorrigiert), Studiendauer, Studienabbruch
    - Unterschiede zw. Vordiplomern + Abgebrochenen oder Verspäteten: nicht in Abinote + Leistungsmotivation, sondern in N-Wert, Informiertheit + Belastungen
- Auswahlgespräche: Meta-Analyse zu Einstellungsinterviews
  - Schmidt+Hunter (1998): Kriterium berufl. Ausbildungserfolg: IQ .56, Interviews .35
  - Schmidt-Atzert: Metaanalyse Studienerfolg (2008)
    - alle Schulnoten .41-.52, studienfachbezogen .36
    - Studierfähigkeitstests .48

- Strukturierte Interviews .21, unstrukturierte .11
- Gründe f. unterschiedl. Validität i. Vgl. zur Berufserfolg-Vorhersage
  - Messqualität d. Kriteriums (Erfolge), weniger reliable Auswahlgespräche bei Uni
  - Bessere Interviewdurchführung durch Personalfachleute
  - Widerspiegelung v. soz. Kompetenzen durch Gespräche (wichtiger f. Job?)
  - Kürzere Biografie v. Studienplatzbewerbern (weniger Aussagekraft)
- ⇒ Gesamtnote als bester Prädiktor (.40-.50), gewichtete Schulnoten eher nicht sinnvoll
  - Studierfähigkeitstests eher niedrigere Validität als Noten
  - Inkrementelle Validität bei Kombination v. Noten+Tests
  - Studienabbruch eher psychosoziale Erklärungen
  - Interviews eher wenig valide

### Aspekte d. Auswahl

- Ökonomie
  - Zeitl. Und finanzieller Aufwand f. Bewerber, Prüfer und Organisation
  - z.B. Abinote keine Extra-Kosten, Tests f. Anfahrt und Durchführung vielen Stunden gebraucht + Auswertung nötig, Gespräche f. Prüfer Schulungen und Zeitaufwand!
  - Mit recht billigem validem Verfahren Ausgewählte am Ende evtl. „günstiger“ als teure Verbesserung d. Ausbildung, bis schlecht Ausgewählte auf gleichem Stand sind
- ⇒ Vgl. Taylor-Russel
  - Trefferquote/Erfolgsrate wird durch valideres Verfahren bei niedriger BR und niedriger SQ am meisten erhöht
  - Insg. EQ natürlich bei höherer BR höher, und steigend mit niedrigerer SQ
- Verfälschbarkeit
  - Kommerzielles Training f. Studierfähigkeitstest mögl., evtl. Itemdiebstahl und immerwährender Austausch d. Items nötig
  - Verfälschbarkeit auch v. Interview-Antworten (Tipps, Erfahrungen anderer Bewerber, bei Standardisierung immer gl. Fragen, schwere Generierung neuer Fragen)
  - z.B. Fragenkatalog f. Psychologiestudienplatz in München
    - Motivation, interessierende psych. Fragestellungen, bereits gelesene Literatur + Tipps f. Bewerber
- Fairness: Gerechtigkeit d. Abinote, des Testes und der Gespräche?
- Akzeptanz
- ⇒ DGP-Empfehlungen
  - Vorerst nach Abinote auswählen, dann spezifische Studierfähigkeitstests entwickeln (Module) => inkrementelle Validität d. Kombination verwenden!
  - Informationen über Studium und Standort abfragen + Freiwillige Selbsttests

### Module

- Anforderungsanalyse, Abstimmung auf Ausbildungsprofil d. Instituts durch Gewichtungen
- Evtl. Verwendbarkeit f. andere Fächer
- Qualifikationen
  - Fächerübergreifend: Muttersprache, 1 Fremdsprache (meist Englisch), Mathe
  - Psychologisch: soz. Kompetenz, Selbststeuerung d. Lernens, Eigenverantwortlichkeit
- 4 Module
  - Analytisches + schlussfolgerndes Denken (FNV)
  - Psychologierelevante Mathe- und Biokenntnisse (Stochastik, Funktionen, Algebra)
  - Fachspezifischer Verständnistest Psychologie
  - Englishtest