

## 2. Testkonstruktion: Von der Planung zum ersten Entwurf

### 2.1 Merkmalsbereich & Geltungsbereich

<b>Merkmalsbereich</b>	<ul style="list-style-type: none"> <li>- Was soll der Test messen</li> <li>- Zu erfassende Konstrukte</li> <li>- Hilfsmittel zu Abgrenzung: Literaturrecherche, qualitative Interviews, strukturierte Varianten der Merkmalsabgrenzung (AFA, CIT)</li> </ul>
<b>Geltungsbereich</b>	<ul style="list-style-type: none"> <li>- Einsatzbereich des Tests</li> <li>- Entscheidungsspielraum häufig durch Auftraggeber eingeschränkt</li> <li>- Einhaltung ethischer Richtlinien liegt bei Testautorin</li> </ul>

Strukturierte Varianten der Merkmalsabgrenzung

<b>AFA</b>	act frequency Approach: Facettentheorie
<b>CIT</b>	critical incident technique: Methode der kritischen Ereignisse

### 2.2 Festlegung des generellen Konstruktionsprinzips

<b>Rational</b>	Deduktiv & theoriegeleitete Konstruktion
<b>External</b>	empirisch & kriteriumsorientierte Strategie
<b>Internal</b>	induktiv & faktorenanalytisches Vorgehen
<b>Seltene Prinzipien</b>	Typologisierende & Prototypenansatz

*Rationale Testkonstruktion*

<b>Ausgangspunkt</b>	Definition des Zielkonstrukts
<b>Deduktion</b>	Ableitung der spezifischen (direkt beobachtbaren oder ertragbaren) Indikatoren aus der allgemeinen Definition
<b>Theoriegeleitete Entwicklung</b>	Formulierung & Auswahl von Items und Zuordnung von Subtest, erfolgt aufgrund von inhaltlichen Erwägungen und nicht hauptsächlich auf Grundlage empirischer Daten

Retrospective Behavioral Self-Control Scale

<b>Definition Selbstkontrolle</b>	Verhaltensnah & utilitaristisch: Meidung von Handlungen die langfristig eher schaden als nutzen
<b>Itemgenerierung</b>	Items fragen konkret nach solchen Handlungen, aber nicht nach Generalisierungen/ Einstellungen

## Berliner Intelligenzstruktur - Test (BIT)

<b>Grundannahme</b>	Hierarchische Struktur der Intelligenz
<b>Bimodale Intelligenzleistungen</b>	Kombination mindestens einer <i>inhaltsgebundenen</i> und einer <i>operationaler</i> Komponente

## Externale Testkonstruktion

<b>Grundproblem</b>	Atheoretische Anpassung an die Stichprobe
<b>Nachteile</b>	<ul style="list-style-type: none"><li>- schwer interpretierbare Skalen</li><li>- Erforderlichkeit großer Stichproben</li><li>- Kreuzvalidierung</li></ul>
<b>Vorteile</b>	Gewährleistet auch bei geringem theoretischen Vorwissen eine gewisse kriterienbezogene Validität und prakt. Nutzen
<b>Kriterienorientierung</b>	Auswahl der Items nach dem Prinzip der Maximierung des empirischen Zusammenhang mit einem Kriterium

## Minnesota Multiphasic Personality Inventory

<b>Kontrastgruppenmethode</b>	Auswahl der Items aus großem Pool, welche signifikant zwischen psychisch auffälligen/ kranken & psychisch unauffälligen Gruppen unterschieden
-------------------------------	---

## CPI-So: California Psychological Inventory - Socialization Skala

<b>Ausgangspunkt</b>	Entwicklung der CPI-So zunächst unabhängig von CPI um psycholog. Prädisposition von kriminellen Verhalten zu identifizieren
<b>Itemauswahl</b>	Ausschließlich nach dem empirischen Kriterium der Unterscheidung überführter Delinquenten von unauffälligen Kontrastgruppen
<b>Ergebnis der Auswahl</b>	<ul style="list-style-type: none"><li>- Inhaltliche sehr heterogene Skala</li><li>- Validität zur Vorhersage devianten Verhaltens ist empirisch gut belegt</li></ul>

## Internale Testkonstruktion

<b>Ausgangspunkt</b>	Generierung eines Itempools, der häufig breiten Merkmalsbereich abdeckt
<b>Untersuchung der Items</b>	<ul style="list-style-type: none"><li>- Items werden einer Stichprobe zur Bearbeitung vorgelegt</li><li>- Die erhaltenen Daten werden hinsichtlich der Korrelationsmuster untersucht (explorative Faktorenanalyse)</li></ul>
<b>Ergebnis</b>	Mehrdimensionale Inventare, deren Subskalen jeweils homogene Konstrukte messen. Insgesamt wird ein breiter und heterogener Merkmalsbereich abgedeckt.

Fünf-Faktoren-Modell (z.B. Five Factor Personality Inventory/ FFPI)

<b>Ausgangspunkt</b>	Idee des <i>lexikalischen Ansatzes</i> : Menschliche Eigenschaften von überindividueller Bedeutung schlagen sich als Adjektive in der natürlichen Sprache wieder.
<b>Vorgehen</b>	An die Stelle „nackter“ Adjektive treten kurze Statements

*Seltene Prinzipien der Testkonstruktion*

<b>Typologisierende Testkonstruktion</b>	<ul style="list-style-type: none"> <li>- Verbindung von Elementre der externalen und internalen Strategie</li> <li>- Identifikation von Personengruppen/ Objektgruppen auf Grundlage empirischer Information</li> <li>- Dominierendes statistisches Verfahren: Clusteranalyse</li> <li>- Beispiel: Myers-Briggs-Type Indikator</li> </ul>
<b>Prototypenansatz</b>	- Ähnlichkeit mit idealtypischem Vertreter wird als kognitiver Ankerreiz für die Einschätzung anderer Objekte vorgegeben

**2.3 Generierung eines Itempools**

<b>Inhaltliche Aspekte</b>	Was wird gefragt
<b>Formale Aspekte</b>	Sprachliche Formulierung, Festlegung des Antwortformats
<b>Objektive Infos</b>	z.B. In Leistungstest
<b>Subjektive Infos</b>	Selbst- vs. Fremdeinschätzung

*Inhaltliche Aspekte*: Persönlichkeitsitems nach Angeleinter/ John/ Löhr

<b>Reaktionen (Offen, verborgen, Symptome)</b>	<b>Einstellungen &amp; Überzeugungen</b>
<b>Eigenschaftszuschreibungen</b>	<b>Reaktionen anderer auf die eigene Person</b>
<b>Wünsche/ Interessen</b>	<b>Bizarre Items</b>
<b>Biografische Fakten</b>	

International Personality Item Pool ist unter Einhaltung bestimmter Standards frei zugänglich. Bei neuen Testverfahren wird in der Regel ein neuer Itempool generiert.

„Regeln“ bei Generierung eines Itempool:

- jedes Item sollte innerhalb eine definierten Bereichs liegen
- Die Itemmenge sollte diesen Bereich repräsentativ abdecken

*Act Frequency Approach*: Facettentheorie, oder auch Ansatz der Verhaltenshäufigkeiten. Stellt eine Art Metatheorie der Persönlichkeit dar.

<b>Dispositionen</b>	Erfassung als Kognitive Kategorien, wobei Verhaltensweisen nach der Häufigkeit ihres Auftretens (ohne eigentlichen Erklärungswert) zusammengefasst werden
<b>Vorgehensweise</b>	<ul style="list-style-type: none"> <li>- Psychologischen Laien wird eine Eigenschaft vorgegeben und sie sollen sich eine Person aus dem Bekanntenkreis mit dieser Eigenschaft (in besonders hoher Ausprägung) vorstellen</li> <li>- Anschließend sollen sie konkrete Verhaltensweisen aufschreiben, die sie bei ihnen beobachtet haben</li> <li>- Die daraus entstehende Itemsammlung wird einer weiteren unabhängigen Gruppe vorgelegt und von dieser hinsichtlich ihrer Prototypizität bewertet</li> </ul>
<b>Bewertung der AFA</b>	<ul style="list-style-type: none"> <li>- AFA erlaubt Generierung &amp; Zuordnung von Items zu Skalen</li> <li>- Beurteilungen der Prototypizität stimmen nach empirischen Befunden meist gut überein</li> <li>- Unterschiedliche Eigenschaften in nach AFA konstruierten Tests korrelieren tendenziell höher miteinander, als bei konventionell konstruierten Persönlichkeitsinventaren</li> </ul>

Critical incident technique: Methode der krischen Ereignisse. Entammt der Organisationspsychologie, wurde ursprünglich als Verfahren der beruflichen Leistungsbeurteilung konzipiert.

<b>Ansatzpunkt</b>	Gute/ unbefriedigende Leistungen schlagen sich in konkreten erfolgskritischen Ereignissen nieder
<b>Kritische Ereignisse</b>	<ul style="list-style-type: none"> <li>- werden durch Einschätzungen von Befragungen konstruiert, welche vom Testautor unabhängig sind</li> <li>- Meist umfangreicher als AFA</li> <li>- Einschluss des situativen Kontextes mit Vor- /Nachgeschichte</li> </ul>
<b>Bewertung der CIT</b>	<ul style="list-style-type: none"> <li>- Eignungsdiagnostische Instrumente haben sich hinsichtlich Validität zur Vorhersage beruflicher Leistung empirisch bewährt</li> <li>- Erlaubt Abgrenzung des relevanten Merkmalsbereichs, wenn das mangels Daten (mit den Mitteln der externalen Testkonstruktion) nicht möglich ist</li> </ul>

## 2.4 Sprachliche Formulierungsregeln und Antwortformate

Itembestandteile

<b>Itemstamm</b>	Eigentliche Frage/ Aufgabe/ Reiz
<b>Antwortformat</b>	Mögliche Reaktionen/ Antworten/ formale Aspekte

Formulierungsrichtlinien (unvollständig)

<b>Verständlichkeit</b>	<ul style="list-style-type: none"> <li>- Vermeidung doppelter Verneinung</li> <li>- Vermeidung von Fachausdrücken</li> <li>- Sachverhalt so kurz wie möglich und so lange wie nötig</li> </ul>
<b>Eindeutigkeit</b>	<ul style="list-style-type: none"> <li>- Zugrundelegen von jeweils nur einem sachlichen Gedanken</li> <li>- Vermeidung starker Verallgemeinerung</li> <li>- Verwendung eindeutiger zeitlicher &amp; situativer Bezüge</li> </ul>
<b>Meidung von Antworttendenzen</b>	<ul style="list-style-type: none"> <li>- Vermeidung von Suggestivfragen und unnötig wertender Formulierung</li> <li>- Meidung von Gemeinplätzen</li> </ul>

Aufgabentypen

<b>Freies Format</b>	<ul style="list-style-type: none"> <li>- Kurzsatzaufgaben</li> <li>- Ergänzungsaufgaben</li> </ul>	
<b>Gebundenes Format</b>	<ul style="list-style-type: none"> <li>- Ordnungsaufgaben</li> <li>- Auswahlaufgaben</li> <li>- Beurteilungsaufgaben</li> </ul>	<ul style="list-style-type: none"> <li>- Zuordnungsaufgaben</li> <li>- Umordnungsaufgaben</li> <li>- Dichotomie Aufgaben</li> <li>- Mehrfachauswahl</li> <li>- Analogskalaaufgaben</li> <li>- Ratingskalen</li> </ul>
<b>Atypische Formate</b>		

Ratingskalen: Elemente

<b>Differenzierung der Abstufung</b>	<ul style="list-style-type: none"> <li>- dreistufig, vierstufig, zehnstufig...</li> <li>- Psychometrische Optimum zw. 5-9 Stufen</li> <li>- Sonderfall: Stufenlos einstellbare Analogskala</li> </ul>
<b>„Begrenzungen“</b>	<ul style="list-style-type: none"> <li>- unipolar: logisches Minimum/ Nullpunkt</li> <li>- Bipolar: Extremausprägungen nach oben und unten</li> </ul>
<b>Skalenniveau</b>	<ul style="list-style-type: none"> <li>- Es ist fraglich ob mit mehrstufigen Ratingformaten ein metrisches Skalenniveau erreicht werden kann (Analogskala: ja)</li> <li>- Unterstellung Intervallskala vertretbar, desto             <ul style="list-style-type: none"> <li>- Mehr Skalenniveaus pro Item vorgegeben werden</li> <li>- Mehr items zu einem Testwert aggregiert werden</li> <li>- Gleichmäßiger die Abstände zwischen den Skalenniveaus „sind“</li> </ul> </li> </ul>

Kernprobleme von Ratingskalen:

<b>Soziale Erwünschtheit</b>	Tendenz im Sinne sozialer Normen zu antworten
<b>Akquieszenz</b>	Generelle Tendenz zur Zustimmung

Mehrfachwahlaufgaben

<b>1 aus y richtig</b>	Genau eine Antwort ist richtig. Antworten müssen disjunkt sein (keine Schnittmengen)
<b>x aus y richtig</b>	Zahl der richtigen Antworten unbekannt
<b>Distraktoren</b>	Falsche Antwortalternativen
<b>Forced-choice-Item</b>	Antwortalternativen die bei freier Auswahl ähnlich häufig bevorzugt würden, werden zusammen gruppiert, mit der Instruktion genau eine auszuwählen