

Fragebogen 3

Einführung in die Psychologische Diagnostik, Westmeyer

Ausarbeitung: Johannes Geffers [auf Grundlage vieler Vorarbeiten...]

1. Worin sehen Sie die wesentlichen Unterschiede zwischen L-Daten, Q-Daten und T-Daten und was lässt sich aufgrund welchen Datentyps am besten erfassen?

- Der wesentliche Unterschied liegt darin, dass sie mittels verschiedener Verfahren erfasst werden. Damit verbinden sich unterschiedliche Vorstellungen zu deren Einsetzbarkeit.
- **T-Daten** (Test-Data aus ›objektiven‹ Tests).
 - resultieren aus objektiven Tests (objektiv heißt mangelnde Durchschaubarkeit der Messintention, wie sie z.B. bei projektiven Tests zu finden ist)
 - dienen der Erfassung von Fähigkeiten, Temperament, Persönlichkeitseigenschaften, Interessen
 - Die Messintentionen bleiben dem Pb absichtlich verborgen; Proband weiß nicht, welcher Aspekt des Verhaltens beobachtet wird • bewusste Verfälschungstendenzen werden so vermieden
 - Erfassung des unmittelbaren Verhaltens eines Individuums in einer standardisierten Situation, ohne dass dieser sich selbst beurteilen muß
 - Geeignete Dimensionen: *Fähigkeiten, generelle Persönlichkeitseigenschaften, Temperament, Dynamische Eigenschaften*
- **Q-Daten** (Questionnaire Data, subjektive Testdaten, aus Antworten aus Beratungsgesprächen, Fragebögen, ›Selbst-Ratings‹).
 - werden über Fragebögen erhoben
 - basieren auf Introspektion und Selbstbeurteilung
 - zentrale Bedeutung: Selbsteinschätzung (self report)
 - Geeignete Dimensionen: Erfassung von *Temperament, Persönlichkeitseigenschaften*
 - anfällig für Verfälschungen
 - Faking bad und faking good problemlos durchführbar, weil subjektive Tests (also v.a. Fragebögen und Selbstbeurteilungsskalen) in ihrem Aufbau, ihrer Zielsetzung und Interpretation relativ leicht durchschaubar sind.
- **L-Daten** (Life-record Data [Lebensprotokolldaten], Beobachtung in natürlicher Umgebung).
 - werden durch Fremdeinschätzung oder Selbsteinschätzung im Rahmen eines Interviews erfasst
 - Daten sind durch Beobachterperspektive gefiltert.
 - Geeignete Dimensionen: Erfassung von Fähigkeiten, Temperament, Persönlichkeitseigenschaften, Interessen
- Eine Beschränkung auf eine Datenebene ist mit den Erfordernissen diagnostischer Praxis in der Regel unvereinbar, **Multimodalität** und damit **Multimethodalität** werden zum Standard in der Diagnostik.
 - **mulimodal** = Die diagnostischen Informationen werden auf mehreren Datenebenen (Modalitäten) erhoben. Unterschieden werden: vergangenes und gegenwärtiges Erleben, vergangene und gegenwärtige Leistungen, psychophysiologische, neurophysiologische und biochemische Reaktionen

2. Erläutern Sie den Unterschied bei T- und Q-Daten zwischen üblichen Interpretationen des Antwortverhaltens und seiner Auffassung als checking behavior.

- ◆ Die **übliche Interpretation des Antwortverhaltens** bei T- und Q-Daten basiert auf der Annahme der Richtigkeit des Antwortverhaltens eines Probanden, d.h. der Übereinstimmung zwischen Ankreuzverhalten und der Ausprägung des gemessenen psychologischen Konstrukts.

- ◆ Die Auffassung des Antwortverhaltens als **checking behavior** erwächst aus der Annahme, dass sich das psychologische Konstrukt, das hinter dem Messinstrument steht, durch das Antwortverhalten manifestiert. D.h. checking behavior wird als Indikator für das dahinterstehende psychologische Konstrukt verstanden. Dabei wird nicht unüberprüft von der Richtigkeit des Antwortverhaltens eines Probanden ausgegangen, sondern von instruktionsbezogenem Antwortverhalten dieses Probanden bei T- und Q-Daten. Bei T-Daten wird der Bezug zum Konstrukt als unproblematisch unterstellt.
 - Beim checking behavior muß der Bezug zum Konstrukt erst über eine Konstruktvalidierung hergestellt werden. Zunächst wird das Verhalten lediglich als instruktionsbezogenes Antwortverhalten gesehen.


3a. Mit welchen Validitätswerten für diagnostische Verfahren ist bei der Vorhersage von Erfolgskriterien in unterschiedlichen Bereichen zu rechnen?

- Niedrige empirische Validitäten psychometrischer Persönlichkeitstests im Bereich der Personalpsychologie: Leistungsbeurteilungen, r_{tc} 0.206; Fluktuation, r_{tc} 0.121; Ausbildungsleistung, r_{tc} 0.152; Statusveränderung, r_{tc} 0.126; Gehalt, r_{tc} 0.268

3b. Worauf ist der Umstand zurückzuführen, dass es so viele konkurrierende Vorstellungen zur Struktur der Intelligenz gibt?

- Intelligenz ist eine nicht unmittelbar beobachtbare Eigenschaft.
- Intelligenz ist eine soziale und methodische (FA) Konstruktion. So werden etwa von verschiedenen Autoren verschiedene Grundfaktoren angenommen.
- Intelligenz ist ein komplexes Phänomen und kann sich verschieden ausdrücken (z.B. analytisch, praktisch, kreativ).

◆ methodische Gründe (Methodenvarianz!):

1. **Variablenauswahl** (Welche Variablen setze ich ein? breites oder schmales Herangehen?)
 - Intelligenz ist eine soziale und persönliche Konstruktion, d.h. je nachdem aus welcher Perspektive sie betrachtet wird und welche Grundfaktoren dem Intelligenz-Konstrukt zugeordnet werden, bestehen verschiedene Intelligenz-Struktur-Theorien.
 - Intelligenzkonzepte sind jeweils der Versuch, die Komplexität dieses Phänomens zu ordnen
 - ◆ Beispiele:
 - WECHSLER: verbale Intelligenz  allgemeine Intelligenz
 - CATTELL: geht bis zu Faktoren 3. Ordnung sekundär Faktor
 - GUILFORD: Würfel-Schema: Operationen, Produkte, Inhalte
 - EYSENCK: 3 Intelligenzen A, B, C
 - A: biologisch-genetisch / B: Alltagsintelligenz / C: psychometrische Intelligenz
 - Konkurrierende Vorstellungen zur Struktur der Intelligenz haben gemein, dass fast alle Konstrukte der differentiellen Psychologie (z.B. Intelligenz) nicht vollständig operational definiert sind.
 - Sie bieten aufgrund ihres Bedeutungsüberschusses die Möglichkeit, Hypothesen abzuleiten, die sich wiederum auf empirische Sachverhalte (offene Konstrukte) beziehen.
2. **Personenstichprobe** (homogen vs. heterogen).
 - Welche Gruppe von Personen bildet meine Population?
 - heterogene Stichprobe (von der Wiege bis zur Bahre)
 - homogene Stichprobe (alle Jugendlichen)
 - Durch Stichprobeneffekte ergeben sich unterschiedliche Strukturmodelle.
3. **Extraktionskriterien** (z.B. principle components, principal axis factoring, maximum likelihood).
 - Wieviele Faktoren extrahiere ich?
 - Kriterien für die Extraktion: Screeplot/ Screeplot
 - Kaiser-Guttman-Kriterium
 - Eigenvalue
 - Aus der unterschiedlichen Anzahl der Faktoren ergeben sich unterschiedliche Strukturen.

4. Rotationsmethode.

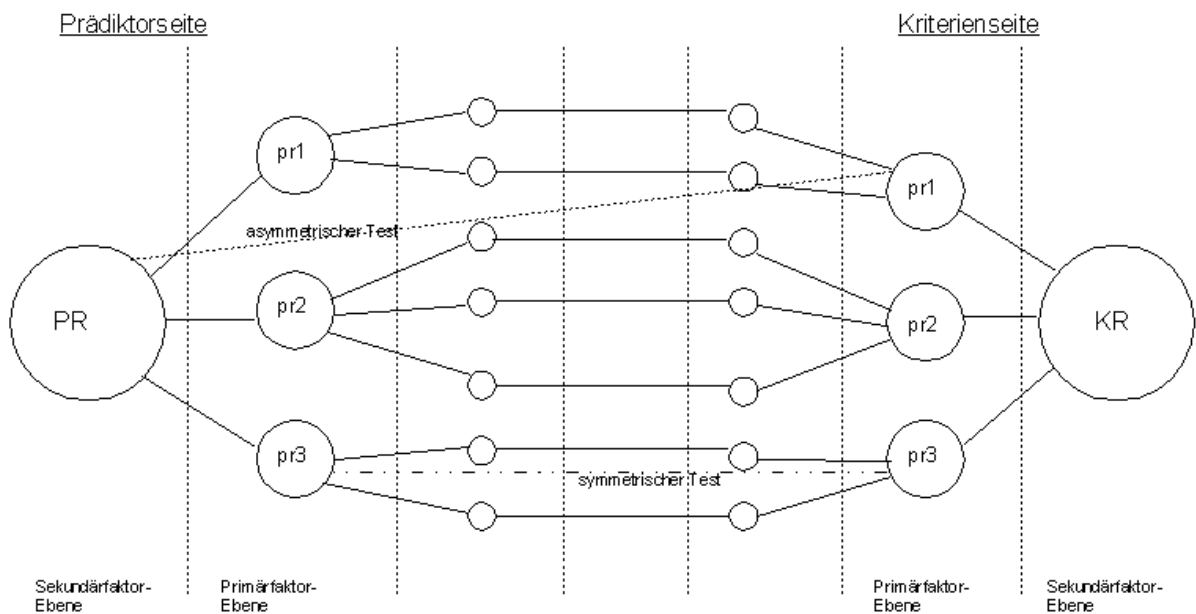
- Die Faktoren können unterschiedlich rotiert werden: orthogonal oder oblique.
- Je nach Rotation können unterschiedliche Faktoren resultieren.

5. exploratorische oder konfirmatorische FA

- In den Aspekten 1-4 und den vielfältigen Kombinationen der Aspekte miteinander liegt die große Variationsmöglichkeit/Kombinationsmöglichkeit. So ergeben sich sehr unterschiedliche Methoden in die das Konstrukt Intelligenz eingebettet ist.

4. Was ist unter symmetrischen, was unter asymmetrischen Prädiktor-Kriteriums Beziehungen zu verstehen?

- Die Unterscheidung bezieht sich auf die **Aggregationsebene**.
- Eine **symmetrische Beziehung** zwischen **Prädiktor und Kriterium** ist gegeben, wenn beide auf derselben Ebene der Aggregation der Faktorenanalyse angesiedelt sind (vergleichbare Komplexität; z.B. sowohl Prädiktor als auch Kriterium auf der Ebene von Primärfaktoren).
 - Symmetrie impliziert höhere Validitätswerte, d.h. je symmetrischer desto höher die Validität. Auch höhere Stabilität der Messungen durch Aggregation über Messzeitpunkte hinweg.
- Auf der **untersten Ebene** liegen die Testitems. Die Ebene, auf der eine bestimmte Anzahl von Prädiktoren eine bestimmte Anzahl von Kriterien durch jeweilige Testitems vorhersagen, wird **Primärfaktorebene** genannt.
- Werden diese Gruppen von Faktoren und Prädiktoren aggregiert zu neuen Prädiktoren und Kriterien, so befindet man sich auf der **Sekundärfaktorebene**.
- Beispiel / Veranschaulichung.



Hierarchisches Brunswik'sches Linsenmodell

A) Intelligenz

- Intelligenz — Einzelne Schulnote (geringe Validität, da asymmetrisch)
- Intelligenz — nach Fachgebieten zusammengefasste Schulnoten (höhere Validität, da symmetrisch)

B) Berufserfolg

- Berufserfolg — asymmetrisch — Selbstwirksamkeitsskala von Schwarzer (Skala aus 10 Items) (niedrig aggregiert)
- Berufserfolg (hoch aggregiert) — • geringe Validität der Vorhersage des Berufserfolgs allein auf der Basis der Selbstwirksamkeitsskala

- Zum Berufserfolg tragen verschiedene Faktoren bei, Selbstwirksamkeit ist nur eine davon

- ◆ **Asymmetrische Prädiktor-Kriterium-Beziehung:** Korrelation von Prädiktor und Kriterium auf unterschiedlichen Faktorebenen.
 - Bsp.: Korrelation von Intelligenztestleistung (Prädiktor, Sekundärfaktorebene) und Note im Mathetest (Kriterium, Primärfaktorebene)
- ◆ **Vergleich symmetrisch/asymmetrische Prädiktor-Kriterium-Beziehung.**
 - Optimal wäre es, wenn Prädiktor und Kriterium auf derselben Aggregationsebene angesiedelt sind. Dies impliziert höhere Validitätswerte.
 - Je höher aggregiert, umso bessere Möglichkeit der Vorhersage (höhere Reliabilität und Validität).
 - Je symmetrischer, umso höhere Validitätskoeffizienten.

5. Erläutern Sie den Grundgedanken des adaptiven Testens. Worin sehen Sie die Vorzüge dieser Vorgehensweise? Warum ist das adaptive Testen auf probabilistische Testmodelle angewiesen und kann sich nicht auf die KTT stützen?

- Grundgedanke ist eine Strategie der Testvorgabe, bei der sich die **Auswahl der Items am Leistungsniveau** (oder der Reaktion) **des Pbn** orientiert.
- **Vorgehen:** Beim adaptiven Testen werden Aufgaben gleichen Schwierigkeitsgrades in eine Testgruppe aufgenommen (optimal 5 Aufgaben). Je nachdem wie viele Aufgaben in einer Gruppe gelöst werden, wird anschließend eine schwierigere, gleich schwierige oder leichtere Testgruppe vorgelegt bis ein festgelegtes Abbruchkriterium erreicht ist (z.B. 15 Items).
- **Vorteile** sind die Ökonomie des Verfahrens und höhere Messgenauigkeit. Durch das probanden-spezifische Vorgehen kann aus weniger Items mehr Information gewonnen werden. Durch die Verkürzung des Verfahren wird die Motivation und Konzentration der Pbn positiv beeinflusst.
- **Problematisch** ist das Erstellen eines umfangreichen Aufgabenpools, der mit dem RASCH-Modell konform sein muss (d.h. die Anzahl gelöster Aufgaben stellt eine erschöpfende Statistik dar).
- Die Items müssen gut differenzieren und einen weiten Schwierigkeitsbereich abdecken.
- ◆ **Stichprobeninabhängigkeit und spezifische Objektivität** (notwendige Vorzüge der PTT gegenüber der KTT):
 - Im Rahmen der adaptiven Testens sowie in der PTT hängt der Vergleich zweier Versuchspersonen nicht von der Itemstichprobe ab (Stichprobenunabhängigkeit). Dieser Vergleich zweier Personen bzw. einer Person mit dem Kriterium hinsichtlich eines Merkmals nennt sich dann spezifisch objektiv im Sinne von RASCH, wenn das Ergebnis des Vergleichs nicht von dem für den Vergleich benutzten Messinstrument (also der Itemauswahl) abhängt.
 - Folglich ermöglichen diese Grundannahmen der PTT es der adaptiven Diagnostik unterschiedliche Vpn auf derselben Dimension zu lokalisieren, obwohl sie im Rahmen der Testung verschiedene Items bearbeitet haben.
- ◆ **Fazit oder Warum ist adaptives Testen auf IRT-Modelle angewiesen?**
 - Adaptives Testen, z.B. branched testing, ist innerhalb der KTT nicht möglich, denn benötigt wird die spezifische Objektivität. Es wird ausgegangen von wechselseitiger Stichprobenunabhängigkeit. Bei der KTT können Personen nur verglichen werden, wenn sie dieselben Items erhalten.
 - Gütekriterien der KTT greifen nicht: Die Aufgabenfolgen sind individuell verschieden, es gibt also für jeden Pb eine bestimmte Wahrscheinlichkeit, wie der nächste Aufgabenblock gelöst werden kann. Und es ist keine Standardisierung der Pb möglich.
- **Ausführlichere Erläuterung.**
 - ◆ ›Konventionelle Herangehensweise‹.
 - Damit eine Aussage als wissenschaftlich gelten kann, muß sie der Forderung nach Generalisierbarkeit entsprechen, d.h. beispielsweise, dass die Ergebnisse, die ein Pb zu einem oder mehrerer Zeitpunkte erzielt, nicht unkontrolliert schwanken dürfen. Die Testleistung muß – bis auf einen un-

vermeidlichen Testfehler – zuverlässig bestimmbar sein • in der KTT ist die Reliabilität das Maß für diese Zuverlässigkeit, und ein Teil der Reliabilität ist abhängig von der Objektivität.

- Die KTT geht vom Vorliegen eines Testergebnisses (als Messung einer zugrundeliegenden Eigenschaft) aus und setzt damit Testergebnis und zugrundeliegende Eigenschaft de facto gleich. Im Rahmen dieses Ansatzes wird maximale Objektivität am ehesten durch genaue Prozedurregeln gewährleistet, die möglichst viele Störregeln ausschalten. • Jeder Pb muß in derselben Weise instruiert werden, dieselben Items in gleicher Reihenfolge und unter gleichen Bedingungen bearbeiten; die Testantworten müssen in derselben Weise ausgewertet und interpretiert werden.
- ◆ **Nachteile der konventionellen Strategie.**
 - Die Testökonomie ist nicht optimal, weil Zeit für wenig informative Items verwendet wird.
 - Die für alle Pbn *identische Itemmenge* enthält nur einen Bruchteil von Aufgaben, die für eine konkrete Person relevant sind: Sie enthält Aufgaben, die leistungsstarke Pbn (nahezu) mit Sicherheit lösen und leistungsschwache nicht.
 - Die Standard-Itemmenge berücksichtigt auch nicht die konkrete diagnostische Fragestellung: Wenn z.B. nur nach Eignung gefragt wird, wird bei leistungsstarken und –schwachen Pbn zuviel, in Grenzfällen aber unter Umständen zu wenig Information erhoben.
 - Sie sind nicht optimal, was die Messgenauigkeit betrifft, weil sich die Messung nicht an der konkreten Fragestellung orientiert.
 - Sie sind nicht optimal, was die Motivation des Pb betrifft, weil leistungsstarke Pbn gelangweilt und leistungsschwache entmutigt werden können
- ◆ **Adaptive Test-Strategien** versprechen einen Ausweg aus diesem Dilemma.
 - Unter **adaptivem Testen** („maßgeschneidertes Testen“ (*tailored testing*)) werden die Strategien der Testvorgabe verstanden, bei denen sich die Auswahl der Testitems am Leistungsniveau der betreffenden Person (z.B. altersspezifischer Testeinstieg) oder an der Reaktion des Pb während der Testdurchführung orientiert: Gibt eine Person falsche Antworten, folgen leichtere Items, bei richtigen schwierigere.
 - ◆ **Vorteile des adaptiven Testens.**
 - Ökonomie des Verfahrens. Wegen des probandenspezifischen Vorgehens • Rückgriff auf begrenzte Itemzahl • Konzentration auf die informativen Items.
 - Positive Beeinflussung der Konzentration und Motivation durch *Verkürzung des Verfahrens*: leistungsstarke Pbn sind in konventionellen Tests gelangweilt, leistungsschwache eher entmutigt • wenn Person in Aufgabenschwierigkeitsgruppe bleibt, kann nach dem 2. Durchgang abgebrochen werden.
 - Höhere Messgenauigkeit, weil es weniger Grenzfälle und mehr diagnostische Informationen in adaptiven Tests gibt; es werden für alle Trait-levels nur sehr geringe SMF gemacht.
 - ◆ **Vorgehen.**
 - **branched testing.**
 - Einzelne Items werden zu Gruppen zusammengefasst, die den gleichen Schwierigkeitsgrad haben.
 - Itemgruppe enthält z.B. 5 gut diskriminierende Aufgaben (geringe, mittlere u. hohe Schwierigkeit).
 - Je nach dem, wie viele (0-1, 2-4, 5) und wie schwierige Items gelöst wurden, wird Pb zur nächsten Itemgruppe weitergeleitet.
 - Pb muss von 60 Aufgaben nur ca. 15 lösen, um Fähigkeitsbestimmung vornehmen zu können.
 - **tailored testing.**
 - computergestütztes Verfahren sinnvoll, denn für eine optimale adaptive Strategie muß:
 - jede einzelne Antwort sofort verrechnet werden,
 - es muß der Personenparameter neu geschätzt,
 - das Konfidenzintervall dieser Schätzung bestimmt,
 - über Fortsetzung oder Abbruch der Testung entschieden und
 - bei Fortsetzung das für die erzielte Schätzung informativste Item ausgewählt werden.
 - All dies möglichst rasch, weil der Pb durch lange Wartezeit nicht gelangweilt und der Ökonomiegewinn nicht gefährdet werden soll
 - ◆ **Unvereinbarkeit von KTT und adaptivem Testen.**
 - Es ergeben sich 3 grundlegende testtheoretische Fragen:
 1. **Messwertskala.** Wie können 2 Pbn miteinander verglichen werden, wenn beide unterschiedliche Itemmengen mit unterschiedlichen Schwierigkeiten beantwortet haben? • Frage nach der Messwertskala
 2. **Itemauswahl.** Welches Item soll als jeweils nächstes vorgelegt werden? • Frage nach der Itemauswahl
 3. **Abbruchkriterium.** Wann ist das Leistungsniveau des Pb hinlänglich genau bestimmt, so dass die Testung beendet werden kann? • Frage nach dem Abbruchkriterium
 - Diese sind mit der KTT nicht zufriedenstellend beantwortbar:

1. Die KTT setzt Testergebnisse und latente Dimensionen de facto gleich, also erzeugen unterschiedliche Itemmengen neue Skalen, die nicht ohne Weiteres ineinander überführbar sind • es fehlen Kriterien, wann Items als homogen anzusehen sind.
 2. Die Definition der Reliabilität als Verhältnis der Varianz der wahren Werte zur Gesamtvarianz des Tests bezieht sich auf eine Referenzpopulation • ändert diese sich, ändert sich die Reliabilität und scheinbar auch die Genauigkeit des Testergebnisses.
 3. Die Reliabilität impliziert, dass ein Test bei allen Personen gleich misst • tatsächlich aber bei sehr leistungsstarken und –schwachen ungenauer als bei durchschnittlich Pbn.
- ◆ **PTT als Voraussetzung für adaptives Testen.**
- **Ausgangspunkt:** Das *beobachtbare Testverhalten* wird als Ausdruck *latenter Eigenschaften* aufgefasst, und aus den beobachteten Reaktionen sollen Rückschlüsse auf die Eigenschaften gezogen werden. Solche Rückschlüsse sind aber nicht einfach, weil die Gesetzmäßigkeiten des Zusammenhangs zwischen *Reaktionen*, *latenten Eigenschaften der Person* und *Itemparametern*, wie der Itemschwierigkeit, nicht von vornherein bekannt sind.
 - Die probabilistische Testtheorie beruht auf Modellvorstellungen, wie die Wahrscheinlichkeit für eine bestimmte Reaktion von Parametern der Person und des Items abhängt.
- ◆ Auf die obengenannten 3 Fragen bietet die PTT befriedigende Lösungen:
1. **Messwertskala.** Die Schätzung des Personenparameters stellt eine Messung einer latenten Eigenschaft dar. Unterschiedliche Ausprägungen können damit verglichen werden, auch wenn die Pbn nicht dieselben Itemmengen bearbeiten.
 2. **Itemauswahl.** Für jedes Item kann angegeben werden, welchen Beitrag es für die Einschränkung des Konfidenzintervalls der Messung leistet. So ist eine optimale Itemauswahl möglich.
 3. **Abbruchkriterium.** Die Vorgabe von Items kann abgebrochen werden, wenn das Konfidenzintervall der Messung einen vorgeschriebenen Wert unterschreitet.
- ◆ **Adaptives Testen und IRT.**
- **spezifische Objektivität:** Der Vergleich zweier Personen hinsichtlich ihrer Fähigkeit ist nicht abhängig davon, dass diesem Vergleich die gleichen Items zugrunde liegen (muß nicht dieselbe Itemstichprobe sein). Leistungsvergleich ist auch nicht abhängig davon, welche Leistung andere Personen der Bezugsgruppe bei diesem Test erbracht haben.
 - **Wechselseitige Stichprobenunabhängigkeit:** Schätzung von Item- und Personenparameter erfordert nicht, dass in einem Fall die gleiche Personenstichprobe, im anderen Fall die gleiche Itemstichprobe zugrundeliegt
- ◆ **Schwierigkeiten.**
- Hohe Anforderungen an den Itempool. Adaptive Strategien sind umso effektiver, je mehr Informationen die Items in bezug auf den Personenparameter liefern können, d.h. es müssen viele homogene Items mit bekannten Itemcharakteristiken vorhanden sein, die das relevante Schwierigkeitskontinuum etwa gleichmäßig abdecken • i.d.R. muß der Itempool für adaptives Testen eigens konstruiert werden
 - Konformität mit dem RASCH-Modell. Dieser umfangreiche Aufgabenpool muß mit dem RASCH -Modell konform sein, d.h. jede Aufgabe bildet eine erschöpfende Statistik nach dem RASCH -Modell.

6. Wie kann der Prozess der Beantwortung von Items eines Fragebogens verstanden werden? Welche Eigenschaften von Items sind für Verlauf und Ergebnis eines solchen Prozesses von Bedeutung?

- i. Der Trait als Konstrukt beeinflusst den Inhalt der Items.
 - ii. Der Inhalt der Items beeinflusst die von den Pbn wahrgenommene Itemform.
 - iii. Versteht der Pb den Iteminhalt, wird in Abhängigkeit von verbaler Kompetenz und individueller Geschichte eine mentale Repräsentation des Items aufgebaut.
 - iv. Diese Repräsentation enthält lexikalische, syntaktische, semantische und pragmatische Aspekte.
 - v. Die Reaktion auf das Item gibt Aufschluss über die kognitive Repräsentation. Sie wird beeinflusst vom aktuellen Zustand des Traits zum Messzeitpunkt und des persönlichen Antwortstils.
- ◆ **Itemeigenschaften, die den Beantwortungsprozess beeinflussen können sind:**
- Verständlichkeit
 - Eindeutigkeit/Mehrdeutigkeit
 - Implizite soziale Bewertung
 - Abstraktheit
 - Selbstbezug
 - semantische Item-Trait-Beziehungen
 - logische Item-Trait-Beziehungen
 - Oberflächeneigenschaften der Items: Länge, Syntax und Antwortformat
 - Endorsement

◆ **Einige Fehlerquellen bei der Beantwortung von Fragebögen.**

- **Unterschiedliche Interpretation der Items.** Z.B. durch abstrakte Formulierungen und inhaltliche Ambiguität (»Ich habe volles Selbstvertrauen« • von konkreten Situationen und Verhaltensweisen völlig abstrahierende Selbstbeurteilungen).
- **Formal-syntaktische Item-Merkmale.** Doppelte Verneinung, Aktiv-Passiv-Formulierung, lange Sätze, komplexe Formulierungen.
- **Gedächtnisprobleme, Erinnerungsfehler.** Bei Items über alltägliches Verhalten.
- **Defizite der Selbstbeobachtung und –erkenntnis, Selbsttäuschung.** Kontrollskalen: »Ich habe schon mal gelogen!«, »Nie!« • verdächtig.
- **Tendenz zur konsistenten Selbstdarstellung.** Die Beantwortung bestimmter Items wird z.B. auf Antworten zu vorhergehenden abgestimmt; u.U. unter Orientierung an subjektiven Hypothesen impliziter Persönlichkeitstheorien.
- **Anker- und Halo-Effekte** (in impliziten Persönlichkeitstheorien). Manche Items können eine besonders zentrale Stellung mit „breiter Ausstrahlung“ auf weitere Antworten haben.
- **Herstellung von Zusammenhängen zwischen Itemantworten.** Mehr oder weniger unwissentlich durch Erinnerungs-, Selbstbeobachtungs- und Selbstdarstellungsprobleme ›verzerrte‹ Beantwortung.
- **Absichtliche Verfälschung** (Lügen, Simulation, Dissimulation). Kaum auszuschließende Fehlerquelle; selbst bei Leistungstests kann man sich zumindest inkompetenter darstellen (faking bad); bei Persönlichkeitsfragebögen ist das Problem besonders gewichtig, v.a. bei diagnostischen Situationen mit einem Macht-Dilemma (Bewerberauswahl, Rentenbegehren...); Personen erkennen z.T. die Skalenkonstruktion und vermögen so zu antworten, dass sie gezielt hohe/niedrige Werte erreichen.
- **Tendenz zu sozial erwünschten Antworten.** In den SD-Skalen (soziale Erwünschtheit, social desirability, SD) wird versucht, die Neigung zu sozial erwünschten Antworten durch Items zu erfassen, die entweder allgemein übliche, geringfügige Normverletzungen enthalten oder sozial erwünschte Verhaltensweisen die kaum jemand immer zeigt (Notlügen, Streiten, Tischmanieren, Pünktlichkeit). SD-Skalen oder verwandte Skalen über Offenheits-, Lügen-, Leugnungstendenzen sind in der Praxis die häufigsten Kontrollmethoden. Bei extremen SD-Werten sollten die übrigen Fragebogenergebnisse nicht interpretiert werden.
- **Tendenz zur unkritischen Zustimmung (Ja-Sagen).** Tendenz relativ unabhängig vom Item-Inhalt zustimmend zu antworten.
- **Tendenz zur unkritischen Ablehnung (Nein-Sagen).** Tendenz relativ unabhängig vom Item-Inhalt ablehnend zu antworten.
- **Tendenz zu extremen Antworten (Schwarz-Weiß-Malen).** Tendenz bei Stufenantworten aus Unsicherheit, mangelndem Verständnis, geringer Differenzierungsfähigkeit extreme Skalenwerte zu bevorzugen.
- **Tendenz zu indifferenten Antworten (Tendenz zur Mitte, Ambivalenz).** Tendenz bei Stufenantworten aus Unsicherheit, mangelndem Verständnis, geringer Differenzierungsfähigkeit mittlere Skalenwerte zu bevorzugen (Tendenz zur Mitte).
- Der i.d.R. als unproblematisch unterstellte Übergang vom instruktionsbezogenen Ankreuzverhalten bei der Bearbeitung eines subjektiven Tests zum eigentlich interessierenden Persönlichkeitsmerkmal bedarf einer Konstruktvalidierung, bei der auch die sog. Antworttendenzen oder –stile zu berücksichtigen sind.

7. Erläutern Sie an wenigstens 3 Aspekten, warum die übliche Gegenüberstellung von Verhaltensdiagnostik und traditioneller Diagnostik einer näheren Überprüfung nicht standhält?

- Durch eine kontrastierende Charakterisierung des behavioralen und des sogenannten traditionellen Ansatzes in der psychologischen Diagnostik wird die beträchtliche Variation innerhalb beider Ansätze außer Acht gelassen.

1. Persönlichkeitskonzeption.

- In der Verhaltensdiagnostik würden Persönlichkeitskonzepte allenfalls als Zusammenfassung spezifischer Verhaltensmuster dienen. Der traditionelle Ansatz soll von stabilen Eigenschaften ausgehen, die dem Verhalten zugrunde liegen.
- Einwand: Hier ist nicht berücksichtigt, dass im sogenannten traditionellen Ansatz zwischen Traits und States unterschieden wird. Mischel nimmt in seiner „Rekonzeptualisierung der Persönlichkeit“ (1968) fünf Personvariablen an, die als Disposition Ähnlichkeit mit traditionellen Konstrukten haben

2. Verhaltensursachen.

- In der Verhaltensdiagnostik sind Verhaltensursachen aufrechterhaltende Bedingungen in der Umwelt oder Umwelt/Person – Interaktion. Im traditionellen Ansatz lägen die Verhaltensursachen hingegen in der Person.

- Einwand: Auch im traditionellen Ansatz sind Verhaltensursachen situationsspezifisch. Es werden auch probabilistische Zusammenhänge zwischen Situation und Verhalten formuliert. Andererseits gibt es dispositionelle Erklärungen auch im behavioralen Ansatz (z.B. in der Verhaltensgleichung von Kanfer).
- Verhaltensursachen sind im sog. traditionellen Ansatz nicht in kontextfrei und situationsspezifisch konzipierten Personeigenschaften zu sehen, sondern in von vornherein situationsbezogen angelegten Persönlichkeitskonstrukten, die probabilistische Zusammenhänge zwischen situativen Bedingungen und Verhaltensweisen zum Ausdruck bringen. Zu diesen „strukturellen“ Ursachen muß das Vorliegen einer bestimmten Situation, die zu den Manifestationsbedingungen der Disposition gehört, als notwendiger Bestandteil dieser Art der Verhaltensklärung hinzukommen.
- Verhalten wird i.d.R. als ein bestimmtes Verhalten einer Person in Reaktion auf einen Reiz in einer Situation konzipiert und durch ein Maß ausgedrückt. Das grundlegende Erklärungsmodell für Verhalten läuft nicht darauf hinaus, dass Verhalten eine Funktion von Persönlichkeitseigenschaften ist, sondern eine Funktion der Interaktion von Persönlichkeit und Situation.

3. Rolle der Geschichte.

- Die Vergangenheit sollte in der Verhaltensdiagnostik relativ unbedeutend sein, während sie im traditionellen Ansatz als entscheidend betrachtet würde.
- Einwand: Auch im traditionellen Ansatz wird oft lediglich ein gegenwärtiger Status beschrieben, während in der Verhaltensdiagnostik Verhalten in bestimmten Situationen als Resultat einer Interaktionsgeschichte zwischen Person und Umwelt angesehen wird.
- Tatsächlich werden gerade in behavioralen Theorien verhaltensdeterminierende Bedingungen in der genetischen Ausstattung, der Verhaltensgeschichte und der aktuellen Situation gesehen. Wenn es darum geht, konkretes Verhalten einer Person zu erklären, ist nicht nur auf aktuelle Bedingungen (S und C) zu rekurrieren, sondern auch auf die funktionalen Qualitäten dieser Bedingungen, die in der Regel im Zuge von Lernprozessen entstanden sind und deshalb ihrerseits nur unter Bezugnahme auf die Verhaltensgeschichte der betreffenden Person erklärt werden können.

4. Erfassungsmethoden.

- Bevorzugte Erfassungsmethoden des behavioralen Ansatzes seien direkte Methoden (z.B. Verhaltensbeobachtung in natürlicher Umgebung), während im sogenannten traditionellen Ansatz indirekte Methoden (z.B. Interviews und Selbstbericht) im Vordergrund stünden.
- Einwand: Nach dem Dictionary of Behavioral Assessment Techniques: 53,47% Selbstberichte, 26,04% Beobachtungen; Behavior Therapy und Behaviour Research and Therapy 76% bzw. 53% Verwendung standardisierter Trait-Fragebögen in den Beiträgen.
- Die kontrastierende Bezeichnung der Methoden als direkt/indirekt unterstellt, dass im sogenannten traditionellen Ansatz auch nur Verhalten erfasst werden soll oder kann, dies aber nicht durch Beobachtung (direkt), sondern durch Befragung (indirekt) geschieht. Daß sich ein Selbstbericht statt auf Verhalten auf Gedanken, Gefühle, Stimmungen, Befindenszustände etc. beziehen kann und diese so direkt erfasst wie möglich, wird übersehen. Außerdem wird die Direktheit der Verhaltensbeobachtung überschätzt • sie sind auch abhängig von der Konstruktion des Beobachters

8. Wie lässt sich das Ergebnis einer Verhaltensbeobachtung von Personen genauer charakterisieren?

- Nach WESTMEYER lässt sich das Resultat einer Verhaltensbeobachtung in Form einer Beobachtungsstruktur (BS) wiedergeben.
- ◆ z ist eine **Beobachtungsstruktur** genau dann, wenn es S, CS, O, T • \cdot_t , OF gibt, so dass
 - $z = \langle S, CS, O, T, \cdot_t, OF \rangle$
 - **S** eine nicht-leere Menge von Personen ist
 - **CS** eine nicht-leere Menge von Verhaltensdeskriptoren/-kategorien
 - **O** eine Menge mit einem einzigen Element: Beobachter
 - **T** eine nicht leere endliche Menge von Zeitpunkten ist, geordnet durch die zweistellige Relation \cdot_t
 - **OF**: $O \times S \times T \cdot \text{Pot}(CS)$ [eine dreistellige Funktion]
- Durch die Beobachtungsfunktion OF wird der Beobachtungsvorgang verbegrifflicht, durch den

ein Beobachter O einer Person aus S zu einem Zeitpunkt T eine Teilmenge von Verhaltensdeskriptoren/-kategorien aus CS zuordnet.

- Wichtig ist, dass der Beobachter nicht das Verhalten einer Menge von Deskriptoren zuordnet, sondern der Person zu einem Zeitpunkt.
 - »Es fällt auf, daß das Verhalten, da den Gegensatz der Beobachtung darstellt, in der Beobachtungsstruktur selbst gar nicht vorkommt. Der Beobachter ordnet also, genau genommen, nicht dem Verhalten, über das in unbeschriebener Form gar nicht gesprochen werden kann, eine Menge von Deskriptoren zu, sondern der Person zu einem Zeitpunkt.« (WESTMEYER 1994, 282).
- Bei 2 Beobachter = immer 2 Beobachtungsstrukturen.
- Zur **Potenzmenge** gehören immer alle Teilmengen, die aus CS gebildet werden können

$$OF(O, S_1, t_1) = \{c_1, \dots, c_n\}$$

↑
Beobachter

↑
Pers.

↑
Zeitpkt. d. Beob.

↑
Stirn i. Falten

↑
Augen n. oben

↑
Mund n. unten

◆ Kodierungsformen.

- **Isomorphe Deskription:** vollständige Erfassung (z.B. Video)
- **Reduktive Deskription:** Zeichen- oder Kategoriensysteme
 - Zeichensysteme: *nur ausgewählte Verhaltensweisen werden registriert*, d.h. immer nur, wenn sie auftreten, z.B. mit Strichliste, Zeitintervall (von 10 Sek.) als Beispiel.
 - Kategoriensystem: *der gesamte Verhaltenstrom soll vollständig kodiert werden.* Die Kategorien müssen erschöpfend sein, denn für jede möglicherweise auftretende Verhaltensweise wird eine Kategorie gebraucht. Wenn für ein bestimmtes Verhalten noch keine Kategorie existiert: Restkategorie.

9. In welche theoretische Rahmenkonzeption lässt sich die Verhaltensdiagnostik relativ gut einordnen?

- In der Verhaltensdiagnostik dominiert gegenwärtig eine kognitive Perspektive (Cone 1986), gekennzeichnet durch die Verwendung von Trait-Begriffen, statistischen Normen und psychometrischen Kriterien. Daneben existiert in weit geringerem Maß die sogenannte radikal-behaviorale Perspektive, in der es um idiographische Analysen von Verhalten mit konsequent naturwissenschaftlicher Orientierung geht. Frage: Wie lassen sich beide Perspektiven innerhalb einer integrativen theoretischen Rahmenkonzeption miteinander verbinden?
- Die Verhaltensdiagnostik lässt sich in die theoretische Rahmenkonzeption von **MAGNUSSON** (1980) einordnen. Es handelt sich hierbei um eine sehr **integrative Rahmenkonzeption**, die **problem- und gegenstandsbezogen erweitert werden kann** und Raum lässt für die Verfolgung einer kognitiven oder einer radikal-behavioralen Perspektive, aber auch für eine Verbindung beider Sichtweisen.
- ◆ Das **interaktionistische Verhaltenskonzept** von MAGNUSSON expliziert vier Annahmen:
 - (1) Für jede Person gibt es eine bestimmte Population möglicher Verhaltensweisen. Diese Populationen sind bis zu einem gewissen Grad personenspezifisch.
 - (2) Die Wahrscheinlichkeit für ein bestimmtes Verhalten ist eine bedingte und determiniert durch den situativen Kontext, d.h.
 - a) durch den situationalen Bezugsrahmen (die intersituationalen Faktoren) und
 - Die intersituationalen Faktoren charakterisieren den Bezugsrahmen, in dem die intrasituationale Interaktion zwischen Zielperson s_1 und Interaktionspartner s_2 stattfindet und der sich während dieser Interaktion nicht ändert. Setzt sich zusammen aus Personvariablen von s_1 und s_2 und der Menge der anderen, den situativen Kontext kennzeichnenden Variablen: u.a. häusliche, schulische oder Arbeitsumgebung.
 - b) durch die sich fortlaufend ändernden situativen Bedingungen, die der Ursprung jeder spezifischen Verhaltensweise sind (die intrasituationalen Faktoren)
 - (3) Mit einer bestimmten Situation, in der sich eine Person befindet, ist zweierlei vorgegeben:
 - a) eine umgrenzte Stichprobe möglichen und wahrscheinlichen Verhaltens, die situationspezifisch ist, und
 - b) – bis zu einem gewissen Grad – die Wahrscheinlichkeit für einzelne Verhaltensweisen.
 - (4) Innerhalb des durch eine bestimmte Situation vorgegebenen Bezugsrahmen ändern sich die Wahrscheinlichkeiten für bestimmte Verhaltensweisen in Abhängigkeit von Änderungen der intrasituationalen Faktoren (= physikalische Bedingungen, Verhalten anderer Personen, eigenes Verhalten der Person oder eine Kombination mehrerer dieser Faktoren).

10. Welches grundlegende Problem ergibt sich bei der Anwendung der Wissensgrundlagen des psychometrischen Ansatzes auf den Einzelfall?

- Der psychometrische Ansatz bedient sich einer normativ-nomothetischen Forschungsstrategie. **Einzelne Testergebnisse werden dabei anhand einer Stichprobe bezüglich Erwartungswert, Konfidenzintervall, Fehlervarianz usw. relativiert.**
- Bei der Anwendung auf den Einzelfall müsste eine individuelle Fehlervarianz ermittelt werden, denn die Annahme der Äquivalenz der Fehlervarianzen ist nicht empirisch prüfbar. Der Test müsste also bei einer Person x-mal angewendet werden, um den Testwert im Sinne der KTT interpretieren zu können. Da dies aber ohne Veränderung der Person (und damit der Messwerte) nicht möglich ist, wird auf Verteilungscharakteristika von Gruppen zurückgegriffen, die aber für den Einzelfall nicht gelten müssen.
- Außerdem gibt es das Problem der Zuordnung zu Gruppennormen.
- [vgl.: Diagnostischer Prozess, Fragebogen 1, Frage 6]

11. Wie ist bei der psychologischen Begutachtung vorzugehen? Wie lässt sich das Gutachten als Produkt sinnvoll gliedern?

- i. **Auswertung von Unterlagen und Akten.**
 - **Verhältnis Versuchsleiter – Versuchsperson** sollte neutral, freundlich, wohlwollend sein
 - ii. **Exploration bzw. Anamnese** (abhängig von der Fragestellung; frei, strukturiert, Möglichkeit zur Hypothesenbildung)
 - iii. **Explizite Formulierung von Hypothesen.**
 - iv. **Datenerhebung.** Grundregeln:
 1. Es sollten nur Tests eingesetzt werden, die hinsichtlich der Fragestellung relevant sind.
 2. Es sollten so wenig Tests wie möglich eingesetzt werden (Belastung).
 3. Je zentraler eine Hypothese ist, desto mehr Tests sind zu ihrer Absicherung heranzuziehen. *Niemals nur auf einem Bein stehen – mindestens auf zwei Beinen!*
 4. Die Reihenfolge der Tests richtet sich nach Abwechslung und Motivierung.
 5. Vermeidung belastender, unüblicher Testzeiten.
 6. Geeigneter Raum (Ruhe, Licht).
 7. Pausen.
 - v. **Auswertung** (schnell; Angabe von Standardwerten, Prozentrang, Konfidenzintervalle, kritische Differenzen, Profilvergleiche)
 - vi. **Zusätzliche Datenerhebung bei Widersprüchen oder neuen Hypothesen.**
 - vii. **Interpretation** (durch Daten belegt; verständlich, nachvollziehbar, widerlegbar).
 - viii. **Entscheidung** (Psychologe als Berater)
- ◆ Eine **sinnvolle Gliederung** des Gutachtens könnte wie folgt aussehen:
1. **Übersicht** (Gutachter, Problemstellung, Verfahren, Datenquellen, Adressaten)
 2. **Vorgeschichte** (für Fragestellung relevante Informationen; Quellen kenntlich machen!)
 3. **Untersuchungsbericht** (erhobene Daten, gesonderte Berichte zu den eingesetzten Verfahren [u.U. mit Testbeschreibung], Verhaltensbeobachtung bzw. -beschreibung, Ergebnisbericht [wichtigste Ergebnisse in Scores], Interpretation)
 4. **Befund** (Interpretation, neutrale Darstellung ohne Diagnose/ Prognose)
 - Informationen aus Vorgeschichte und Untersuchungsbericht werden thematisch zusammengekommen (*Befundliste*).
 - Integration von Einzelheiten zu übergreifenden, einheitlichen Aussagen • (Beschreibung von relativ invariantem Verhalten [im Gegensatz zum Untersuchungsbericht]): *Befundskizze*.
 - Feststellungen, keine diagnostischen oder prognostischen Erklärungen (nicht erklären, sondern Befundlage darstellen).
 5. **Stellungnahme** (Diagnose/ Prognose, transparente Argumentation)

- Das **psychodiagnostische Gutachten** – Definition (JÄGER & PETERMANN).
 - »Psychodiagnostische Gutachten für (meist fachfremde) Dritte sind eine selbständige (in sich geschlossene) zusammenfassende Darstellung der psychodiagnostischen Vorgehensweise, der Befunde und Schlussfolgerungen in bezug auf eine hinsichtlich einer konkreten Fragestellung zu begutachtende Person, Institution oder Situation, basierend auf einem der Fragestellung gemäßen, angemessenen komplexen diagnostischen Prozess für einen Gutachtenempfänger (Auftraggeber). Mit Hilfe des Gutachtens soll sein Empfänger Entscheidungen in seinem System (seinem diagnostischen Prozess) fundierter treffen können.« Ein Gutachten ist eine wissenschaftliche Leistung, bei der mit Hilfe von anerkannten wissenschaftlichen Theorien und Kriterien Aussagen zu konkreten Fragestellungen gemacht werden.

12. Welchen Beurteilungsfehlern und Störvariablen ist der psychologische Gutachter häufig ausgesetzt? Nennen und erläutern Sie einige typische Beispiele.

- ◆ **Beurteilungsfehler** können sich ergeben aus
 - **Informationsverarbeitung** (primacy/ recency Effekte, bessere Nutzung konsistenter Infos, Überbewertung negativer Infos, positiver bias, Informationsmenge usw.).
 - **Variablen des Diagnostikers** (Persönlichkeit, implizite Persönlichkeitstheorien, Übertragung/ Gegenübertragung, diagnostische und therapeutische Ausrichtung usw.)
 - **Variablen des Klienten** (Persönlichkeit, Rasse, Art der Symptome, Infovermittlung bei Anamnese, sozialer Status, Bedeutung der Begutachtungssituation für den Klienten)
 - **Interaktion Diagnostiker-Klient** (Reaktionen in der Anamnese)
- ◆ **Störvariablen** ergeben sich aus
 - **Auftraggeber** (schildert das Problem subjektiv/ selektiv)
 - **Diagnostiker** (verarbeitet die Infos abhängig von theoretischer Position, Erfahrung etc.)
 - **Testsituation** (Verlauf, Atmosphäre)
 - **Klienten** (Unterdrückung von Infos, mangelnde Ausdrucksfähigkeit, Verzerrung von Infos)

13. Welche Richtlinien (Normen, Guidelines) existieren, die die Praxis psychologischer Diagnostik regulieren?

- ◆ **ITC (International Test Commission)** erarbeitet Richtlinien für die fachgerechte Testanwendung:
 - Zur Gewährleistung der Übereinstimmung in der Qualität von Tests, die zum Einsatz in unterschiedlichen Kultur- und Sprachbereiche adaptiert werden.
 - Erarbeitung von Richtlinien für eine faire und ethisch korrekte Anwendung von Tests, aus denen sich Standards für die Ausbildung und spezifische Kompetenzanforderungen an die Testanwender (Fachwissen, Fertigkeiten, Fähigkeiten und andere persönliche Merkmale) ableiten lassen.
 - Es wird nicht beabsichtigt, neue Richtlinien zu „erfinden“, sondern die übereinstimmenden Stränge der Arbeiten von Experten verschiedener Nationen zusammenzufassen, die bereits existierenden Richtlinien, Durchführungsvorschriften, Standards und andere Dokumente durchziehen, und eine zusammenhängende Struktur zu schaffen, innerhalb derer sie nachvollzogen und angewendet werden können.
 - Richtlinien als Orientierungspunkt, anhand dessen vorliegende lokale Standards im Hinblick auf Vollständigkeit und internationale Vergleichbarkeit beurteilt werden können
- ◆ Die **European Association of Psychological Assessment (EAPA)** sponsert die Entwicklung von Richtlinien für den Assessment Process (Beurteilungsprozess, etc.) / Guidelines for the Assessment Process (GAP).
- ◆ **Deutschland: DIN 33430 Projekt.** 1995 initiierte der BDP das DIN 33430 Projekt, indem es beim Deutschen Institut für Normung (DIN) eine Standardisierung psychologischer Tests beantragte. (Sollte seit Sommer 2001 existieren).
- ◆ **verschiedene (nationale) europäische Initiativen** (Großbritannien, Frankreich, Schweden, Holland etc.). Die jeweilige hinter den einzelnen Initiative stehende Organisation kann dabei recht unterschiedlich sein.