

## 3.2 Überblick über die Probabilistische Testtheorie

<b>Oberbegriff</b>	Umfasst verschiedene Modelle
<b>Grundannahmen</b>	<ul style="list-style-type: none"> <li>- In der Testsituation kann nur ein Verhaltensausschnitt erfasst werden, welcher auf eine Verhaltensdisposition zurückgeht</li> <li>- Korrelation der manifesten Variable soll auf eine einzige latente Variable zurückgeführt werden</li> <li>- Lokale stochastische Unabhängigkeit &amp; probabilistischer Zusammenhang</li> </ul>

<b>Verhaltensausschnitt</b>	Manifeste Variable
<b>Verhaltensdisposition</b>	Latente Variable

### 3.2.1 Grundlagen des Rasch-Modells

Lösungswahrscheinlichkeit  $p$  ergibt sich aus

<b>Personenparameter <math>\Theta</math></b>	Fähigkeit/ Eigenschaftsausprägung
<b>Itemparameter <math>\sigma</math></b>	Schwierigkeit des Items
$\Theta = \sigma$	$p = 0,5$
$\Theta > \sigma$	$p > 0,5$
$\Theta < \sigma$	$p < 0,5$
<b>Erschöpfende Statistik der Personenfähigkeit</b>	<ul style="list-style-type: none"> <li>- Der Summenwert einer Person liefert alle Informationen über die Fähigkeitsausprägung einer Person</li> <li>- Wird das Rasch-Modell nicht abgelehnt, sagt der Summenwert der Itemantworten auch etwas über den Ausprägungsgrad aus</li> </ul>

*Aspekte des Modells*

<b>Erschöpfende Statistik der Personenfähigkeit</b>	- Der Summenwert einer Person liefert alle Informationen über die Fähigkeitsausprägung einer Person
<b>Eindimensionalität</b>	Unter Geltung des Rasch-Modells ist der Test eindimensional. Das heißt ein Item ist ein guter Indikator für eine Fähigkeit, wenn die Antwort auf das Item auf eine Fähigkeitsausprägung zurückzuführen ist.
<b>Lokale Unabhängigkeit</b>	Eindimensionalität ist formal gegeben, sobald die Inter-Item-Korrelation eines Tests nach Auspartialisierung der latenten Eigenschaft verschwindet.

<b>Lokale stochastische Unabhängigkeit</b>	Ist gegeben, wenn für jede Person die Lösungswahrscheinlichkeiten zweier beliebiger Items multipliziert werden dürfen um die kombinierte Lösungswahrscheinlichkeit beider Items zu ermitteln.
<b>Spezifische Objektivität</b>	<ol style="list-style-type: none"> <li>1. Interpersonella Vergleiche sind invariant über spez. Items &amp; Maße die verwendet werden</li> <li>2. Inter-Item-Vergleich sind invariant über die spez. Personen an denen die Items kalibriert werden</li> </ol>
<b>Additivität</b>	- Parameter haben dasselbe Skalenniveau und dieselbe Einheit, sodass sie additiv/ substraktiv miteinander verknüpft werden können

#### Modellparameter

<b>Personenparameter</b>	<ul style="list-style-type: none"> <li>- Schätzung mittels conditional Maximum-Likelihood-Methode (cML)</li> <li>- <math>\theta</math> hat Differenzskalenniveau (logarithmierte Verhältnisskala): Logit-Einheit</li> <li>- Kenntnis der Itemparameter wird für cML vorausgesetzt</li> </ul>
<b>Logit-Einheit</b>	<ul style="list-style-type: none"> <li>- nicht-lineare Transformation der Rohwerte in eine abstrakte Einheit</li> <li>- Negative Werte = leichte Items/ geringe Fähigkeit</li> <li>- Positive Werte = schwere Items/ hohe Fähigkeit</li> </ul>

#### Item characteristic Curve ICC

<b>y - Achse</b>	Itemlösungswahrscheinlichkeit
<b>x - Achse</b>	Gemeinsame Einheit von Personen-/ Itemparameter
<b>Itemschwierigkeit</b>	Schnittpunkt einer x-Achsen-Parallelen mit Kurve

Durch Modelltests wird festgestellt ob Rasch-Modell abgelehnt werden muss.

*Grafischer Modelltest:* Streudiagramm in welchem die geschätzten Parameter aus 2 Stichproben dargestellt werden. X-/ y-Achse sind die Logit-Einheiten der Itemparameter. Ansteigende Linie zeigt Regressionsgerade der Parameter einer Teilstichprobe auf die andere (Idealfall: Regressionsgerade = Winkelhalbierende). Nachteil: kein Signifikanztest

*Andersen- Likelihood-Quotienten-Test:* einziger bewährter Signifikanztest

*Bootstrap-Methode:* simuliert für verschiedene Modelltests eine Verteilung der jeweiligen Prüfgröße unter Geltung des Rasch-Modells. Die Verwendung der asymptotischen Prüfgrößen der  $\chi^2$ -Verteilung ist an Voraussetzungen geknüpft die im Rasch-Modell selten erfüllt sind.

Annahmen, wenn Rasch-Modell nicht abgelehnt:

- Lösungswahrscheinlichkeit wird durch logistische Funktion beschrieben
- Summenwerte sind suffiziente/ erschöpfende Statistiken der Personenfähigkeit
- Vergleiche zwischen Items und Personen sind spezifisch objektiv
- Items sind (mit Ausnahmen) eindimensional
- Alle Items besitzen die gleiche Trennschärfe (nur bei dichotomen Modell nötig)

### 3.2.2 Weitere probabilistische Testmodelle für dichotome Daten

2-Parameter-Modell (Birnbaum-Modell) / Erweiterung des Raschmodells

<b>Personenparameter <math>\Theta</math></b>	Fähigkeit/ Eigenschaftsausprägung
<b>Itemparameter <math>\sigma</math></b>	Schwierigkeit des Items (Stelle auf der x-Achse an der $p = 0,5$ )
<b>Trennschärfeparameter <math>\beta_i</math></b>	<ul style="list-style-type: none"> <li>- Steigung (einer Tangente am Wendepunkt) der ICC</li> <li>- Je höher die Trennschärfe, desto steiler der Anstieg der ICC</li> <li>- Trennschärfen einzelner Items unterscheiden sich in diesem Modell</li> </ul>
Ungewichtete Summenwertbildugn	Nicht mehr zulässig, da wegen der unterschiedlichen Trennschärfen eine Gewichtung mit der Itemtrennschärfe vorgenommen werden müsste.

3- Parameter-Modell

<b>Personenparameter <math>\Theta</math></b>	Fähigkeit/ Eigenschaftsausprägung
<b>Itemparameter <math>\sigma</math></b>	Schwierigkeit des Items (Stelle auf der x-Achse an der $p = 0,5$ )
<b>Trennschärfeparameter <math>\beta_i</math></b>	<ul style="list-style-type: none"> <li>- Steigung (einer Tangente am Wendepunkt) der ICC</li> <li>- Je höher die Trennschärfe, desto steiler der Anstieg der ICC</li> <li>- Trennschärfen einzelner Items unterscheiden sich in diesem Modell</li> </ul>
<b>Ratewahrscheinlichkeit <math>\gamma_i</math></b>	Je nach Beschaffenheit der Distraktoren wird die Ratewahrscheinlichkeit für jedes Item und bei jeder Person in Abhängigkeit von der Personenfähigkeit anders ausfallen

### 3.2.2 Probabilistische Modelle für Ratingskalen

*Ordinales Rasch-Modell* für Items mit mehr als 2 geordneten Antwortalternativen  
 --> Ordinalskalenniveau!

<b>Thresholds</b>	Grenzen/ Schwellen/ Übergang an dem es wahrscheinlicher ist Kategorie 1 anstatt Kategorie 0 zu wählen
<b>Wahrscheinlichkeitsfunktion <math>p(0)</math></b>	Entspricht der Wahrscheinlichkeit Kategorie 0 zu wählen und somit der gespiegelten logistischen Funktion $p(1)$ zu wählen
<b>Schnittpunkt <math>p(0) \times p(1)</math></b>	Schwelle

Wird das Schwellenkonzept auf mehr als zwei geordnete Antwortkategorien übertragen, wird nicht mehr von ICCs gesprochen sondern von CCCs: Category characteristic Curies.

*Partial-Credit-Modell:* Damit soll die Wahrscheinlichkeit bestimmt werden, mit der eine Person eine bestimmte Antwortkategorie wählt.

Voraussetzung: Antwortschwellen müssen geordnet sein und dürfen sich nicht überschneiden.

*Mixed-Rasch-Modelle:* Modelle quantifizieren und klassifizieren zugleich. Sucht nach Personengruppen, die sich in ihrem Antwortmuster maximal unterscheiden. Innerhalb einer jeden Klasse gilt jedoch das Rasch-Modell. Innerhalb einer Klasse wird nur eine Fähigkeit gemessen. Mixed-Rasch-Modelle können Verletzungen von Eindimensionalitätsannahme für jedes Item anzeigen.