



**Universität Stuttgart**  
Zentrum für Lehre und  
Weiterbildung | zlw

**Klausurmodellierung  
in der Technischen  
Thermodynamik mittels  
Item Response Theory**  
*Ergebnisse einer  
hochschuldidaktischen  
Begleitforschung*

Thorsten Braun  
Basel, 17. Feb 2018

## Ziele des Vortrags



Nachvollziehen der wichtigsten Schritte und Ergebnisse einer Kompetenzmodellierung anhand einer realen Klausur in der Technischen Thermodynamik.



Einen grundlegenden Übertrag des Vorgehens in andere, vergleichbare Forschungskontexte ermöglichen.



Die Modellierung von studentischen Kompetenzen anhand bestehender Prüfungen anregen.

## Vorbemerkungen



Diese Präsentation samt Literaturliste finden Sie unter:  
<https://www.uni-stuttgart.de/hd/ueber/mitarbeiter/Braun-00007/>



Die vorgestellten Ergebnisse sind Teil einer umfangreicheren  
Modulentwicklung, die hier nicht dargestellt werden kann.  
Siehe dazu Braun 2015.



Alle Berechnungen und Abbildungen wurden mit R  
durchgeführt und erstellt (`{tam}`, `{sirt}` und `{ggplot2}`)

- Literaturhinweise zum Rahmenprojekt und zur verwendeten Statistik-Software:
  - BRAUN, Thorsten: Die Lehrsituation adäquat beforschen : Methodenintegrative Anregungen für hochschuldidaktische Begleitforschung. In: KLEUSBERG, Alfred; KRÖBER, Edith; NICKOLAUS, Reinhold; WEIDL, Timo (Hrsg.): Ausgewählte Ergebnisse zur Qualitätsoptimierung der Lehre an der Universität Stuttgart. Stuttgart, 2015, S. 152–166
  - ROBITZSCH, Alexander: Supplementary Item Response Theory Models ('sirt'). R Package. URL <https://cran.r-project.org/web/packages/sirt/sirt.pdf>. – Aktualisierungsdatum: 2016-06-19 – Überprüfungsdatum 2017-01-15
  - ROBITZSCH, Alexander: Test Analysis Modules ('TAM'). R Package. URL <https://cran.r-project.org/web/packages/TAM/TAM.pdf>. – Aktualisierungsdatum: 2016-05-31 – Überprüfungsdatum 2017-01-14

## Gliederung

- Beschreibung der Klausur und Stichprobe
- Bedeutung des Codierverfahrens für die Kompetenzmodellierung
- Modellvergleiche (1PL, 2PL, BF)
- Testletstruktur und essentielle Eindimensionalität
- Erkenntnisse aus der bifaktoriellen Struktur
- Herausfordernde Aufgaben
- Zusammenfassung und Konsequenzen

# **Beschreibung der Klausur und Stichprobe**

## Beschreibung der Klausur und Stichprobe

### Testlet 1: gemischte Aufgaben

- a) Isobare Wärmekapazität
- b) p-v-Zustandsdiagramm
- c) Wärmekapazität ideales Gas
- d) T-s-Diagramm ideales Gas
- e) Leistungsziffer Wärmepumpe
- f) Thermodyn. Gleichgewicht
- g) Isotherme Kompressibilität

### Testlet 2: ideales Gas/Entropie

- a) molare Zusammensetzung
- b) Entropieproduktion

**Klausurdauer:** 3 Stunden

**Klausurinhalte:** 2 Semester

### Testlet 3: Kreisprozess

- a) p-v-Diagramm
- b) Wärmestrom
- c) Wärmestrom
- d) Drücke
- e) Massenstrom
- f) Technische Leistung
- g) Dampfgehalt

### Testlet 4: Feuchte Luft

- a) Wasserbeladung
- b) Volumenstrom
- c) Enthalpie
- d) Massenstrom
- e) Druck

### Testlet 5: Kaltluftprozess

- a) Leistungsziffer

## Beschreibung der Klausur und Stichprobe

### Beispielaufgabe 2: Ideales Gas / Entropie

In einem stationär durchströmten Prozess werden zwei Stoffströme unterschiedlicher reiner Gase adiabat miteinander vermischt. Die Zustände vor dem Mischen sind:

- Gasstrom A:  $T_{A1} = 400\text{ K}$ ,  $p_{A1} = 1\text{ bar}$
- Gasstrom B:  $T_{B1} = 300\text{ K}$ ,  $p_{B1} = 10\text{ bar}$

Am Ausgang der Mischkammer stellt sich die Temperatur  $T_2 = 351,7\text{ K}$  und der Druck  $p_2 = 1\text{ bar}$  ein. Die Gase können als perfekte Gase mit folgenden molaren isobaren Wärmekapazitäten behandelt werden:  $c_{p,A} = 20,79 \frac{\text{J}}{\text{mol K}}$ ,  $c_{p,B} = 29,10 \frac{\text{J}}{\text{mol K}}$ .

Skizze:



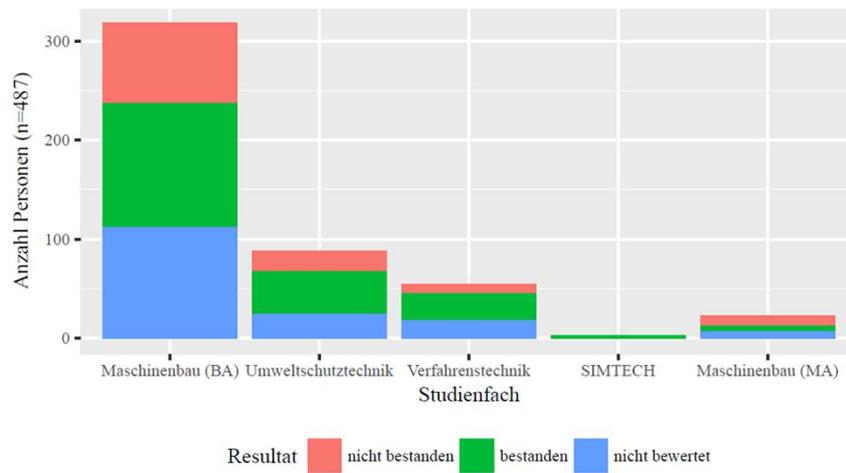
- Welche molare Zusammensetzung ( $x_A$ ,  $x_B$ ) hat das Gemisch? (6 Punkte)
- Berechnen Sie die molare Entropieproduktion des Vorgangs  $\sigma_{12}^{spez} = \frac{\dot{\sigma}_{12}}{n_A + n_B}$ . (5 Punkte)

*Hinweis:*

- Änderungen der kinetischen und potentiellen Energie sollen vernachlässigt werden.

## Beschreibung der Klausur und Stichprobe

### Modulergebnis nach Studienfächer

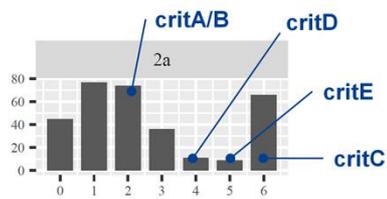


# **Bedeutung des Codierverfahrens**

## Bedeutung des Codierverfahrens

### Fünf Codierstrategien bzw. Codierkriterien

- critA: ab Median, inklusive null Punkte
- critB: ab Median, exklusive null Punkte
- critC: nur volle Punktzahl
- critD: charakteristische Modalitäten
- critE: inhaltsanalytische Erwägungen (qualitativ) auf Grundlage der Musterlösung



## Bedeutung des Codierverfahrens

### Gütekriterien der fünf Codierstrategien

Codierkriterium	AICc	AIC3	WLE-Reliabilität der Sigma-Parameter
critA	7960	7941	0,72
critB	8005	7986	0,76
critC	6708	6689	0,74
critD	7830	7811	0,76
<b>critE</b>	<b>6509</b>	<b>6491</b>	<b>0,77</b>

- **critE zeigt sich als die beste Codierstrategie**, bei schlechtestem Personenfit. Allerdings verletzen nur 13 Personen den Personenfit signifikant.

- Literaturhinweise zu Datenrecodierung, Informationskriterien und WLE-Reliabilität:
  - BÜHNER, Markus: Einführung in die Test- und Fragebogenkonstruktion. 3. Auflage : Pearson Studium, 2011, S. 41ff.
  - BÜHNER, Markus ; KÜCHENHOFF, Helmut; GÄRTNER, Manuela (Mitarb.); HEINE, Jörg-Henrik (Mitarb.); HOFER, Sarah (Mitarb.) : Das Rasch-Modell : Modellprüfung und Informationskriterien. Skript. München, 2008. – Skript, S. 4
  - ROBITZSCH, Alexander: Test Analysis Modules ('TAM'). R Package. URL <https://cran.r-project.org/web/packages/TAM/TAM.pdf>. – Aktualisierungsdatum: 2016-05-31 – Überprüfungsdatum 2017-01-14, S. 168ff.
  - ROST, Jürgen: Lehrbuch Testtheorie - Testkonstruktion. 2., vollst. überarb. u. erw. Aufl. Bern, Göttingen : Huber, 2004 (Statistik und Methodenlehre bei Hans Huber), S. 89f., 313f., 339ff.
  - WARM, Thomas A.: Weighted likelihood estimation of ability in item response theory. In: Psychometrika 54 (1989), Nr. 3, S. 427–450. URL <http://link.springer.com/article/10.1007/BF02294627>, S. 429f.

## Bedeutung des Codierverfahrens

### Einfluss der Codierstrategie

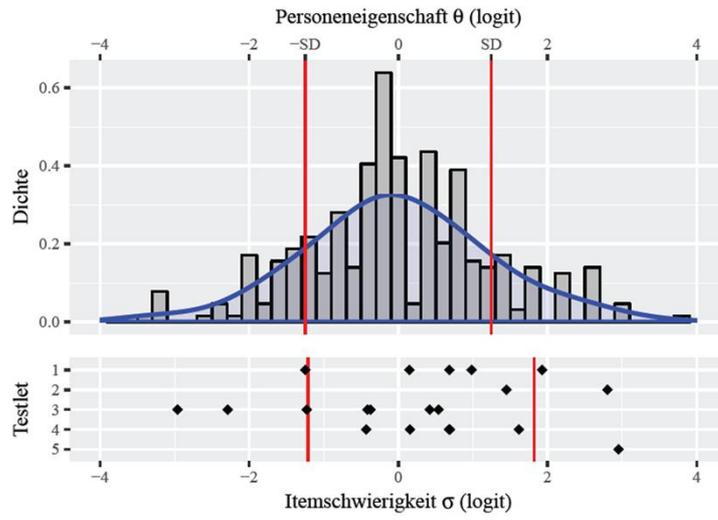
- Codierstrategie beeinflusst die relative Testschwere sowie Boden- oder Deckeneffekte.
- Toleranz oder Strenge der Codierung beeinflusst die Spannweite der Sigma-Parameter.
- Zusammenhang mit Item- und Personenfit: liberale Codierung bedroht Modellgüte, strenge Codierung bedroht Personenfit.
- Je strenger codiert wird, desto mehr wird die Personenhomogenität diszipliniert (im grafischen Modellgeltungstest).
- „Tolerantes Maximum“ als taugliche Faustregel zur Codierung.

- Literaturhinweise zum Guttman-Fehler, grafischen Modellgeltungstest, Item- und Personenfit:
  - BOONE, William J. ; STAVER, John R. ; YALE, Melissa S.: Rasch analysis in the human sciences. Dordrecht : Springer, 2014, S. 163ff., 172f.
  - BÜHNER, Markus: Einführung in die Test- und Fragebogenkonstruktion. 3. Auflage : Pearson Studium, 2011, S. 547
  - BÜHNER, Markus ; KÜCHENHOFF, Helmut; GÄRTNER, Manuela (Mitarb.); HEINE, Jörg-Henrik (Mitarb.); HOFER, Sarah (Mitarb.) : Das Rasch-Modell : Modellprüfung und Informationskriterien. Skript. München, 2008. – Skript, S. 5
  - DIEKMANN, Andreas: Empirische Sozialforschung : Grundlagen Methoden Anwendungen. 7. Aufl. Reinbek bei Hamburg : Rowohlt, 2001 (Rowohlt's Enzyklopädie 55551), S. 237f.
  - MACCALLUM, Robert C. ; BROWNE, Michael W. ; SUGAWARA, Hazuki M.: Power Analysis and Determination of Sample Size for Covariance Structure Modeling. In: Psychological Methods 1 (1996), Nr. 2, S. 130–149. URL <http://www.statpower.net/Content/312/Handout/MacCallumBrowneSugawara96.pdf>
  - MEIJER, Rob R.: The Number of Guttman Errors as a Simple and Powerful Person-Fit Statistic. In: Applied Psychological Measurement 18 (1994), Nr. 4, S. 311–314
  - ROBITZSCH, Alexander: Supplementary Item Response Theory Models ('sirt'). R Package. URL <https://cran.r-project.org/web/packages/sirt/sirt.pdf>. – Aktualisierungsdatum: 2016-06-19 – Überprüfungsdatum 2017-01-15, S. 284f., 278ff.
  - ROBITZSCH, Alexander: Test Analysis Modules ('TAM'). R Package. URL <https://cran.r-project.org/web/packages/TAM/TAM.pdf>. – Aktualisierungsdatum: 2016-05-31 – Überprüfungsdatum 2017-01-14, S. 46f., 71f., 84f.
  - ROST, Jürgen: Lehrbuch Testtheorie - Testkonstruktion. 2., vollst. überarb. u. erw. Aufl. Bern, Göttingen : Huber, 2004 (Statistik und Methodenlehre bei Hans Huber), S. 124
  - TENDEIRO, Jorge N.: Person Fit ('PerFit'). R Package. URL <https://cran.r-project.org/web/packages/PerFit/PerFit.pdf>. – Aktualisierungsdatum: 2016-10-18 – Überprüfungsdatum 2017-01-14, S. 13f.
  - WRIGHT, B. D. ; LINACRE, J. M.: Reasonable mean-square fit values. In: Rasch Measurement Transactions (1994), 8:3, S. 370. URL <http://www.rasch.org/rmt/rmt83b.htm>
  - WRIGHT, B. D. ; MASTERS, G. N.: Computation of OUTFIT and INFIT Statistics. In: Rasch Measurement Transactions (1990), 3:4, S. 84–85. URL <http://www.rasch.org/rmt/rmt34e.htm> – Überprüfungsdatum 2017-01-15
- **Guttman-Fehler** erweist sich hier als ein sehr gutes Maß für den Personenfit. Zum Guttman-Error vgl. Diekmann 2001, S. 237 f.; Rost 2004, S. 124.; Tendeiro 2016-10-17, S. 13 f.; Meijer 1994.
- Zum **grafischen Modellgeltungstest** vgl. Bühner und Küchenhoff 2008, S. 5.; Bühner 2011, S. 547.
- Zu **Itemfit** vgl. Boone, Staver und Yale 2014, S. 164 ff.; Berechnungsgrundlagen bei Wright und Masters 1990. – R: `tam.fit{tam}`; Robitzsch 2016, S. 84 f.; `msq.itemfit{tam}`; ebd., S. 71 f.; `pcm.itemfit{sirt}`; Robitzsch 2016-07-19, S. 284 f.; `RMSEA IRT.itemfit{tam}`, vgl. Robitzsch 2016, S. 46 f.
- Zum **Personen-Fit** vgl. `pcm.fit{sirt}`; Robitzsch 2016-07-19, S. 278 ff.; Boone, Staver und Yale 2014, S. 163, 172f.

# **Modellvergleiche (1PL, 2PL, BF)**

## Modellvergleiche

### 1PL Wright-Map



## Modellvergleiche

Vergleich nach Gütekriterien und Informationskriterien

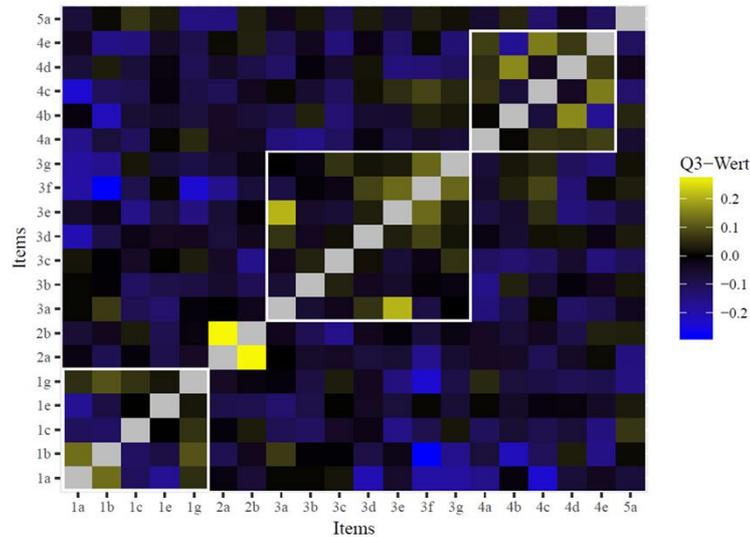
Modell	SRMR	AIC	AICc	EAP-Reliabilität	WLE-Reliabilität
1PL	0.0567	6488	6509	0.7824	0.7656
2PL	0.0484	6474	6486	0.8019	0.7848
BF/1PL	0.0520	6432	6437	0.6922	na

- Die Modelle kommen nicht zu signifikant verschiedenen Item- oder Personenparametern (t-Test und Wilcoxon Signed Rank Test).
- Das 2PL-Modell entwickelt die Modellierung nicht grundlegend weiter. Interpretationsprobleme überwiegen die Vorteile.
- Die Gültigkeit des bifaktoriellen Modells konnte insb. durch Test auf Invarianz der Itemparameter nachgewiesen werden. Problematisch ist jedoch die verschlechterte Reliabilität und der Verzicht auf das überlegene WLE-Verfahren zur Parameterschätzung.

# Testletstruktur und essentielle Eindimensionalität

## Testletstruktur und essentielle Eindimensionalität

### Explorative Q3-Matrix



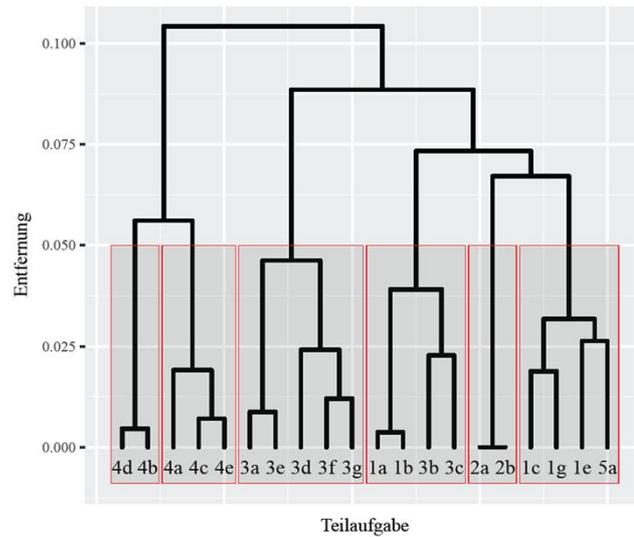
Universität Stuttgart

17

- Literaturhinweise zu Testlets und Q3-Matrix
  - CHEN, Wen-Hung ; THISSEN, David: Local Dependence Indexes for Item Pairs Using Item Response Theory. In: Journal of Educational and Behavioral Statistics 22 (1997), Nr. 3, S. 265–289
  - WAINER, Howard ; BRADLOW, Eric T. ; WANG, Xiaohui: Testlet response theory and its applications. Cambridge : Cambridge University Press, 2007, S. 44
  - WANG, Wen-Chung ; WILSON, Mark: The Rasch Testlet Model. In: Applied Psychological Measurement 29 (2005), Nr. 2, S. 126–149
- Zu **Testlets im Allgemeinen**: Wainer, Bradlow und X. Wang 2007, S. 44.
- Zur **Q3-Matrix**: tam.modelfit{tam} vgl. Robitzsch 2016, S. 153.; Chen und Thissen 1997, S. 267 f.

## Testletstruktur und essentielle Eindimensionalität

### Explorative DETECT-Analyse



Universität Stuttgart

18

- Literaturhinweis:
  - ZHANG, Jinming ; STOUT, William: The Theoretical Detect Index of Dimensionality and its Application to Approximate Simple Structure. In: Psychometrika 64 (1999), Nr. 2, S. 213–249

## Testletstruktur und essentielle Eindimensionalität

### Indizes zur konfirmatorischen DETECT-Analyse

Tabelle 7.12: Konfirmatorische DETECT-Analyse

Modell	DETECT	ASSI	RATIO
1A	0.8450	0.4459	0.6404
1B	0.8173	0.5325	0.6072
1C	0.6404	0.4242	0.6981
1D	0.7914	0.5065	0.6064
1E	0.7483	0.4842	0.7022

- Literaturhinweis:
  - ZHANG, Jinming ; STOUT, William: The Theoretical Detect Index of Dimensionality and its Application to Approximate Simple Structure. In: Psychometrika 64 (1999), Nr. 2, S. 213–249
- Zur **Interpretation der Indizes**: Ein DETECT-Index von  $> 0.40$  gilt als moderate, ein Wert von  $> 1.00$  als starke Mehrdimensionalität. Die Parameter ASSI und RATIO können maximale Werte von 1 annehmen. In diesem Fall würden sie eine maximale Abweichung von der Eindimensionalitätsannahme anzeigen. Werte größer 0.25 (ASSI) bzw. 0.36 (RATIO) werden als bedeutsame Abweichung interpretiert. Insofern ist festzustellen, dass die qualitative Testletstruktur analog zu den Erkenntnissen der Q3-Matrix auch als Clusterpartition mit mittelstarker Mehrdimensionalität im DETECT-Index zutage tritt.

## Testletstruktur und essentielle Eindimensionalität

### Statistiken zur essentiellen Eindimensionalität

- Statistiken zur essentiellen Eindimensionalität zeigen die Dominanz des Hauptfaktors auf, bei gleichzeitigem Nachweis bifaktorieller Datenstruktur.
- Auch wenn die Testlets die Klausur markant strukturieren und lokale Abhängigkeiten einführen, **handelt es sich bei der modellierten Kompetenz im Wesentlichen um eine globale und eindimensionale Größe.**

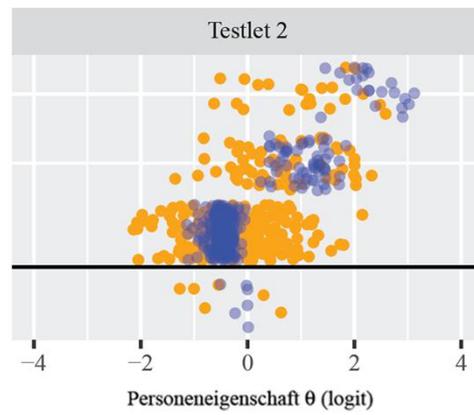
Koeffizient	Wert
Coefficient $\alpha$	0.79
Coefficient $\omega$	0.91
Coefficient $\omega_H$	0.79
Explained Common Variance ECV	0.87
Percent Uncontaminated Correlations PUC	0.78

- Literaturhinweise:
  - REISE, Steven P.: The Rediscovery of Bifactor Measurement Models. In: Multivariate Behavioral Research 47 (2012), Nr. 5, S. 667–696
  - REISE, Steven P. ; RODRIGUEZ, Anthony ; HAVILAND, Mark G.: Evaluating Bifactor Models : Calculating and Interpreting Statistical Indices. In: Psychological Methods 21 (2016), S. 137–150
  - WAINER, Howard ; BRADLOW, Eric T. ; WANG, Xiaohui: Testlet response theory and its applications. Cambridge : Cambridge University Press, 2007, S. 44
  - WANG, Wen-Chung ; WILSON, Mark: The Rasch Testlet Model. In: Applied Psychological Measurement 29 (2005), Nr. 2, S. 126–149
- Hier verwendetes Bifactor- bzw. Testlet-Modell: **Rasch Testlet Model**, vgl. W.-C. Wang und Wilson 2005.
- Zur ess. Eindimensionalität: Reise, Rodriguez und Haviland 2016, S. 137ff.
- Zum omega-Koeffizient, omegaH, ECV und PUC:vgl. Reise, Rodriguez und Haviland 2016, S. 140-145
- Von »**essentieller Eindimensionalität**« kann man dann sprechen, wenn die Daten einer Erhebung zwar eine bifaktorielle Struktur aufweisen, der Hauptfaktor aber so stark dominiert, dass praktisch von einer eindimensionalen Ladung aller Testaufgaben auf den Hauptfaktor ausgegangen werden kann.

# Erkenntnisse aus der bifaktoriellen Struktur

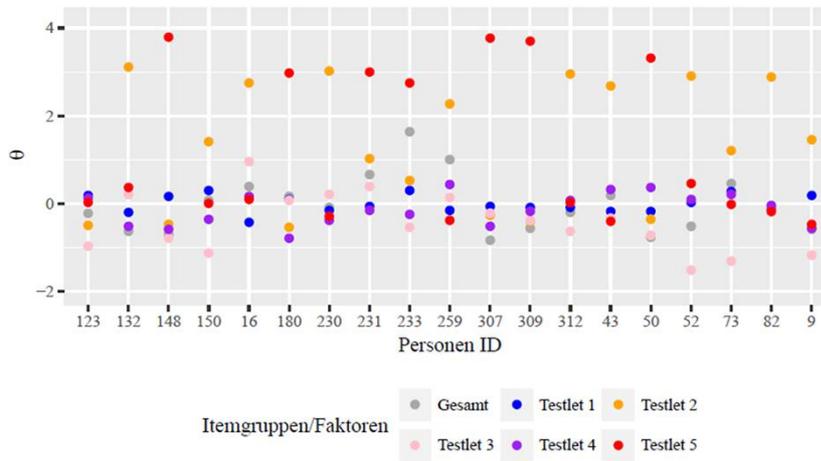
## Erkenntnisse aus einer bifaktoriellen Strukturanalyse

Interpretation der Leistungsdifferenzierung durch Testlets



- Globale Personenleistung
- Leistung im Testlet

## Erkenntnisse aus einer bifaktoriellen Strukturanalyse Identifikation der Extremgruppe mit schlechtem Personenfit



# Herausfordernde Aufgaben

## Herausfordernde Aufgaben

Wodurch zeichnen sich herausfordernde Aufgaben qualitativ aus?

- Anzahl der notwendigen Rechenschritte  $> 3$  (max. 11).
- Schriftliche oder zeichnerische Begründung (Konzeptverstehen).
- Einfache Ableitungen, Integralrechnung; Bestimmung der Entropie.
- Anwendung des 1. oder 2. Thermodynamischen Hauptsatzes notwendig, aber nicht explizit in der Aufgabenstellung gefordert.
- Keine Aufgabenführung (Aufgabe 5).

# Zusammenfassung

## Zusammenfassung

- Die ausführliche und behutsame Modellierung der realen Klausur war ein wichtiger Baustein für die Modulentwicklung.
- Der Überlegung zur angemessenen Recodierung der Daten anhand statistischer und inhaltlicher Merkmale kam besondere Bedeutung zu.
- Die drei verglichenen Modelle kamen nicht zu signifikant unterschiedlichen Parameterschätzungen.
- Auch bei vermuteter, starker Verletzung lokaler Unabhängigkeit kann die Modellierung essentielle Eindimensionalität aufweisen.
- Neben einer robusten 1PL-Schätzung der Personenfähigkeit erhält man differenzierte Strukturinformationen durch das bifaktorielle Modell.

## Konsequenzen für den Modulentwicklungsprozess

- Das Korrekturverfahren wurde angepasst, um Rechen-Punkte von Ansatz-Punkten getrennt zu erfassen.
- Die Klausurhalte wurden verringert, um den störenden Einfluss von Zeitdruck auf die Klausurgüte zu vermeiden.
- Anpassung der Ausbildung von LernbetreuerInnen, mit dem Ziel Konzeptverstehen und Ansatzfindung stärker zu fördern.
- Schwierigkeitsmerkmale der Klausuraufgaben bisher noch nicht zufriedenstellend identifiziert. Aufgabenvariationen sind geplant.
- Eine begründete Aufteilung der einen Modulklausur in zwei Teilklausuren mit unterschiedlichen Anforderungsniveaus wird angestrebt.

**Besonderer Dank für ihre Unterstützung geht an:**



**Prof. Dr. R. Nickolaus**

Institut für  
Erziehungswissenschaft  
Abteilung Berufs-,  
Wirtschafts- und  
Technikpädagogik



**B. Eng. M. Sc. S. Behrendt**

Institut für  
Erziehungswissenschaft  
Abteilung Berufs-,  
Wirtschafts- und  
Technikpädagogik



**Prof. Dr. J. Groß**

Institut für Technische  
Thermodynamik und  
Thermische  
Verfahrenstechnik

- Besonderer Dank geht an Prof. Dr. Reinhold Nickolaus und B.Eng. M.Sc. Stefan Behrendt (BWT) sowie Prof. Dr. Joachim Groß (ITT) von der Universität Stuttgart.



Universität Stuttgart

## Vielen Dank!



Dipl.-Soz. Thorsten Braun

E-Mail [Thorsten.braun@zlw.uni-stuttgart.de](mailto:Thorsten.braun@zlw.uni-stuttgart.de)

Telefon +49 (0) 711 685-82028

Fax +49 (0) 711 685-82023

Universität Stuttgart

Zentrum für Lehre und Weiterbildung | zlw

Azenbergstraße 16

70174 Stuttgart