

GEFÖRDERT VOM



Bundesministerium  
für Bildung  
und Forschung

# Kriteriumsorientiertes adaptives Testen in der Hochschule – KAT-HS

Aron Fink, Sebastian Born, Andreas Frey & Christian Spoden

FRIEDRICH-SCHILLER-  
UNIVERSITÄT  
JENA

  
KAT-HS  
Adaptive University Exams

---

## Hintergrund

- Seit Inkrafttreten der **Bologna-Reform** haben Hochschulen in der EU Studiengänge und Module auf **Erwerb von Kompetenzen** auszurichten
- Lernziele umfassen somit nicht nur **Wissen** , sondern vor allem **wirksame Handlungsweisen** zur Lösung von relevanten Problemen in wissenschafts- und berufsbezogenen Kontexten
- Umfang des Kompetenzerwerbs wird mit den **Modulprüfungen** gemessen
- Vor allem schriftlichen Prüfungen in Form von **Klausuren** gelangen mit anhaltend wachsender Studierendenzahl zu immer höherer Bedeutung
- Möglichkeit zur Identitätsfeststellung unterbindet ein bspw. bei Hausarbeiten mögliches Ghostwriting
- Hohe **persönliche Relevanz** für die Studentinnen und Studenten

---

# Problemstellung

- **Erhebliche Anforderungen** an Klausuren
- Typische Klausuren werden diesen häufig nicht gerecht
- **Psychometrie** und **Pädagogisch-Psychologische Diagnostik** stellen Handwerkszeug zur Optimierung bereit
  
- **Optimierungsmöglichkeiten** in vier Bereichen

---

# Optimierungsmöglichkeiten

1 **Operationalisierung** kompetenzorientierter Lernziele

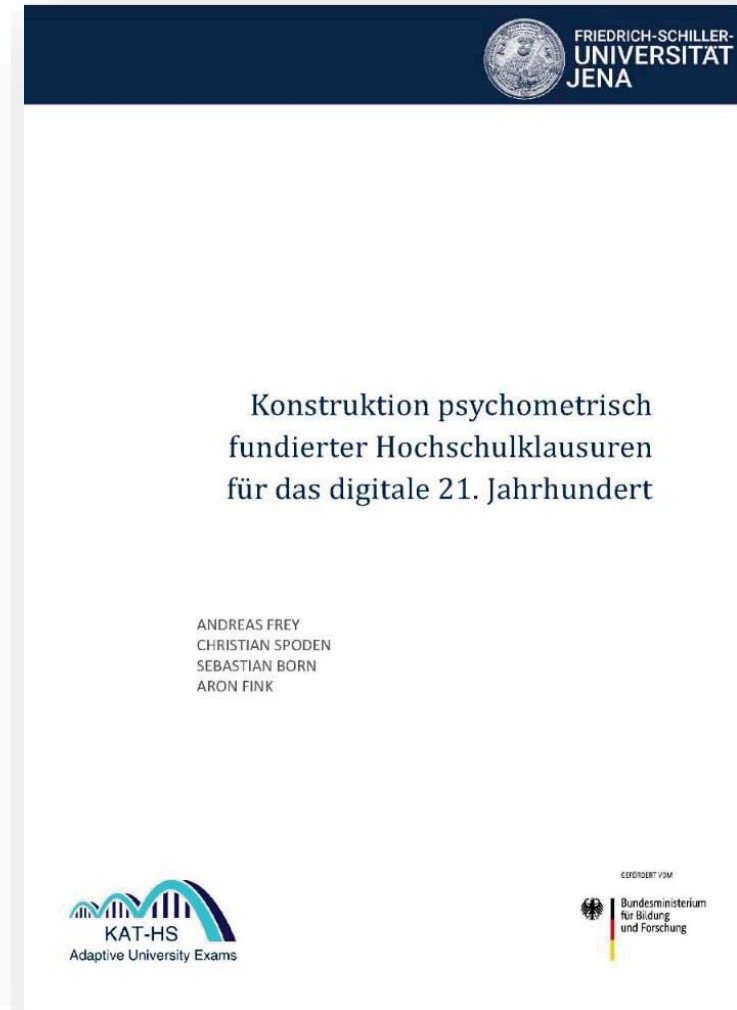
2 **Kriteriumsorientierte** Testwertinterpretation ermöglichen

3 **Verlinkung** zwischen Testzeitpunkten herstellen

4 **Messpräzision** über gemessenen Kompetenzbereich angleichen

# Konzept

- Zusammenhängende Formulierung einer **direkt umsetzbaren Möglichkeit** zur Nutzung qualitativ hochwertiger Klausuren
- **Kombination geeigneter State-of-the-art-Methoden** aus Psychometrie und Päd.-Psychologischer Diagnostik
- **Fächer- und hochschulübergreifende** Einsetzbarkeit
- Bereitstellung von **wissenschaftlich** untermauertem, umfangreichem **Handlungswissen** zur Konstruktion **kriteriumsorientierter adaptiver Klausuren** an Hochschulen



---

## Vorschau

1. Operationalisierung
2. Messmodell
3. Kriteriumsorientierung
4. Statistische Verbindung
5. Angleichung der Messpräzision
6. Einschränkungen und Ausblick

---

# 1. Operationalisierung

- Kompetenzerwerb als Lernziel von Lehrveranstaltungen fordert, dass Klausuren diese Kompetenzen auch prüfen
- **Constructive alignment** (Biggs, 1996)
  - Gegenseitige Abstimmung von: **Lernzielen, Lerngelegenheiten, summativer Messung** und **Rückmeldung**
- Zentral: Differenzierte Formulierung der **Lernziele**
- Klausur als **Kompetenztest**
- Stufen 1 und 2 der Empfehlungen von Frey und Hartig (im Druck)
  - **Spezifikation Messgegenstand** (z. B. Inhaltsbereich x kognitiver Prozess)
  - **Aufgabenentwicklung** (z. B. Haladyna & Rodriguez, 2013)
  - Erstellung **Aufgabendatenbank**

Inhaltsbereich	Kognitiver Prozess					
	1. Wissen	2. Verständnis	3. Anwendung	4. Analyse	5. Synthese	6. Bewertung
<b>1. Forschungsmethodische Grundlagen</b>						
1.1 Wissenschaftliche Erkenntnisse vs. Alltagswissen						
1.2 Ziele wissenschaftlicher Tätigkeit						
1.3 Ablauf empirische Untersuchung & zentrale Begriffe						
1.4 Systematik der Forschungsmethoden						
1.5 Forschungsethik						
<b>2. Planung quantitativer Untersuchungen</b>						
2.1 Das Experiment						
2.2 Versuchspläne						
2.3 Stichprobe und Grundgesamtheit						
<b>3. Quantitative Erhebungsmethoden</b>						
3.1 Skalenniveaus						
3.2 Reaktivität						
3.3 Beobachtung						
3.4 Befragung						
3.5 Gütekriterien						
<b>4. Planung qualitativer Untersuchungen</b>						
4.1 Prinzipien qualitativen Forschens						
4.2 Stichprobenziehung						
4.3 Qualitative Forschungsansätze						
<b>5. Qualitative Erhebungsmethoden</b>						
5.1 Interview						
5.2 Gruppendiskussion						
5.3 Lautes Denken - Cognitive Lab						
5.4 Beobachtung						
<b>6. Qualitative Auswertungsmethoden</b>						
6.1 Codierung						
6.2 Inhaltsanalyse						
6.3 Systematisierung						
6.4 Bewertung qualitativer Methoden						



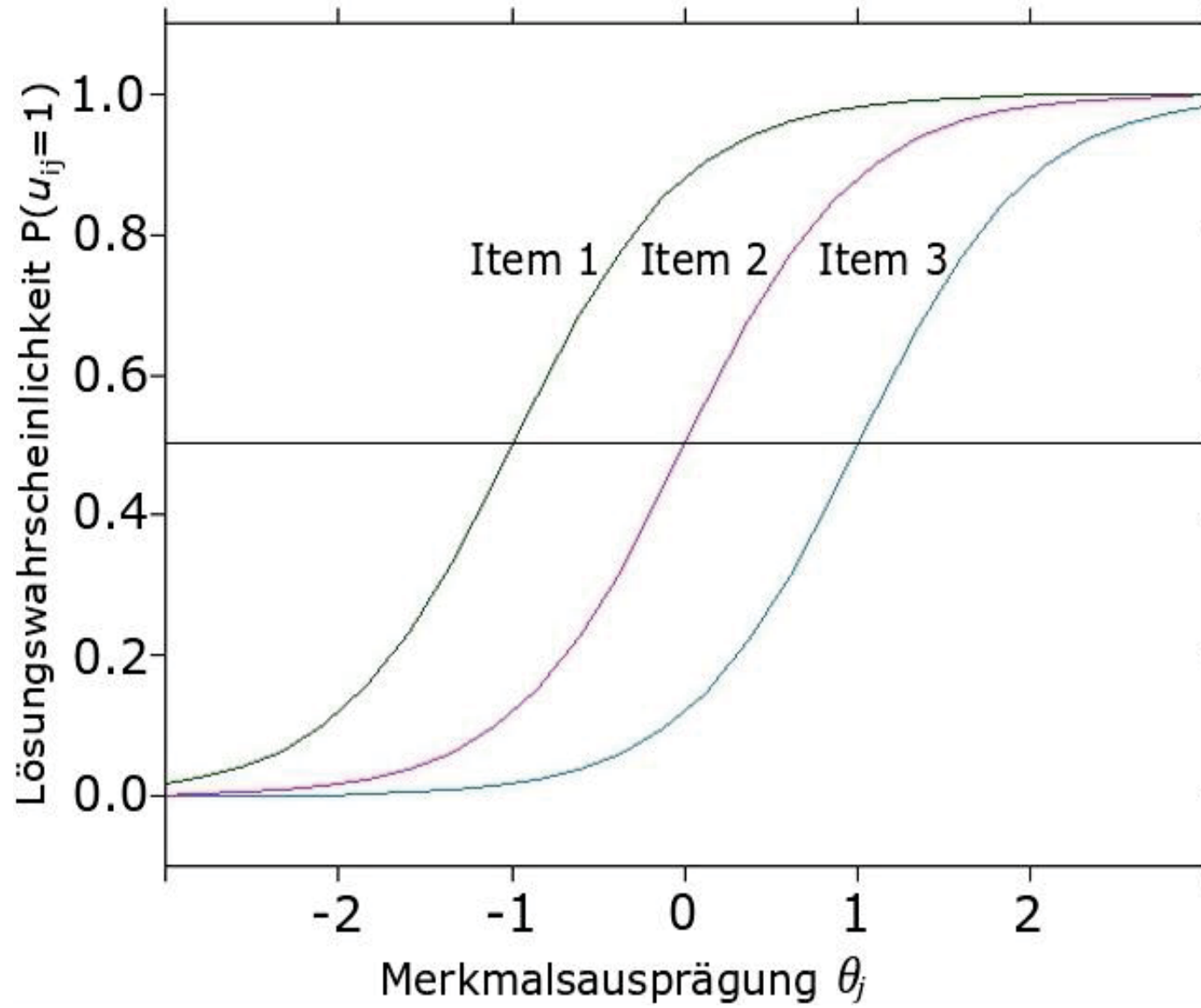
---

## 2. Messmodell

- **Messmodell** zur mathematischen Beschreibung der Beziehung zwischen **beobachtbarem Verhalten** (Item-Response;  $u_{ij}$ ) und **zu messenden Kompetenzen** notwendig
- Für kriteriumsorientierte Tests sind Modelle der **Item-Response-Theorie** (IRT; z. B. van der Linden, 2016) geeignet → *unidimensional logistic response models*
- Annahmen: *Eindimensionalität; lokale stochastische Unabhängigkeit; Monotonizität*
- **Dichotomes Rasch-Modell** (Rasch, 1960) aufgrund der vorzufindenden Rahmenbedingungen an Hochschulen empfehlenswert → **Fähigkeit**  $\theta_j$  und **Itemschwierigkeit**  $b_i$  auf derselben Skala lokalisiert

$$P(u_{ij} = 1 | \theta_j, b_i) = \frac{\exp(\theta_j - b_i)}{1 + \exp(\theta_j - b_i)}$$

# ICC



---

## 3. Kriteriumsorientierung

- momentan häufig verwendete Klausurbewertung ist defizitär, da sie normorientiert bzw. orientiert an willkürlich festgelegten Kriterien erfolgt
- Klausuren als **kriteriumsorientierte Tests** (z. B. Herzberg & Frey, 2011)
- Testergebnis ist ohne sozialen Vergleich interpretierbar: **Relation zum Lernziel**



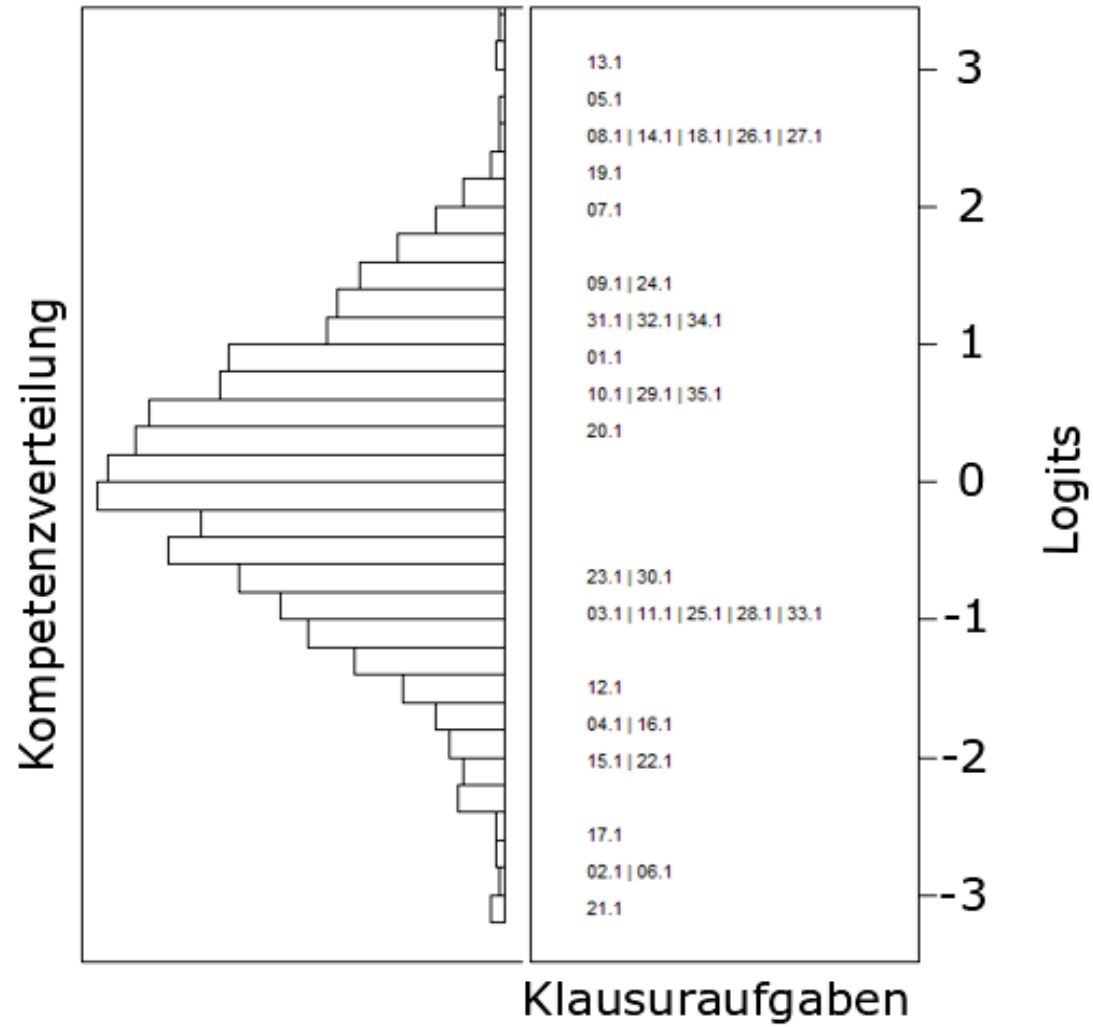
---

### 3. Kriteriumsorientierung

- Notenstufen als direkte Repräsentanten des Ausmaßes der Lernzielerreichung  
→ **Standard Setting** (z. B. Bookmark-Methode; Mitzel et al., 2001)
- verbindet die zunächst rein quantitativ repräsentiere Testskala mit dem Inhalt des Messgegenstands



# Wright Map



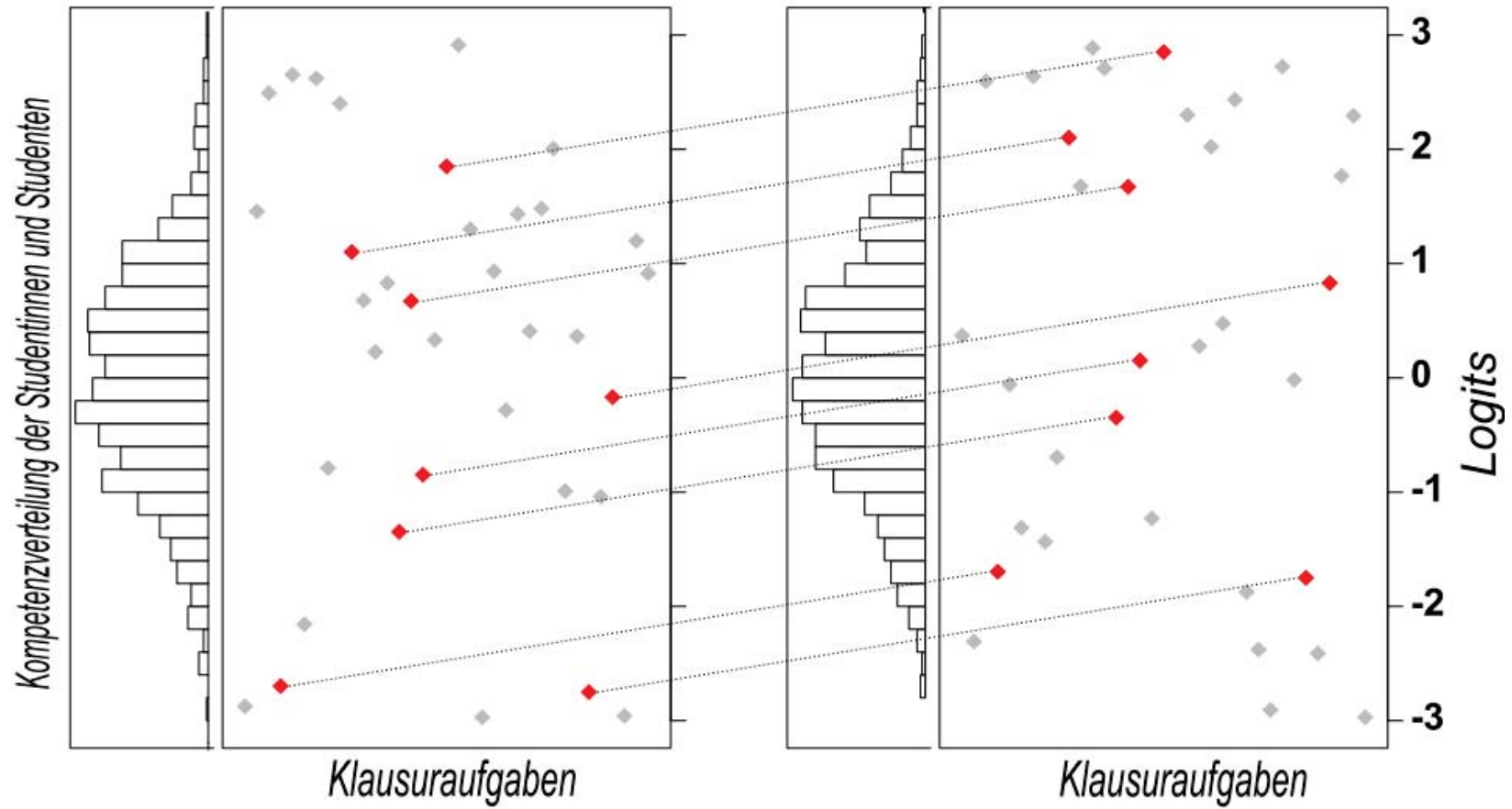
Aus: Frey, Spoden, Born & Fink, 2017, S. 28

---

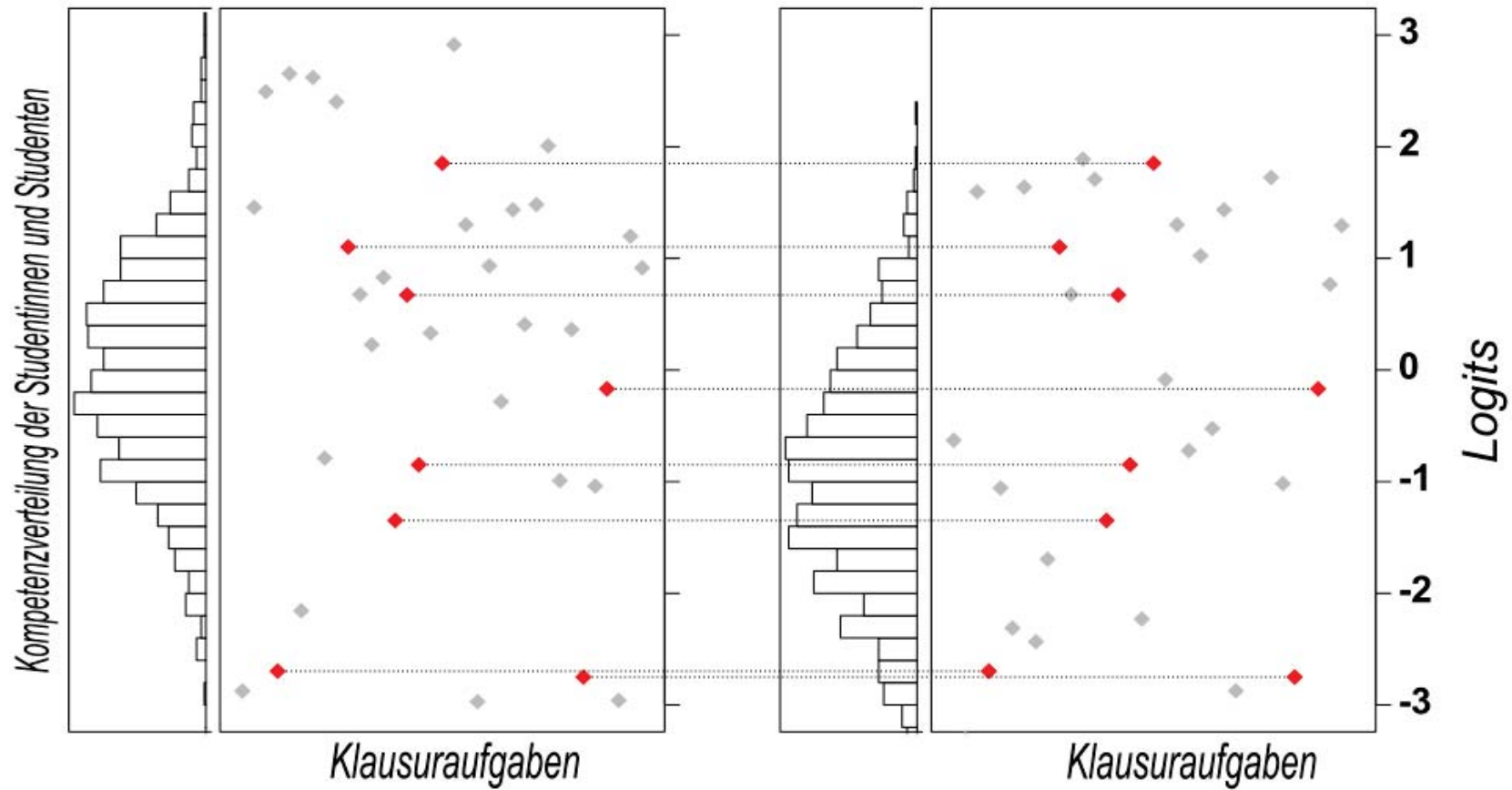
## 4. Statistische Verbindung

- Aus Sicherheitsgründen werden häufig unterschiedliche Klausurzusammenstellungen für verschiedene Testzeitpunkte verwendet
- Nutzung von **Equating-Methoden** um sicherzustellen, dass der gleiche Maßstab zum Erreichen der einzelnen Notenstufen für unterschiedliche Testzusammenstellungen verwendet wird
- Equating-Methoden führen zu einer **Adjustierung der Schwierigkeit** verschiedener Klausurzusammenstellungen, ABER nicht zu einer Adjustierung des Inhalts
- Empfehlung: **Common-Item Equating to a Calibrated Pool** (Kolen & Brennan, 2014)

# Wright Map



## Wright Map



Aus: Frey, Spoden, Born & Fink, 2017, S. 45



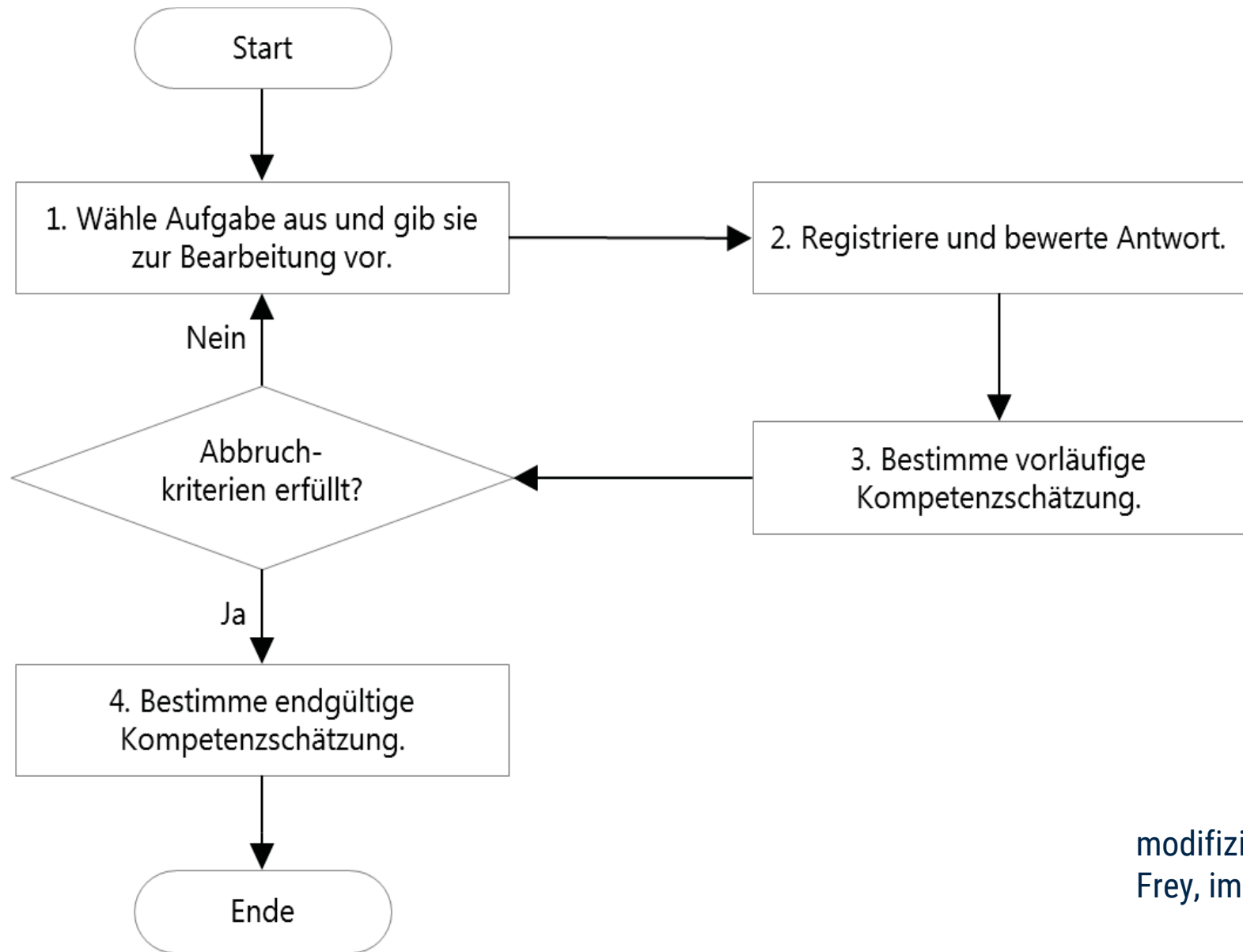
---

## 5. Angleichung der Messpräzision

- Angleichung der **Messpräzision** über gesamte Merkmalsverteilung ist durch Nutzung **computerisierten adaptiven Testens** (CAT) möglich (Frey, 2012)
- Aufgabenauswahl aufgrund des bisherigen Antwortverhaltens der Testperson
- Nur diejenigen Items vorgelegt, die besonders viel **diagnostische Information** enthalten → Iteminformationsfunktion

$$I_i(\theta) = P_i(\theta)(1 - P_i(\theta))$$

- Maximum bei  $P_i(\theta) = 0.5$ , d. h. bei  $\theta = b_i$
- Empfehlung: für Klausuren etwas **höhere Lösungswahrscheinlichkeit** nutzen



modifiziert aus:  
Frey, im Druck

---

## 6. Einschränkungen und Ausblick

- Für CAT benötigt man einen hinreichend großen, kalibrierten **Itempool**
- Rahmenbedingungen an der Hochschule für klassische Kalibrierungsstudie eher ungeeignet → Methode der **kontinuierlichen Kalibrierung** (Fink et al., 2017)
- **Publikation** des Konzepts als Buch in Vorbereitung
- Für kriteriumsorientierte adaptive Klausuren benötigt man außerdem geeignete Software, die möglichst niedrighschwellig implementierbar ist → Entwicklung eines **R-Pakets** zur Umsetzung des Konzepts
- Bereitstellung weiterer Materialien, um **Implementation** zu erleichtern: Manual, Workshop, Online-Portal

---

## Fazit

- Übliche Hochschulklausuren werden hoher Relevanz nicht immer gerecht
- Psychometrie und Päd.-Psychol. Diagnostik bieten Optimierungsmöglichkeiten
  - Operationalisierung
  - Kriteriumsorientierung
  - Verlinkung
  - Adaptivität
- Ziel von KAT-HS: Möglichst niedrighschwellige Angebote zur einfachen Implementation psychometrisch fundierter Klausuren

[www.methods.uni-jena.de](http://www.methods.uni-jena.de)

---

Vielen Dank für  
die Aufmerksamkeit





---

# Literatur

- Bloom, B. S., Engelhart, M. D., Furst, E. J., Hill, W. H. & Krathwohl, D. R. (1956). *Taxonomy of educational objectives: The classification of educational goals. Handbook I: Cognitive domain*. New York, Toronto: Longmans, Green.
- Biggs, J. (1996). Enhancing teaching through constructive alignment. *Higher Education*, 32, 347-364.
- Frey, A. (2012). Adaptives Testen. In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (2. aktualisierte und überarbeitete Auflage, S. 275-293). Berlin, Heidelberg: Springer.
- Frey, A. (im Druck). Computerisiertes adaptives Testen. In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (3. aktualisierte und überarbeitete Auflage. Berlin, Heidelberg: Springer.
- Frey, A. & Hartig, J. (im Druck). Kompetenzdiagnostik. In M. Gläser-Zikuda, M. Harring, & C. Rohlfs (Hrsg.), *Handbuch Schulpädagogik*. Münster: Waxmann.
- Frey, A., Spoden, C., Born, S. & Fink, A. (2017). Konstruktion psychometrisch fundierter Hochschulklausuren für das digitale 21. Jahrhundert. Unveröffentlichtes Manuskript.
- Fink, A., Born, S., Spoden, C. & Frey, A. (in prep) (2017). *A continuous calibration method for computerized adaptive exams*. Manuscript in preparation.
- Haladyna, T. M. & Rodriguez, M. C. (2013). *Developing and validating multiple-choice test items*. New York, NY: Taylor & Francis.
- Herzberg, P. Y. & Frey, A. (2011). Kriteriumsorientierte Diagnostik. In: L. F. Hornke, M. Amelang & M. Kersting (Hrsg.), *Methoden der psychologischen Diagnostik. Enzyklopädie der Psychologie, B/II/2* (S. 281-324). Göttingen: Hogrefe.
- Kolen, M. J. & Brennan, R. L. (2014). *Test equating, scaling, and linking: Methods and practices* (3<sup>rd</sup> ed.). New York, NY: Springer.
- Mitzel, H. C., Lewis, D. M., Patz, R. J. & Green, D. R. (2001). The bookmark method: Psychological perspectives. In G. J. Cizek (Ed.), *Setting performance standards: Concepts, methods, and perspectives* (pp. 249-281). Mahwah, NJ: Erlbaum.
- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen: Danmarks Pædagogiske Institut.
- van der Linden, W. J. (Hrsg.). (2016). *Handbook of item response theory. Volume one: Models*. Boca Raton: Chapman & Hall/CRC.