



**DIPF**

Educational Research  
and Educational Information

*Möglichkeiten und Grenzen der Politik- und  
Praxisberatung auf Basis von PISA-Studien*

# Computerbasiertes Testen – Mode oder neue Erkenntnischance

*Ulf Kröhne & Frank Goldhammer*

Zentrum für Technologiebasiertes Testen (TBA)

# Agenda

- **Mode?**
  - **PISA und CBA**
  - **Modus-Wechsel**
  
- **Neue Erkenntnischancen durch CBA**
  - Innovative Aufgaben
  - Log- und Prozessdaten
  - Steigerung der Messeffizienz
  - Standardisierung
  
- **Resümee**
  - Politik- und Praxisberatung
  - Chancen und Herausforderungen für Schule

# CBA in PISA

2000

- Computer-Based Assessment von **dynamischem Problemlösen (PS)** als Teil der nationalen Ergänzung in Deutschland

2006

- Computer-Based Assessment of **Science (CBAS)** als internationale Option, Auswahl an Teilnehmerstaaten

2009

- Electronic **Reading Assessment (ERA)** als internationale Option, Auswahl an Teilnehmerstaaten

2012

- CBA of **Mathematical Literacy (CBAM)** als internationale Option, Auswahl an Teilnehmerstaaten


2015

- Umstieg auf computerbasiertes Testen für **alle Domänen**, nur noch weniger Länder weiterhin PBA

t ↓

**Question 2: Sound Detective (DM002Q02)**

The monitor screen displays an image of an artery in Anastasia's neck. The colour changes show blood moving at different speeds inside the artery.



Which one of the following statements explains the regular changes in the blood flow in Anastasia's artery?

- The artery is squeezed when she swallows.
- The artery is squeezed when she breathes in.
- The artery is stretched by the ultrasound probe on her neck.
- The artery is stretched when her heart pushes blood through it.

Philosophers' Café - Home



Philosophers have been dealing with these kinds of questions for thousands of years.


**PHILOSOPHERS' CAFÉ: Task 3 (2015/01)**

Go to the Activities page. Find the Timeline. Use the information in the website to find when each of the three philosophers (Bentham, Confucius and Zeno) was born. Then use the drop down menus in the timeline to show when each philosopher was born.

PISA 2015

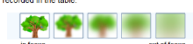
**Adjustable Classes**  
Running the Simulation

In this simulation, you will be able to see how the amount of fluid in the lens affects the student's ability to see a tree clearly from each of the three distances shown below.




To see how all the controls in this simulation work, follow these steps:

1. Move the slider for **amount of fluid in lens**.
2. Select the **distance from tree**.
3. Click the "Run" button to see whether the tree will appear in focus or out of focus to the student. The results will be recorded in the table.



**Anna's View**



**Amount of Fluid in Lens**

Distance from Tree: near, midway, distant

**Run**

Distance from Tree	Amount of Fluid in Lens			
	-2	-1	0	+2
Near				
Midway				
Distant				

# Modus-Wechsel im Assessment

- Äquivalenz

PBA


**R077: Flu**

---

**ACOL VOLUNTARY FLU IMMUNISATION PROGRAM**

As you are no doubt aware the flu can strike rapidly and extensively during winter. It can leave its victims ill for weeks.

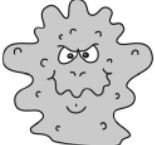
The best way to fight the virus is to have a fit and healthy body. Daily exercise and a diet including plenty of fruit and vegetables are highly recommended to assist the immune system to fight this invading virus.



ACOL has decided to offer staff the opportunity to be immunised against the flu as an additional way to prevent this insidious virus from spreading amongst us. ACOL has arranged for a nurse to administer the immunisations at ACOL, during a half-day session in work hours in the week of May 17. This program is free and available to all members of staff.

Participation is voluntary. Staff taking up the option will be asked to sign a consent form indicating that they do not have any allergies, and that they understand they may experience minor side effects.

Medical advice indicates that the immunisation does not produce influenza. However, it may cause some side effects such as fatigue, mild fever and tenderness of the arm.



CBA

1 / 4

**ACOL VOLUNTARY FLU IMMUNISATION PROGRAM**

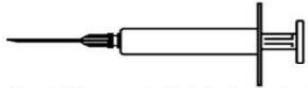
As you are no doubt aware the flu can strike rapidly and extensively during winter. It can leave its victims ill for weeks.

**Frage 1** ?

**Frage 2** ?

**Frage 3** ?


**Frage 4** ?



ACOL has decided to offer staff the opportunity to be immunised against the flu as an additional way to prevent this insidious virus from spreading amongst us. ACOL has arranged for a nurse to administer the immunisations at ACOL, during a half-day session in work hours in the week of May 17. This program is free and available to all members of staff.

Participation is voluntary. Staff taking up the option will be asked to sign a consent form indicating that they do not have any allergies, and that they understand they may experience minor side effects.

Medical advice indicates that the immunisation does not produce influenza. However, it may cause some side effects such as fatigue, mild fever and tenderness of the arm.



<
>
beenden

# PISA 2015 (Feldtest)

- Schüler im Feldtest zufällig CBA oder PBA zugewiesen
- Unterschiede in den Ergebnissen können „kausal“ auf den Modus zurückgeführt werden
- Analyse der Vergleichbarkeit von Items (Invarianz)

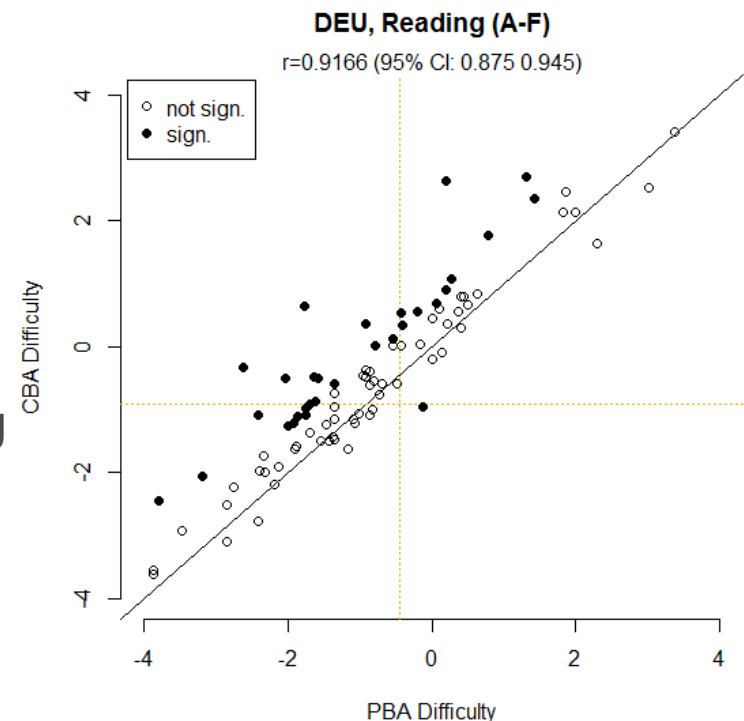


i) Items ohne Änderung von Diskrimination & Schwierigkeit (skalare Invarianz) → Link-Items





ii) Items ohne Änderung der Diskrimination, aber mit Änderung der Schwierigkeit, mehrheitlich schwieriger (metrische Invarianz) (→ nicht invariante Items)



iii) Items mit Änderung der Diskrimination → Ausschluss

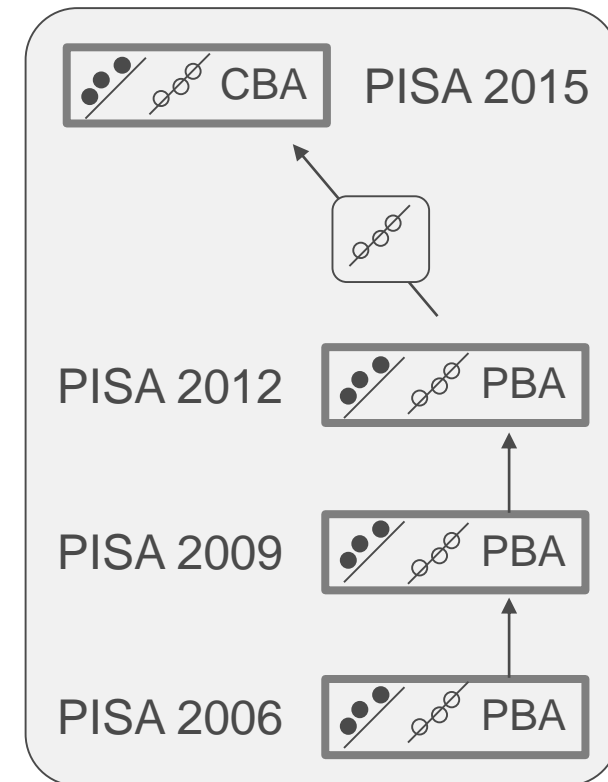


# PISA 2015 (Haupttest)

- Nutzung der identifizierten Link-Items  für Herstellung von Trendinformationen, Modus-Effekte für nicht-invariante Items „schaden“ nicht 
- Weitere methodische Neuerungen (z.B. Nutzung der Daten aller Zyklen)

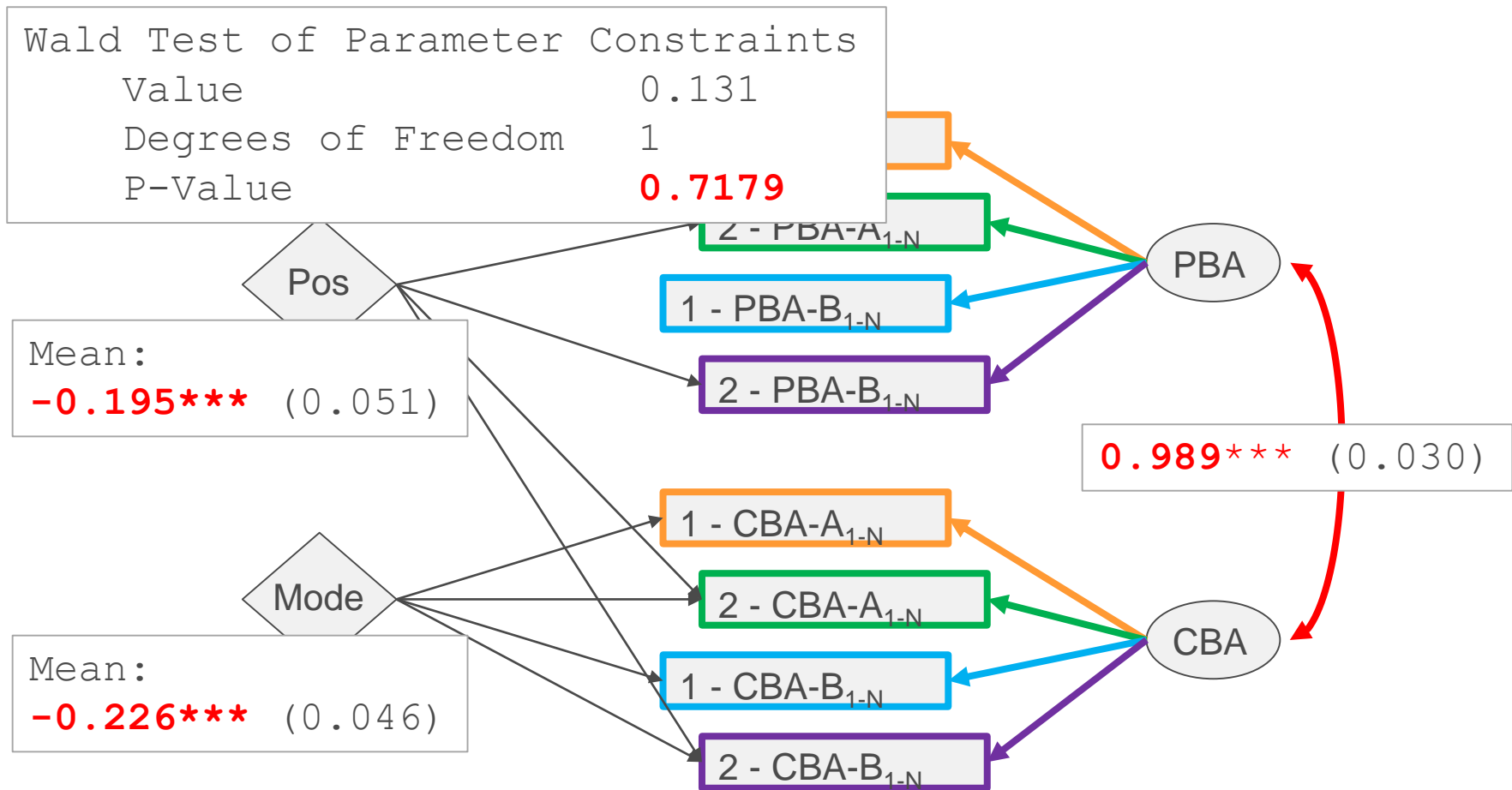
## Mögliche Kritikpunkte

- keine Nutzung des geschätzten Modeffekts aus Feldtest  ↔ 
- keine Betrachtung länderspezifischer Modus-Effekte
- Trennung Link-Items vs. nicht invariante Items
- keine direkte Prüfung der Konstruktäquivalenz





# Konstruktäquivalenz CBA und PBA

Analyse von PISA Lesekompetenz (nat. Ergänzung PISA 2012)



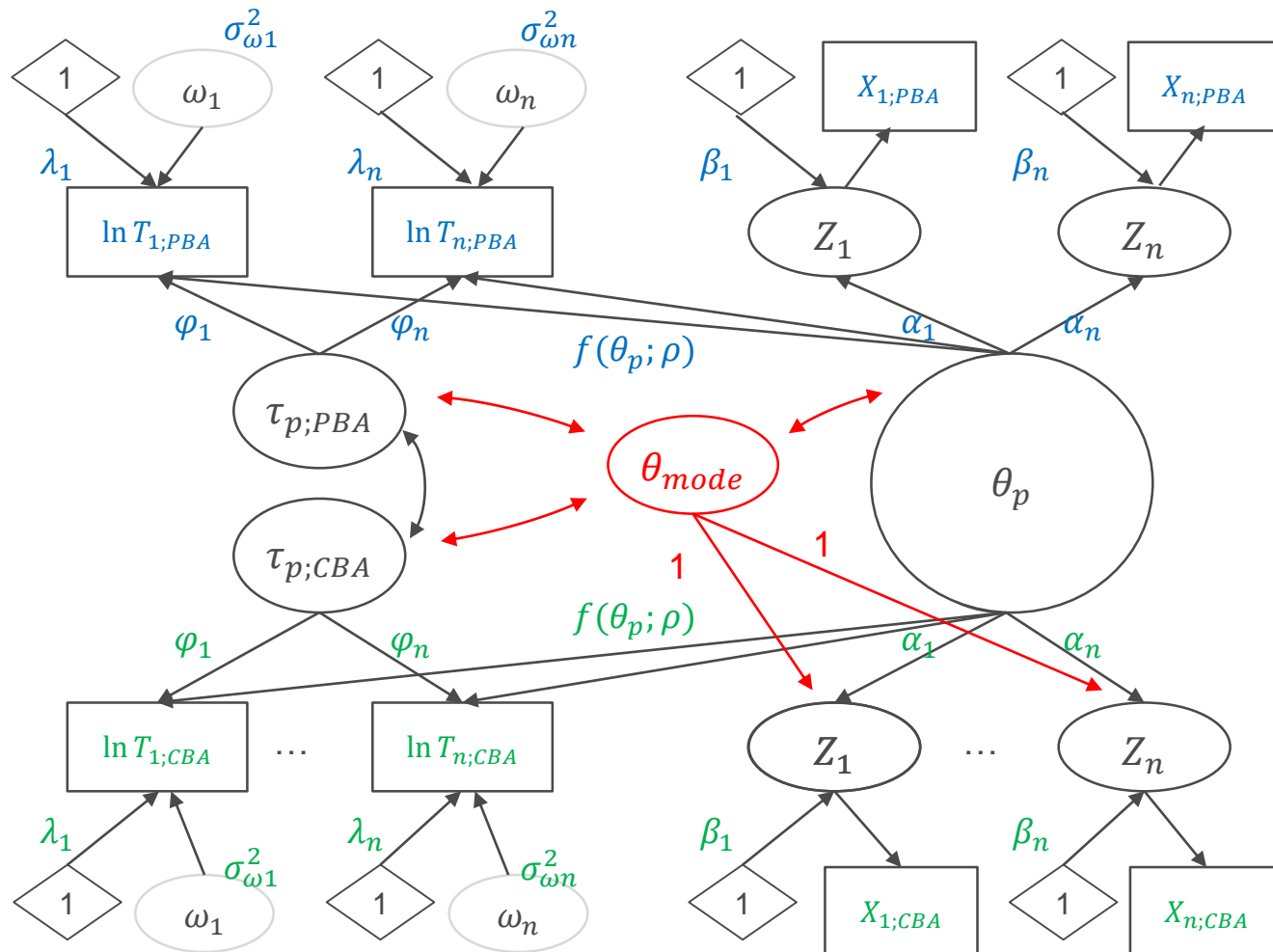
# CBA nur Mode?

- CBA geht mit Mode-Effekten einher, d.h. identische Aufgaben sind in CBA häufig etwas schwerer in PBA
- weiterer Forschungsbedarf zu PISA 2015 bzgl. der Kritikpunkte
  - Wirksamkeit der Modus-Korrektur  vs. 
  - Länderunterschiede?
- weiterer Forschungsbedarf zu Bedeutung von Modus-Effekten
  - allgemeine Erklärungen / theoretische Fundierung
  - Bedeutung der Unterschiede für Schule & Unterricht



# Erklärungen für Modus-Effekte?

Beispiel: Bearbeitungszeiten (PBA vs. CBA) beim Lesen



## Beispiel: Ergebnis des Vergleichs von CBA und PBA (mit E-Pens)

	$MI\ 7\ (\alpha_i, \beta_i, \lambda_i, \sigma_{\omega i}^2)$
Modus-Effekt (Fähigkeitsunterschiede CBA vs. PBA)	-0.249** (0.050)
Interindividuelle Unterschiede im Mode-Effekts (Varianz)	0.437** (0.098)
Korrelation Modus-Effekt und Fähigkeit	-0.683** (0.058)
Intreind. Unterschiede in der Bearbeitungszeit CBA (Varianz)	0.748** (0.045)
Intreind. Unterschiede in der Bearbeitungszeit PBA (Varianz)	0.605** (0.049)
Korrelation Zeit und Modus-Effekt für PBA	-0.243** (0.075)
Korrelation Zeit und Modus-Effekt für CBA	0.306** (0.067)
Korrelation Zeit und Fähigkeit für PBA	-0.462** (0.044)
Korrelation Zeit und Fähigkeit für CBA	-0.591** (0.041)
Korrelation Zeit PBA und CBA	0.423** (0.067)

$$\chi^2 = 18.895$$

$$df=1 \quad p < 0.05$$

# Agenda

- Mode?
  - PISA und CBA
  - Modus-Wechsel
- **Neue Erkenntnischancen durch CBA**
  - **Innovative Aufgaben**
  - **Log- und Prozessdaten**
  - **Steigerung der Messeffizienz**
  - **Standardisierung**
- Resümee
  - Politik- und Praxisberatung
  - Chancen und Herausforderungen für Schule

# Neue Erkenntnischancen durch CBA (1)

## Innovative Aufgaben

- Merkmale messen, die mit auf Papier überhaupt nicht oder nur schwer messbar sind
  - Kompetenz, digitale Texte zu Lesen (Electronic Reading)
  - Umgang mit multiplen Dokumenten (MDL)
- Simulationsbasiertes Assessment / interaktive Items (Diagnostik in authentischen Kontexten)
  - Beurteilung von Glaubwürdigkeit ICT Literacy (z.B. Simulation von Browser und Websites)

## Neue Erkenntnisse durch Validitätssteigerung

- empirische Abgrenzung / Mehrwert muss je Domäne erfolgen (wie verhalten sich alte zu neuen Bestandteile des Konstrukts?)

# Neue Erkenntnischancen durch CBA (2)

## Log- und Prozessdaten

- Spuren der Testbearbeitung (Mauseklicks, Tastatureingaben, ...) entstehen „als Abfallprodukt“ nebenbei
- Systematische Nutzung in Verbindung mit dem diagnostischen Zweck (technologiebasierter) Assessments
- Einblick in Lösungsverhalten, bspw. Auswahl, Sequenz und Dauer von Bearbeitungsschritten usw.

## Herausforderungen

- Inhaltlich: Systemtische Konstruktion von Aufgaben zur Differenzierung von Lösungsverhalten
- Technisch: Datenqualität von Log-Daten, theoretischer Rahmen für Ableitung von Indikatoren, Validierung von Indikatoren, Datenformat für die Handhabung von großen Datenmengen

# Bsp.: Komplexes Problemlösen

“Probieren” vs. systematisches  
Explorieren

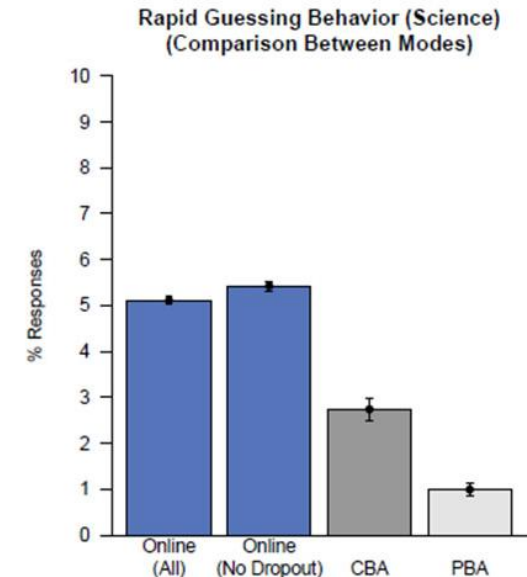
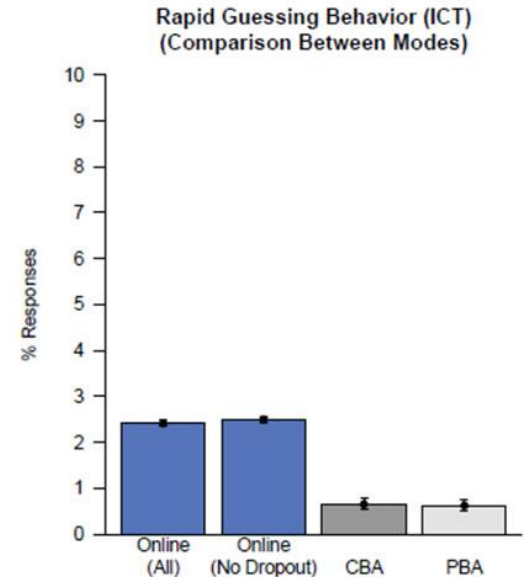
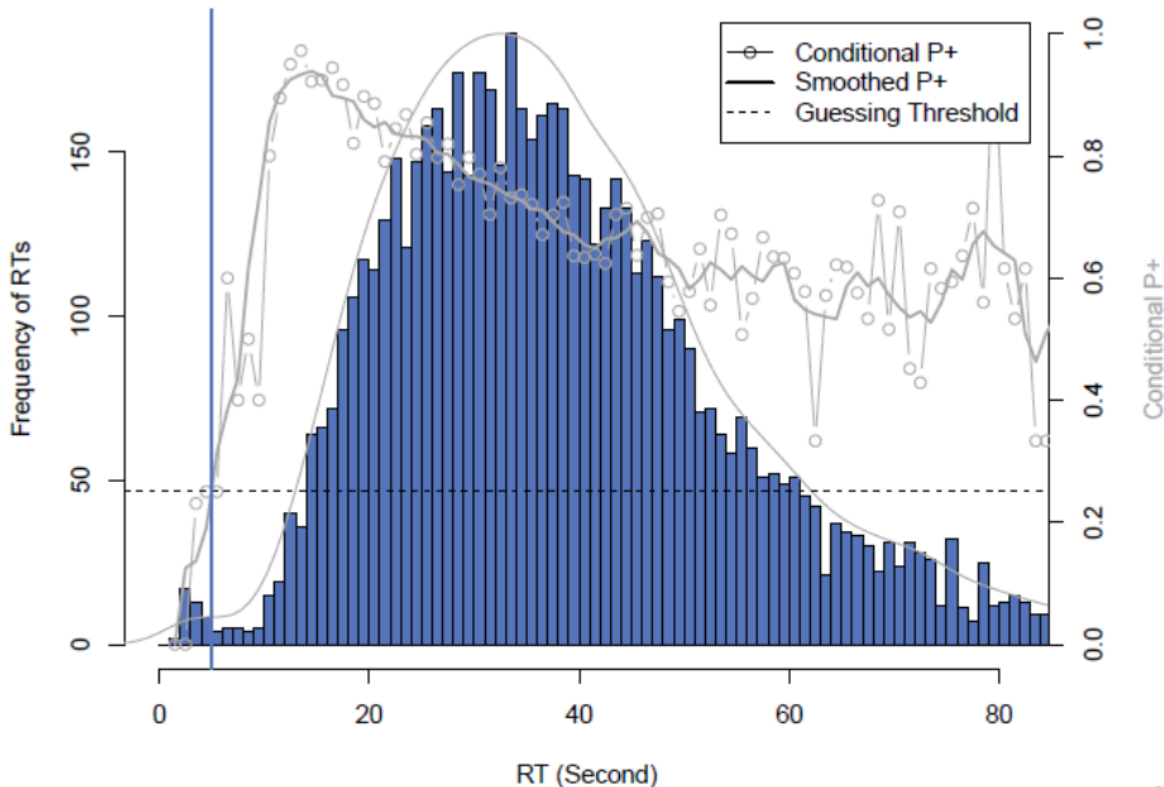
- Identifikation der  
Bearbeitungsstrategie  
“**Vary one thing  
at a time (VOTAT)**”  
aus den Log-Daten

Strategie	Lösung der Aufgabe „Climate Control“ (Teilaufgabe 1)		Gesamt
	falsch	richtig	
ohne VOTAT	42.2%	9.7%	51.9%
VOTAT	6.9%	41.1%	48.1%
Gesamt	49.1%	50.9%	100.0%

# Bsp.: Motivation

Klassifizierung von Antworten nach Antwortzeit

- (sehr) schnelle Antwort können als „Raten“ interpretiert werden



# Agenda

- Mode?
  - PISA und CBA
  - Modus-Wechsel
- Neue Erkenntnischancen durch CBA
  - Log- und Prozessdaten
  - **Steigerung der Messeffizienz**
  - **Standardisierung**
- Resümee
  - Politik- und Praxisberatung
  - Chancen und Herausforderungen für Schule



# Messeffizienz

## Gründe für Messeffizienzsteigerung

- Steigerung der Reliabilität (per se)
- Steigerung der Validität durch Ermöglichung von (zeitintensiveren) Simulationsbasierten Aufgaben
- adaptives testen als Werkzeugkasten  
→ muss an Bedarfe der Diagnostik angepasst werden

	mehr Information	mehr Personen
Populationsparameter (bspw. Mittelwerte)	x	x
Klassifikationsentscheidungen (bspw. Kompetenzstufen / Risikogruppen)	x	
(latente) Veränderung / Change-Score	x	
Individuelle Interpretation von Ergebnissen (z.B. Elternrückmeldungen)	x	

# Standardisierung

## Gütekriterium „Objektivität“

- Audio-Ausgaben in Gruppentestungen (Kopfhörer und CBA vs. Lautsprecher, Audio-CD und Testheft)
- Standardisierung von Instruktionen (Audio-, Videofiles) und Log-Daten vs. Interviewerhandbücher
- Verhinderung von Instruktions-inkonsistentem Verhalten (bspw. Überschreiten von Zeitgrenzen)
- Homogenisieren von Bearbeitungsgeschwindigkeit (bspw. durch Zeitfeedback)
- .....

# Agenda

- Mode?
  - PISA und CBA
  - Modus-Wechsel
- Neue Erkenntnischancen durch CBA
  - Innovative Aufgaben
  - Log- und Prozessdaten
  - Steigerung der Messeffizienz
  - Standardisierung
- **Resümee**
  - **Politik- und Praxisberatung**
  - **Chancen und Herausforderungen für Schule**

# Politik- und Praxisberatung (1)

- Erhebungsmodus primär inhaltlich zu motivieren
  - Handschrift, Geometrie, Formeln usw. (→ PBA),  
interaktive Assessments, Videos, Simulationen (→ CBA)
  - Kompetenzdiagnostik in dem Medium, in welchem Kompetenz  
auch im Alltag angewendet wird
  - wenig gute Gründe für PBA als Referenzkategorie und für  
Beibehaltung von „PBA“-Gewohnheiten bei CBA-Assessments
- Chancen liegen in den klassischen Gütekriterien
  - Reliabilität / Messeffizienz → sorgsamere Umgang mit Testzeit
  - Validität → authentischere Erfassung von Konstrukten
  - Objektivität → Standardisierung / positive Effekte durch CBA

# Politik- und Praxisberatung (2)

- Vorbildfunktion von PISA für Kompetenzdiagnostik?
  - Pro: „Dambruch“ für Anpassung Testmediums
  - Contra: Internationale Vergleichbarkeit des Modus-Effekts offen
  - Stellenwert methodischer Grundlagen für Interpretierbarkeit empirischer Befunde

# Chancen für Schule

- über Gütekriterien hinaus bietet CBA positive Seiteneffekte für Schule
  - Reduktion des operative Aufwands für Assessments (bspw. Chancen für Vergleichsarbeiten, formatives Assessment, ...)
  - schnellere Rückmeldung von Ergebnissen zu Assessments
  - zeitlich flexiblerer Einsatz von Diagnostik zur Individualisierung
  - Einblick in Testbearbeitungsprozess (Log- und Prozessdaten), wenn Aufgaben konsequent für CBA entwickelt werden

# Herausforderungen für Schule

- Schulinfrastruktur notwendig, um CBA außerhalb von PISA zum „Regelfall“ werden zu lassen
- Ausgleich interindividueller Unterschiede (→ Testfairness)
  - Konfundierende Einflüsse von konstruktirrelevanten Variablen, z.B. Computerliteralität, können Subgruppen benachteiligen
  - können bei Testkonstruktion nur bedingt ausgeglichen werden (Tutorials, ggf. stat. Kontrolle oder Ausschluss von Items)
- Effekt des Testmediums
  - Schule (und Wissenschaften) müssen damit „umgehen“, dass identische Aufgaben am Computer „schwieriger“ sind

Vielen Dank!  
– Fragen, Kommentare?

[kroehne@dipf.de](mailto:kroehne@dipf.de)