

# Computer-basiertes Testen in Large Scale Assessments: Mehr und bessere Verhaltensdaten?

Frank Goldhammer, Jean-Paul Reeß

(DIPF - Deutsches Institut für Internationale Pädagogische Forschung, Frankfurt a.M.)



Konferenz für Sozial- und Wirtschaftsdaten (KS WD), Wiesbaden, 14.1.2011

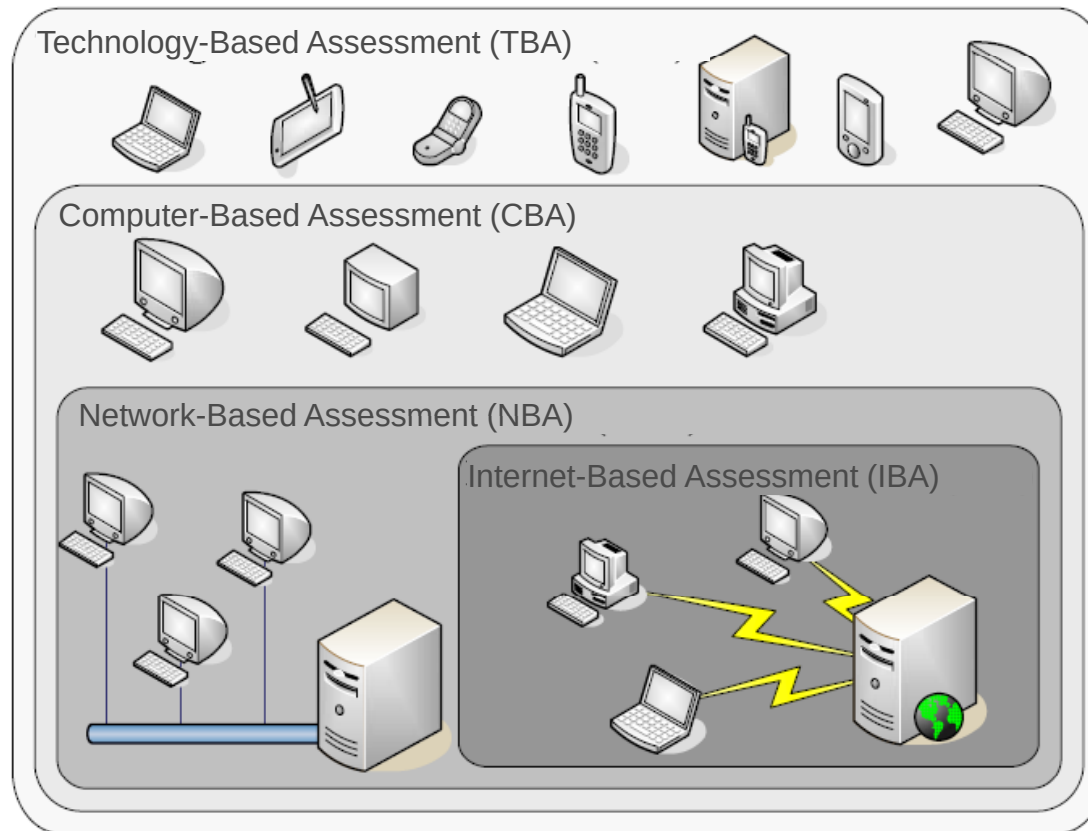
# Inhalt

- Computer-basiertes Testen (CBT) in Large Scale Assessments
- Chancen von CBT
- Herausforderungen von CBT
- Strategien zur Einführung von CBT in Large Scale Assessments
- Fallbeispiel I: PISA 2009 Electronic Reading
- Fallbeispiel II: PIAAC Computerisierung von Leseaufgaben
- Fazit

# Inhalt

- Computer-basiertes Testen (CBT) in Large Scale Assessments
- Chancen von CBT
- Herausforderungen von CBT
- Strategien zur Einführung von CBT in Large Scale Assessments
- Fallbeispiel I: PISA 2009 Electronic Reading
- Fallbeispiel II: PIAAC Computerisierung von Leseaufgaben
- Fazit

# Computer-basiertes Testen (CBT) in Large Scale Assessments



## Konstruktbereiche:

- Kompetenzvariablen  
(kognitives Assessment)
- Persönlichkeitsvariablen
- Sozio-demographische  
Variablen
- ...

# Inhalt

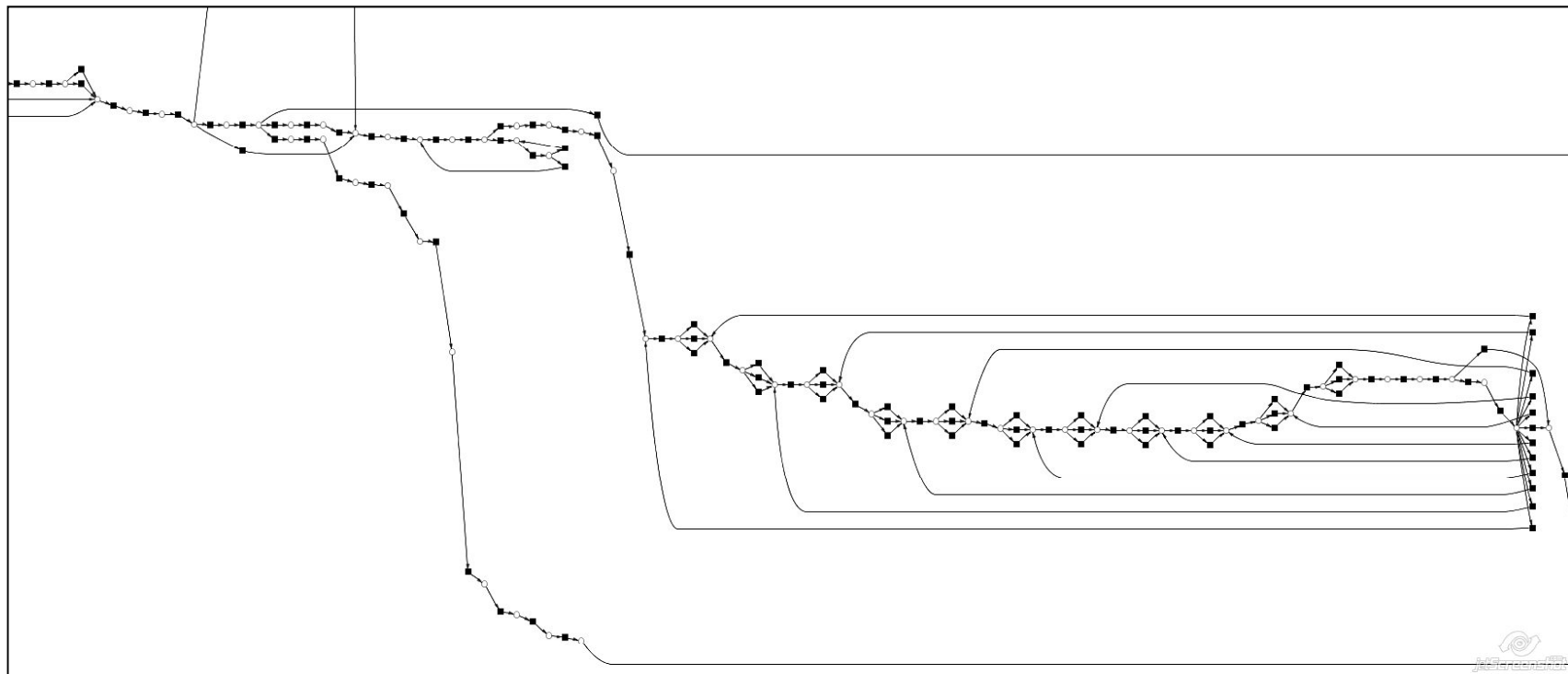
- Computer-basiertes Testen (CBT) in Large Scale Assessments
- Chancen von CBT
- Herausforderungen von CBT
- Strategien zur Einführung von CBT in Large Scale Assessments
- Fallbeispiel I: PISA 2009 Electronic Reading
- Fallbeispiel II: PIAAC Computerisierung von Leseaufgaben
- Fazit

## Chancen von CBT bzgl. Datenqualität

- **Objektivität:** Durch standardisierte Auslieferung, Instruktion und automatische Antwortbewertung geringerer Einfluss von Störgrößen (z.B. Versuchsleitereffekte).
- **Messeffizienz:** Durch adaptive Itemauswahl höhere Reliabilität bzw. Einsparung von Testzeit durch Vermeidung inadäquater Items (z.B. adaptives Testen für Kompetenzassessment, antwortabhängiges Routing in Fragebogen zu Hintergrundvariablen).

# Chancen von CBT bzgl. Datenqualität

- **Beispiel:** PIAAC Workflow.



# Chancen von CBT bzgl. Datenqualität

- **Validität:** Durch neuartige Operationalisierungen verbesserte Repräsentation des Konstrukts bzw. Erfassung eines Konstrukts mit immanentem Technologiebezug (Konstruktvalidität).
- Operationalisierungen, die komplexe, dynamische, realistische Interaktionen erlauben und multimediale Bestandteile beinhalten können.
- Bsp.: Electronic Reading, Dynamisches Problemlösen, Computerbez. Kompetenzen





## Chancen von CBT bzgl. Datenquantität

- **Bearbeitungszeit:** Aufzeichnung von Reaktionszeiten zusätzlich zur Reaktion; Schätzung von individuellen Geschwindigkeitsparametern, die z.B. den Prozeduralisierungsgrad widerspiegeln (zusätzlich zu Fähigkeitsparametern).
- **Bearbeitungsschritte:** Prozessinformationen, d.h. im Prinzip jeder Klick mit der Maus, Tastatureingaben etc., sind erfassbar; mithilfe dieser Informationen können z.B. individuelle Unterschiede im Lösungsverhalten beschrieben und Antwortverhalten erklärt werden.

# Inhalt

- Computer-basiertes Testen (CBT) in Large Scale Assessments
- Chancen von CBT
- Herausforderungen von CBT
- Strategien zur Einführung von CBT in Large Scale Assessments
- Fallbeispiel I: PISA 2009 Electronic Reading
- Fallbeispiel II: PIAAC Computerisierung von Leseaufgaben
- Fazit

# Herausforderungen von CBT

- **(In)Äquivalenz:** Papier- und Computer-basierte Messungen sind hinsichtlich Reliabilität und Validität nicht vergleichbar; (Migration nach CBT birgt z.B. die Gefahr einer Trendunterbrechung in mehrzyklischen Studien, des Verlusts bisheriger Gütekriterien etc.).
- **Testfairness:** Konfundierende Einflüsse von konstruktirrelevanten Variablen, z.B. Computerliteralität, können Subgruppen benachteiligen (Tutorials, statistische Kontrolle).

# Herausforderungen von CBT

- **Validität:** Konstruktbestandeile, die bisher nur Papier-basiert praktikabel messbar sind,  
z.B. das Schreiben von reflektierenden (Kurz)Textantworten,  
Nebenrechnungen.
- **Infrastruktur:** Hohe Anfangsinvestitionen in Hard- und Software,  
technisches Personal.
- Unterstützung des (internationalen) kollaborativen Workflows durch  
Web-Portal (Steuerung von Aktivitäten unterschiedlicher Akteure)

# Inhalt

- Computer-basiertes Testen (CBT) in Large Scale Assessments
- Chancen von CBT
- Herausforderungen von CBT
- Strategien zur Einführung von CBT in Large Scale Assessments
- Fallbeispiel I: PISA 2009 Electronic Reading
- Fallbeispiel II: PIAAC Computerisierung von Leseaufgaben
- Fazit

## CBT in PISA 2000, 2006

- **PISA 2000:** Computer-Based Assessment von dynamischem Problemlösen (PS) als Teil der nationalen Ergänzung in Deutschland
- **PISA 2006:** Computer-Based Assessment of Science (CBAS) als internationale Option, 3 Teilnehmerstaaten

**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.

# CBT in PISA 2009

- **PISA 2009:** Electronic Reading Assessment (ERA) als internationale Option, 17 Teilnehmerstaaten
- Bildung einer gemeinsamen Skala aus papier- und computerbasierten Indikatoren



# CBT in PISA 2012

- **Terms of Reference der OECD für PISA 2012 (Paragraph 213):**

[..] Bidders are also requested to provide an assessment of the potential that **electronically delivered assessments** offer in terms of: (i) broadening the **coverage of the assessment domains** and improving the authenticity of the tasks; and (ii) achieving **efficiency gains** at both national and international levels. While the **establishment of trends** is essential to PISA, the PISA Governing Board . would be willing to discuss a one-time series break if that proved inevitable to advance the intended methodological innovations in test design and test delivery, such as for a move towards a computer-delivered assessment.

- **CBA-Strategie:** Einführung von CBT für Schwerpunktdomäne als Option bei Beibehaltung eines PBT-Teils (Trendfortsetzung), d.h. ERA 2009, CBT für Mathematics 2012



## CBT in PISA 2012, 2015

- **PISA 2012:**

- o Fortführung von ERA aus 2009 (Ziel einer gemeinsamen Skala: PBT + CBT).
- o Neue computer-basierte Skala für Mathematics (PBT + CBT).  
→ 30 Teilnehmerstaaten (internationale Optionen)
- o Kernassessment: Computer-basierte Skala für Problemlösen.

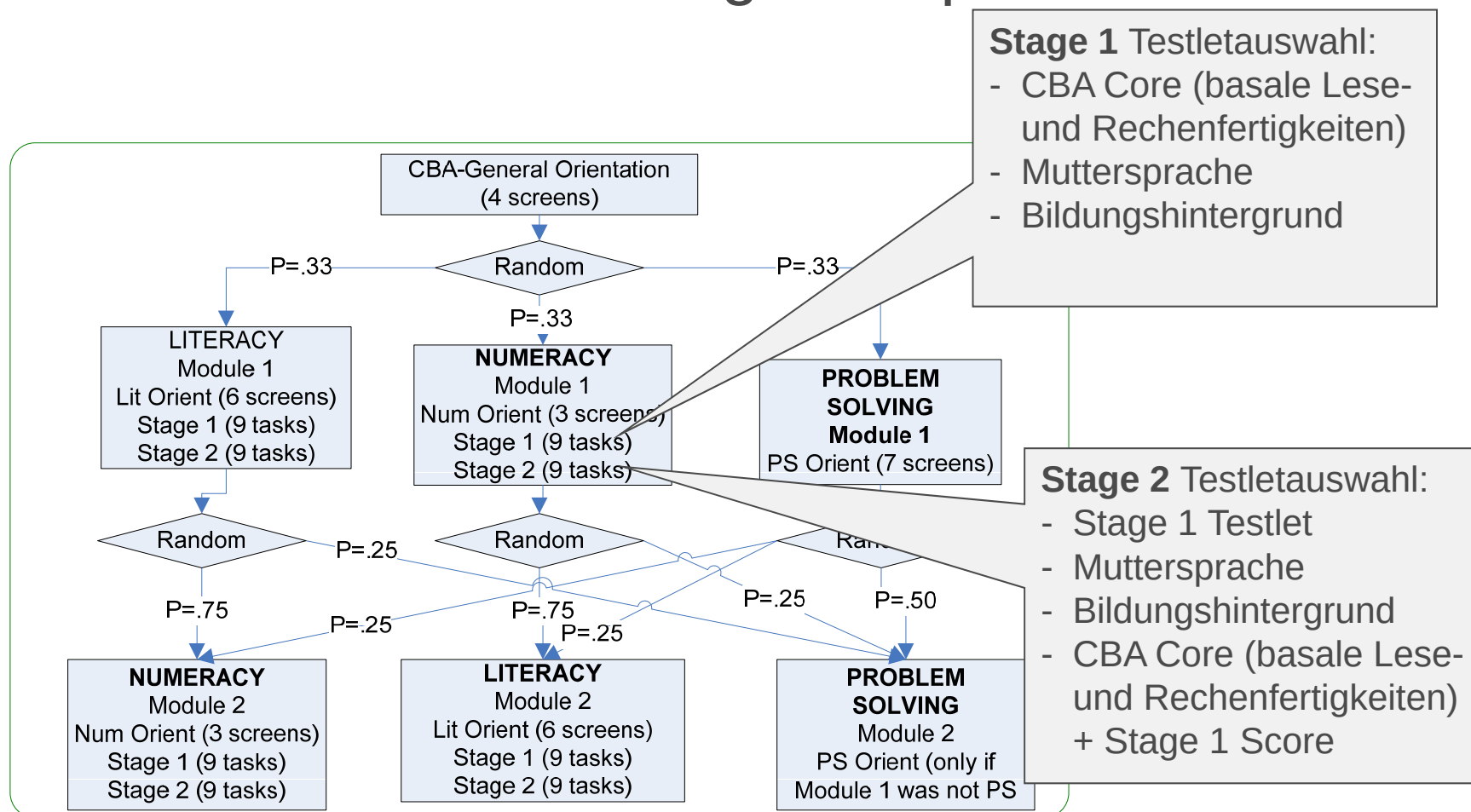
- **OECD-Strategie PISA 2015 (ToR, Paragraph 28):**

[..] PISA has the ambition to **move towards tests that are fully computer-delivered** and fully exploit the potential of computer-delivery in terms of tasks types.

## CBT in PIAAC

- **Innovationen, die computer-basiertes Assessment erfordern:**
  - o Literacy-Domäne: neben gedruckten auch digitale Texte (Simulation von WWW-Umgebungen).
  - o Neues Kompetenzkonstrukt *Problem Solving in Technology-Rich Environments* (Simulation von Computeranwendungen).
  - o Adaptives Testen in der Hauptstudie (*Two-stage Testing*)
  - o Antwortabhängiges Routing im Fragebogen zu Hintergrundvariablen.
- PBT für Teilnehmer mit niedriger Computerliteralität.

# CBT in PIAAC: Testdesign Haupttest



## CBT in NEPS

- **Kernassessment** : Längsschnittliches Erhebungsdesign zunächst Papier-basiert realisiert (Testhefte).
- **Mode Effect (ME)-Studien:** Flankierende Studien zur Prüfung der Äquivalenz von PBT und CBT, d.h. Sicherstellung der Vergleichbarkeit von Test-Scores zwischen CBT und PBT (*cross-mode equating*)
- **Herstellung von Vergleichbarkeit durch *linking*:**
  - o Stabilität der Skalen über den Assessmentmodi hinweg in einem längsschnittlichen Design (Lebensspanne)
  - o für unterschiedliche Kohorten u. Altersgruppen (z.B. differentielle ME)

# Inhalt

- Computer-basiertes Testen (CBT) in Large Scale Assessments
- Chancen von CBT
- Herausforderungen von CBT
- Strategien zur Einführung von CBT in Large Scale Assessments
- **Fallbeispiel I: PISA 2009 Electronic Reading**
- Fallbeispiel II: PIAAC Computerisierung von Leseaufgaben
- Fazit

# Fallbsp. I: PISA 2009 Lesen elektronischer Texte

- **Konzeptualisierung "elektronischer Texte"** in PISA 2009 als Bestandteil des Konstrukts Lesekompetenz:

"For the purposes of PISA [...] we are regarding electronic text as synonymous with hypertext: a text or texts with navigation tools and features that make possible and indeed even require nonsequential reading. Each reader constructs a "customised" text from the information encountered at the links he or she follows.,,  
(Reading Framework, p. 22)



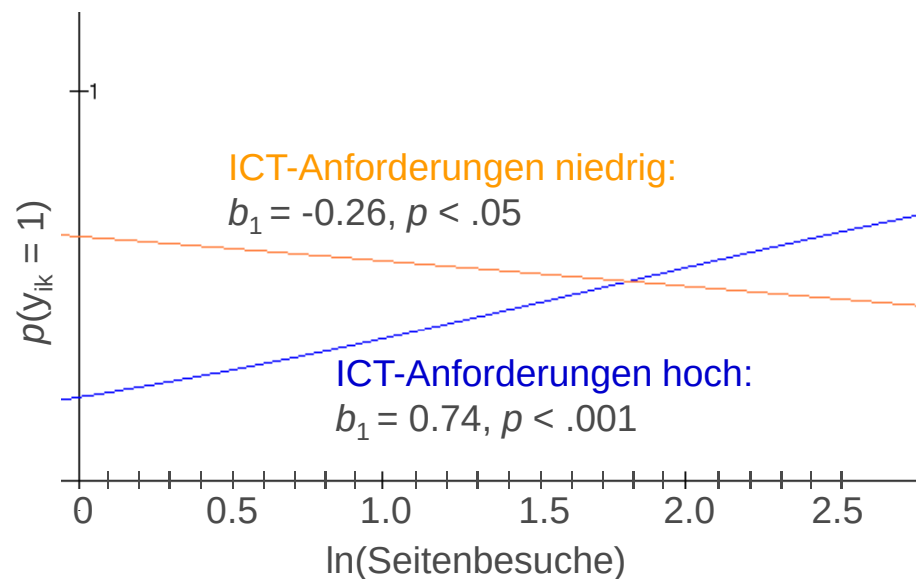
**PHILOSOPHERS' CAFÉ: Task 3** (E007001)  
Go to the Activities page. Find the Timeline. Use the information in the website to find when each of the three philosophers (Bentham, Confucius and Zeno) was born. Then use the drop down menus in the timeline to show when each philosopher was born.

## Fallbsp. I: PISA 2009 Lesen elektronischer Texte

- **Prozessbezogene Verhaltensdaten** mittels CBT, d.h.  
Bearbeitungszeiten und Bearbeitungsschritte (Navigationsverhalten,  
z.B. welche Hypertext-Seiten wurden besucht)
- **Vorhersage des Erfolgs bei der Aufgabenbearbeitung aus:**
  - o Prozessindikatoren und
  - o Aufgabeneigenschaften sowie der
  - o Interaktion von Aufgabeneigenschaften und Prozessindikatoren(Naumann, 2010)

# Fallbsp. I: PISA 2009 Lesen elektronischer Texte

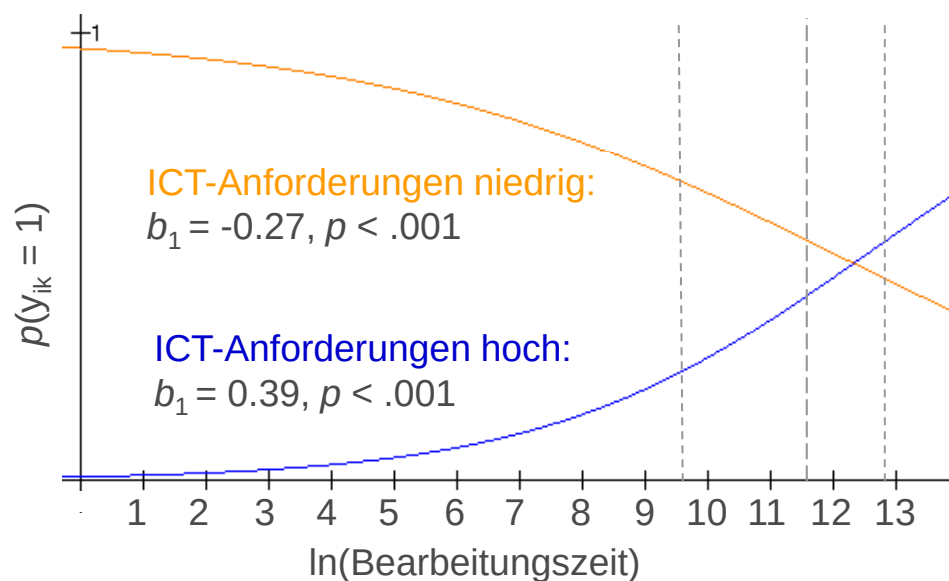
- **Erklärung der Lösungswahrscheinlichkeit:** Effekt des Prozessindikators ‚Anzahl besuchter relevanter Seiten‘ in Abhängigkeit vom Aufgabenmerkmal ‚ICT-Anforderungen‘ (Naumann, 2010)





## Fallbsp. I: PISA 2009 Lesen elektronischer Texte

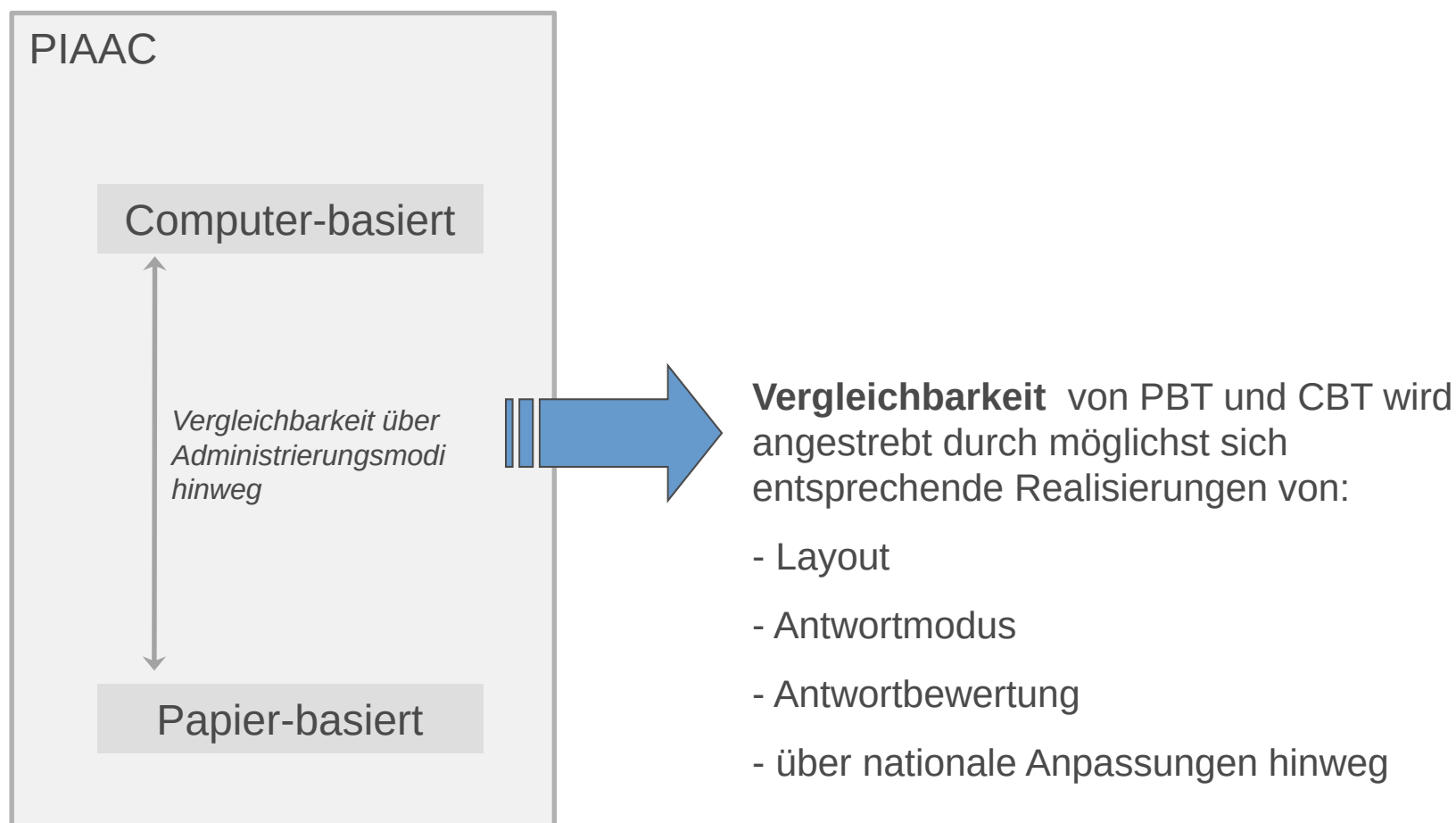
- **Erklärung der Lösungswahrscheinlichkeit:** Effekt des Prozessindikators ‚Bearbeitungszeit‘ in Abhängigkeit vom Aufgabenmerkmal ‚ICT-Anforderungen‘ (Naumann, 2010)



# Inhalt

- Computer-basiertes Testen (CBT) in Large Scale Assessments
- Chancen von CBT
- Herausforderungen von CBT
- Strategien zur Einführung von CBT in Large Scale Assessments
- Fallbeispiel I: PISA 2009 Electronic Reading
- Fallbeispiel II: PIAAC Computerisierung von Leseaufgaben
- Fazit

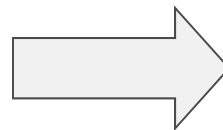
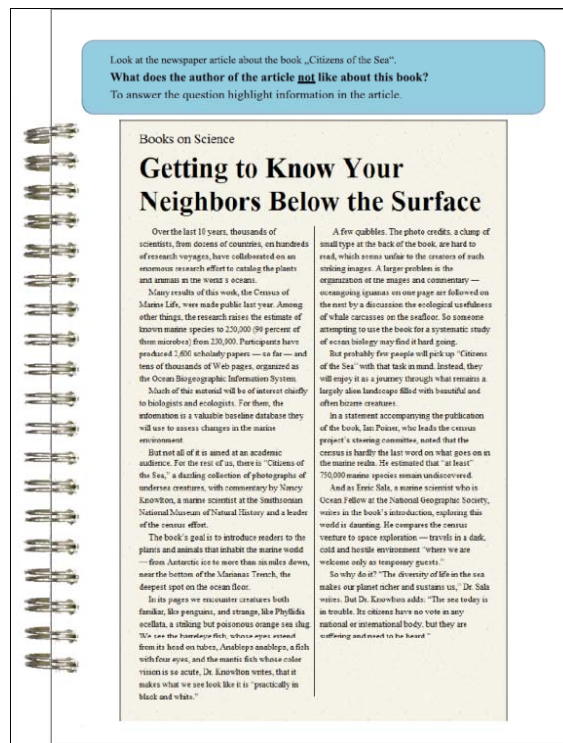
## PIAAC: PBA → CBA



# PBA → CBA: Layout

- **Layout** von Literacy-Aufgaben
  - o PBA: Hochformat
  - o CBA: Querformat, d.h. für CBA in der Höhe geringerer Platz

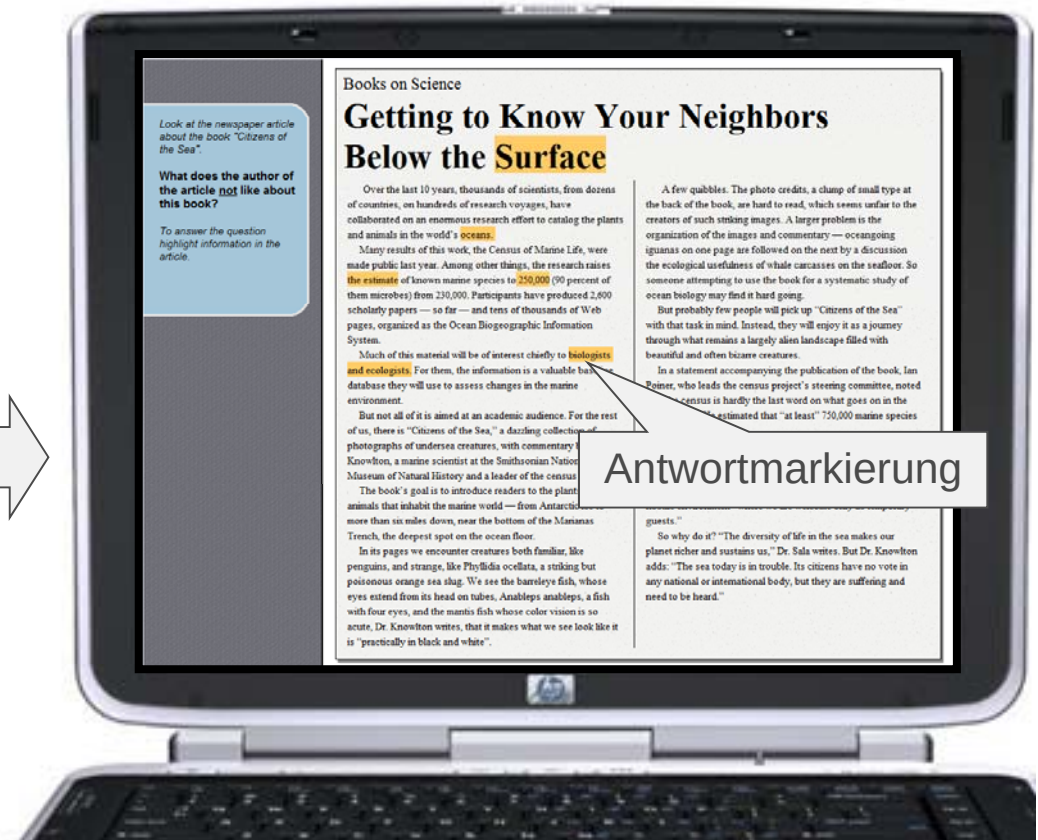
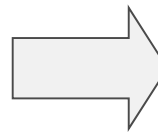
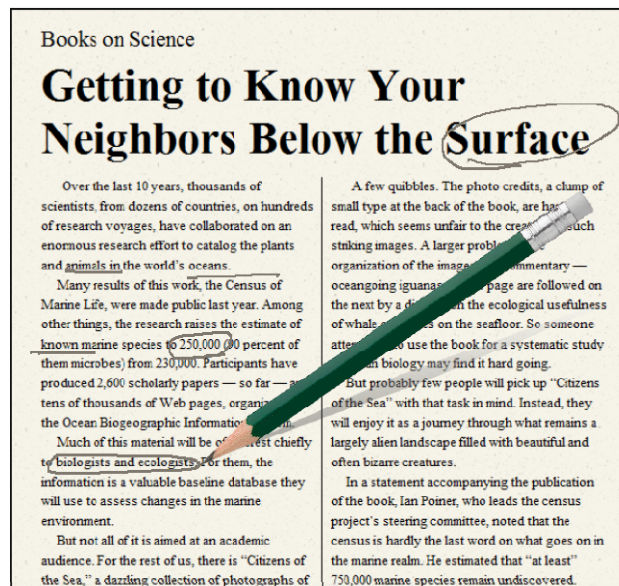
PBA



# PBA → CBA: Antwortmodus

- **Antwortmodus** von Literacy-Aufgaben
  - o PBA: freie Antwort, Einkreisung von Textteilen
  - o CBA: Markieren des Textes durch Klicken und Ziehen per Maus

PBA



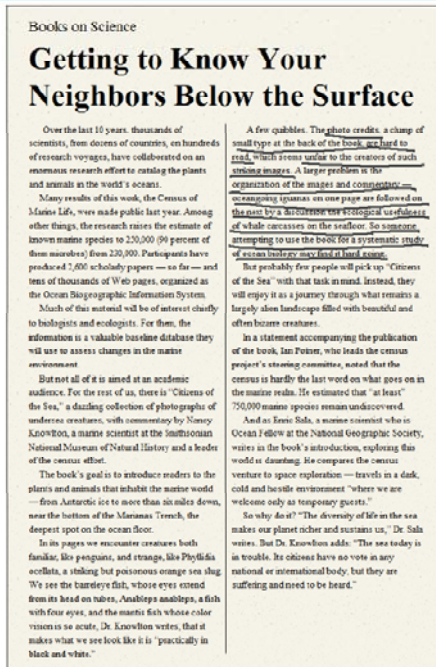


# PBA → CBA: Antwortbewertung

- **Antwortbewertung von Literacy-Aufgaben**
  - o PBA: Beurteilung der Antwort durch geschulte menschliche Beurteiler
  - o CBA: Automatische Beurteilung durch Vergleich der Antwortmarkierung mit dem definierten Minimal- und Maximalbereich für eine richtige Antwort

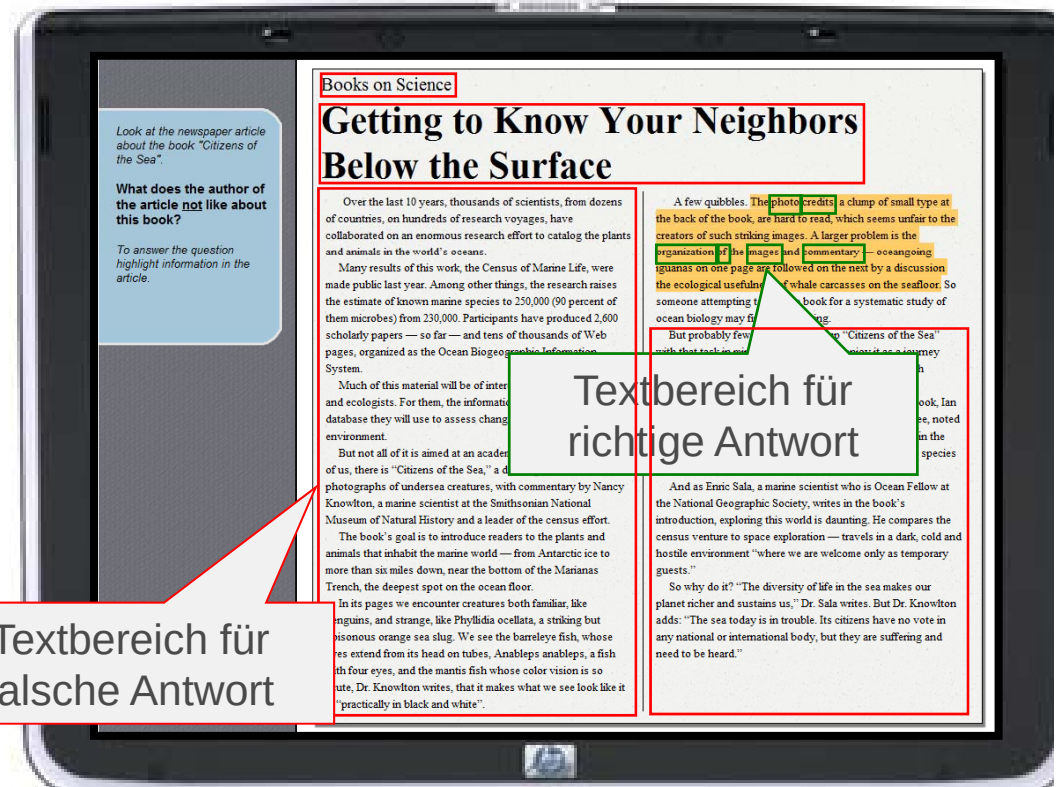
PBA

Look at the newspaper article about the book „Citizens of the Sea”.  
What does the author of the article not like about this book?  
To answer the question highlight information in the article.



Wiesbac

Textbereich für  
falsche Antwort

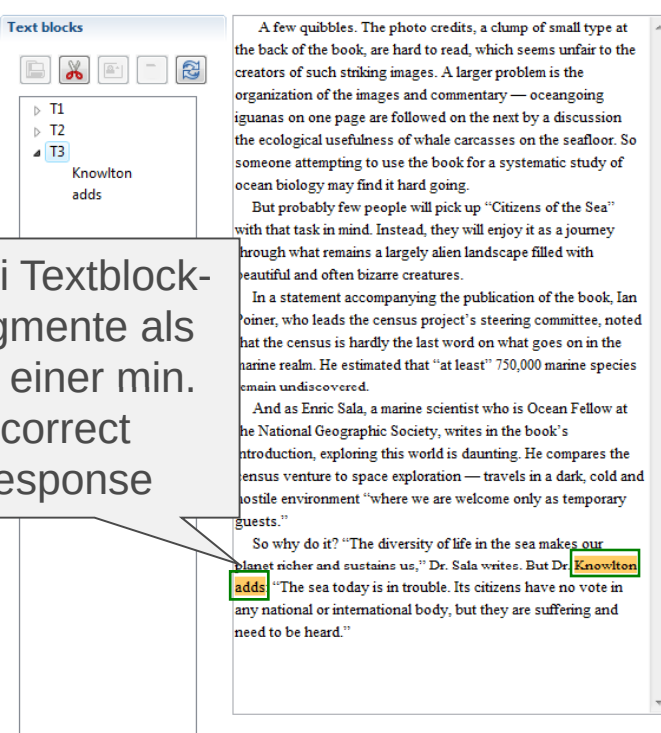


# PBA → CBA: Nationale Anpassung

- **Nationale Anpassung** der Antwortbewertung:
  - o PBA: Schulung nationaler Beurteilerteams mit Hilfe zentral geschulter Beurteiler
  - o CBA: Anpassung der Textbereiche für richtige Antwort im Laufe der Item-entwicklung durch nationale Experten mithilfe eines speziellen graph. Editors

CBA – source EN

CBA – target DE



**Text blocks**

- T1
- T2
- T3

Knowlton  
adds

A few quibbles. The photo credits, a clump of small type at the back of the book, are hard to read, which seems unfair to the creators of such striking images. A larger problem is the organization of the images and commentary — oceangoing iguanas on one page are followed on the next by a discussion the ecological usefulness of whale carcasses on the seafloor. So someone attempting to use the book for a systematic study of ocean biology may find it hard going.

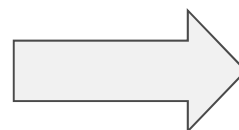
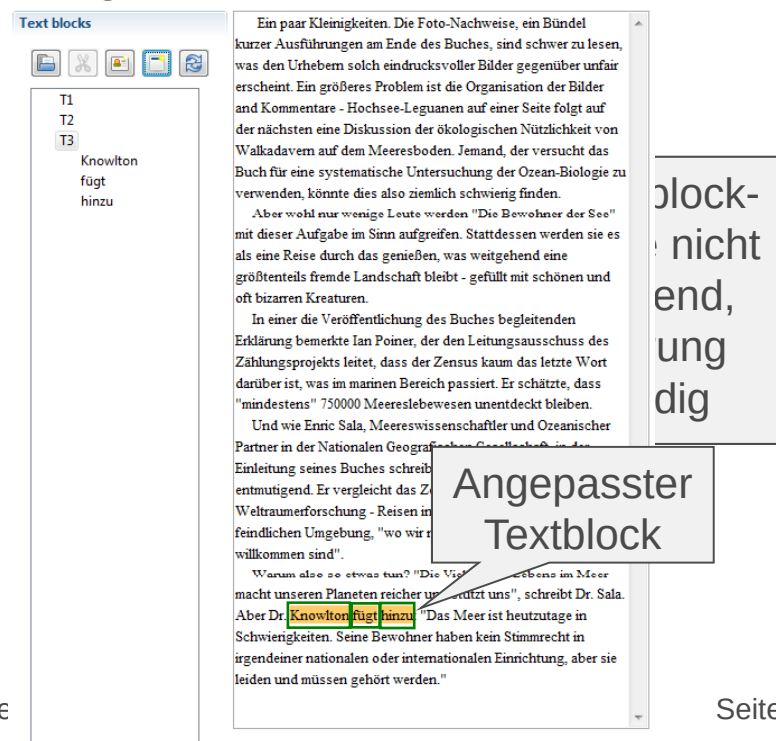
But probably few people will pick up "Citizens of the Sea" with that task in mind. Instead, they will enjoy it as a journey through what remains a largely alien landscape filled with beautiful and often bizarre creatures.

In a statement accompanying the publication of the book, Ian Poiner, who leads the census project's steering committee, noted that the census is hardly the last word on what goes on in the marine realm. He estimated that "at least" 750,000 marine species remain undiscovered.

And as Enric Sala, a marine scientist who is Ocean Fellow at the National Geographic Society, writes in the book's introduction, exploring this world is daunting. He compares the census venture to space exploration — travels in a dark, cold and hostile environment "where we are welcome only as temporary guests."

So why do it? "The diversity of life in the sea makes our planet richer and sustains us," Dr. Sala writes. But Dr. Knowlton adds "The sea today is in trouble. Its citizens have no vote in any national or international body, but they are suffering and need to be heard."

**Zwei Textblock-fragmente als Teil einer min. correct response**

**Text blocks**

- T1
- T2
- T3

Knowlton  
fügt  
hinzu

Ein paar Kleinigkeiten. Die Foto-Nachweise, ein Bündel kurzer Ausführungen am Ende des Buches, sind schwer zu lesen, was den Urhebern solch eindrucksvoller Bilder gegenüber unfair erscheint. Ein größeres Problem ist die Organisation der Bilder und Kommentare - Hochsee-Leguanen auf einer Seite folgt auf der nächsten eine Diskussion der ökologischen Nützlichkeit von Walkadavern auf dem Meeresboden. Jemand, der versucht das Buch für eine systematische Untersuchung der Ozean-Biologie zu verwenden, könnte dies also ziemlich schwierig finden.

Aber wohl nur wenige Leute werden "Die Bewohner der See" mit dieser Aufgabe im Sinn aufgreifen. Stattdessen werden sie es als eine Reise durch das genießen, was weitgehend eine größtenteils fremde Landschaft bleibt - gefüllt mit schönen und oft bizarren Kreaturen.

In einer die Veröffentlichung des Buches begleitenden Erklärung bemerkte Ian Poiner, der den Leitungsausschuss des Zählungsprojekts leitet, dass der Zensus kaum das letzte Wort darüber ist, was im marinen Bereich passiert. Er schätzte, dass "mindestens" 750000 Meereslebewesen unentdeckt bleiben.

Und wie Enric Sala, Meereswissenschaftler und Ozeanischer Partner in der Nationalen Geographischen Gesellschaft, in der Einleitung seines Buches schreibt, ist die Erkundung dieses Weltraums entmutigend. Er vergleicht das Zählungsprojekt mit Weltraumforschung - Reisen in eine feindlichen Umgebung, "wo wir nur als temporäre Gäste willkommen sind".

Warum also so etwas tun? "Die Vielfalt des Lebens im Meer macht unseren Planeten reicher und nützt uns," schreibt Dr. Sala. Aber Dr. Knowlton fügt hinzu "Das Meer ist heutzutage in Schwierigkeiten. Seine Bewohner haben kein Stimmrecht in irgendeiner nationalen oder internationalen Einrichtung, aber sie leiden und müssen gehört werden."

**Angepasster Textblock**

**block-nicht end, ung dig**

# PBA → CBA: Verhaltensdaten

- **Ergebnisse** von Testaufgaben
  - o PBA: Auswertung des Endergebnisses
  - o CBA: Auswertung des Endergebnisses (automatisch)
    - + zusätzlich Zeitdaten und detaillierte Daten

- Ergebnis: richtig/falsch
- Reaktionszeit
- Bearbeitungszeit
- Anzahl der Interaktionen

- Markierter Text nach Beendigung des Items

- Snapshot: vollständige Beschreibung des Stimuluszustandes

```
instance" xmlns:cbascoringresultmm="http://cbascoringresultmm/1.0" xmlns:snapshot="http://snapshot.softcon.de" name="Task1"
result="true" nbHits="1" hitWeight="1" nbMisses="0" missWeight="0" execTime="44" execTimeTotal="44" reactionTime="15" reactionTimeTotal="15"
nbInteractions="5" nbInteractionsTotal="5">
<hitList name="Hit1" weight="1" />
- <highlightOverview>
- <textField id="cbaTextField_6_14061315727487">
  <highlightedFragment startPosition="21" endPosition="365" text="The photo credits, a clip of Simon type at the back of the book, are
  hard to read, which seems unfair to the creators of such striking images. A larger problem is the organization of the images and
  commentary — oceangoing iguanas on one page are followed on the next by a discussion the ecological usefulness of whale
  carcasses on the seafloor." />
</textField>
</highlightOverview>
- <snapshotOverview version="cbaComponentStateStorage_14060918929221">
  <componentStateList xsi:type="snapshot:CBAClipboardState" id="cbaClipboard_1406107313415">
  <componentStateList xsi:type="snapshot:CBAPageManagerState" id="cbaPageManager_14060899096463">
    firstPage="cbaPage_0_14061116697094" />
  <componentStateList xsi:type="snapshot:CBARichTextFieldState" id="cbaTextField_6_14061315727487">
    selectedTextBlocks="2T1,2T2,0T3,0T4" highlightedCharacters="21-365" highlighted="true" />
  <componentStateList xsi:type="snapshot:ScoringManagerState" id="cbaScoringManager_14060899096463" userInteractions="5">
    execTime="44" reactionTime="15" />
</snapshotOverview>
```



# PBA → CBA: Verhaltensdaten

o CBA: + zusätzlich detaillierte Daten über Interaktionen

- <logEntry xsi:type="cbaloggingmodel:LogEntryTimeStamp" timeStamp="2011-01-11T14:08:28.084+0100">
- <logEntry xsi:type="cbaloggingmodel:HighlightTextFieldLogEntry" id="cbaTextField\_6\_1406131572" >
  - <highlightedFragment startPosition="1182" endPosition="1192" text="exploration" />
- <logEntry xsi:type="cbaloggingmodel:LogEntryTimeStamp" timeStamp="2011-01-11T14:08:44.201+0100">
- <logEntry xsi:type="cbaloggingmodel:HighlightTextFieldLogEntry" id="cbaTextField\_6\_1406131572" >
  - <highlightedFragment startPosition="21" endPosition="365" text="The photo credits, a clump of small type at the back of the book, are hard to read, which seems unfair to the creators of such striking images. A larger problem is the organization of the images and commentary — oceangoing iguanas on one page are followed on the next by a discussion the ecological usefulness of whale carcasses on the seafloor." />
- <logEntry xsi:type="cbaloggingmodel:LogEntryTimeStamp" timeStamp="2011-01-11T14:08:49.221+0100">
- <logEntry xsi:type="cbaloggingmodel:HighlightTextFieldLogEntry" id="cbaTextField\_6\_1406131572" >
  - <highlightedFragment startPosition="21" endPosition="365" text="The photo credits, a clump of small type at the back of the book, are hard to read, which seems unfair to the creators of such striking images. A larger problem is the organization of the images and commentary — oceangoing iguanas on one page are followed on the next by a discussion the ecological usefulness of whale carcasses on the seafloor." />
  - <highlightedFragment startPosition="1297" endPosition="1309" text="So why do it?" />
- <logEntry xsi:type="cbaloggingmodel:LogEntryTimeStamp" timeStamp="2011-01-11T14:08:50.924+0100">
- <logEntry xsi:type="cbaloggingmodel:HighlightTextFieldLogEntry" id="cbaTextField\_6\_1406131572" >
  - <highlightedFragment startPosition="21" endPosition="365" text="The photo credits, a clump of small type at the back of the book, are hard to read, which seems unfair to the creators of such striking images. A larger problem is the organization of the images and commentary — oceangoing iguanas on one page are followed on the next by a discussion the ecological usefulness of whale carcasses on the seafloor." />

1. Zustand: „exploration“ wurde markiert (14:08:28)

2. Zustand: „exploration“ wurde demarkiert und markiert wurde „The photo...“ (14:08:44)

3. Zustand: neben „The photo...“ wurde zusätzlich „So why do it?“ markiert (14:08:49)

4. Zustand: „So why do it?“ wieder demarkiert, Markierung von „The photo...“ (14:08:50)

# Inhalt

- Computer-basiertes Testen (CBT) in Large Scale Assessments
- Chancen von CBT
- Herausforderungen von CBT
- Strategien zur Einführung von CBT in Large Scale Assessments
- Fallbeispiel I: PISA 2009 Electronic Reading
- Fallbeispiel II: PIAAC Computerisierung von Leseaufgaben
- Fazit

## Fazit

- Computer-Basiertes Testen verspricht eine verbesserte Datenqualität und ermöglicht die Sammlung von zusätzlichen Verhaltensdaten.
- Die Umstellung von PBT nach CBT erfordert die Einrichtung einer entsprechenden Infrastruktur und muss psychometrisch sowie durch Domänenexperten begleitet werden.
- CBT ermöglicht erweiterte Erkenntnisse über die Struktur eines Konstrukts (z.B. "Kompetenz beim Lesen elektronischer Texte,") und entsprechender interindividueller Unterschiede.
- CBT wird zukünftig einen noch breiteren Raum in LSAs einnehmen.

Ende

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

Herzlichen Dank  
Johannes Naumann und Gabriele Gissler für  
für die Überlassung bzw. Vorbereitung einiger Materialien.