



# Valide Messung akademischer Kompetenzen von Studierenden – Methodische Entwicklungen und Innovationen

**Olaf Köller**

IPN - Leibniz-Institut für die Pädagogik  
der Naturwissenschaften und Mathematik



## Überblick: 4 Punkte



- Woher wissen wir, was Studierende lernen sollen? Vor dem Test kommt die Präzisierung des Konstrukts
- Welches Aufgabenformat für welches Konstrukt?
- Worin ich vielleicht investieren würde
- Ein Plädoyer für intensive Validierungsstudien



# Was soll gelernt werden? Prosa



## **Leitbild universitärer Lehre der Universität Hamburg**

(Ziel universitärer Lehre ist es, **Bildung durch Wissenschaft** zu ermöglichen. Das schließt die Aufgabe ein, alle Voraussetzungen dafür zu schaffen, dass die Studierenden hohe wissenschaftliche Kompetenz erwerben, ihre Fähigkeiten selbsttätig entfalten und sich als mündige Mitglieder der Gesellschaft weiterentwickeln können, die bereit und in der Lage sind, an deren sozial und ökologisch nachhaltiger, demokratischer und friedlicher Gestaltung maßgeblich mitzuwirken und für ihre Zukunftsfähigkeit Verantwortung zu übernehmen. Grundlage der universitären Lehre ist das Humboldt'sche Bildungsideal der Einheit von Forschung und Lehre. Lernendes Forschen, lebenslanges Lernen und die argumentative Verständigung auf wissenschaftlicher Grundlage sind wesentliche Merkmale dieser Lehre.



# Was soll gelernt werden? Prosa



Als Charakteristikum universitärer Lehre betont die **Goethe-Universität** Wissenschaftlichkeit, die in der didaktisch-methodischen Konzeptualisierung des forschenden Lernens ihren Ausdruck findet. Studierende erfahren Wissen hierbei als etwas Offenes, im Werden Begriffenes – eben als Forschung. Sie werden so zu Fragenden und zu Forschenden, die von Anfang an befähigt werden sollen, Wissen und seine Entstehungsbedingungen kritisch zu hinterfragen.

Forschendes Lernen heißt Identifizieren und Kontextualisieren von Problemlagen, beinhaltet stets Skepsis und die Fähigkeit zu distanzierter Betrachtung, bildet Selbständigkeit und methodisch angeleitete Urteilsfähigkeit aus... Universitäre Lehre an der Goethe-Universität vermittelt Fach- und Methodenwissen, das die Grundlage für den Erwerb wissenschaftlicher Kernkompetenzen ist. Zugleich werden die Studierenden in die Lage versetzt, konkret am wissenschaftlichen Gegenstand überfachliche Kompetenzen einzuüben. Dazu gehört nicht zuletzt die Befähigung, sich selbständig Wissen und Informationen anzueignen. Die Absolvierung eines Studiums eröffnet somit sowohl den Weg in die Forschung als auch zu akademisch-qualifizierter Berufstätigkeit, was bei einigen Studiengängen die konkrete Berufsvorbereitung bedeuten kann.



# Was soll gemessen werden? Probleme in der Ausgangslage



- In der Regel wird ein Fach studiert und nicht auf einen Beruf vorbereitet (Ausnahmen: Lehramt, Jura, Medizin, BWL)
- Berufsnahen Tätigkeiten jenseits der Forschungspraxis sind vor allem in der Bachelor-Phase seltene Ereignisse
- Es werden vor allem Inhalte (Wissen) vermittelt
- Gibt es einen Kanon innerhalb des Faches?

▪



# Beispiel Naturwissenschaften – Universitäres Wissen



- Deklaratives Wissen: Definiere! Beschreibe! Was verstehen wir unter?
- Prozedurales Wissen: Wähle aus! Führe durch! Interpretiere! Experimentiere!
- Schematisches Wissen: Erkläre! Prognostiziere! Integriere!
- Strategisches Wissen/Problemlösekompetenzen: Fragen stellen, Wissenstransfer, Anwenden
  
- Wichtig: Wissensfacetten korrelieren in der Regel sehr hoch miteinander ( $r > .70$ )

Quelle: Shavelson, 2005



# Was wird an der Universität in den Naturwissenschaften gelernt?

Erinnere Blooms Taxonomie



1. Factual Knowledge: remember and recall factual information
2. Comprehension: demonstrate understanding of ideas, concepts
3. Application: apply comprehension to unfamiliar situations
4. Analysis: break down concepts into parts
5. Synthesis: transform, combine ideas to create something new
6. Evaluation: think critically about and defend a position

Quelle: Shavelson, 2010



# Projekt KiL

Messung professioneller Kompetenzen in  
mathematischen und naturwissenschaftlichen  
Lehramtsstudiengängen

C | A | U

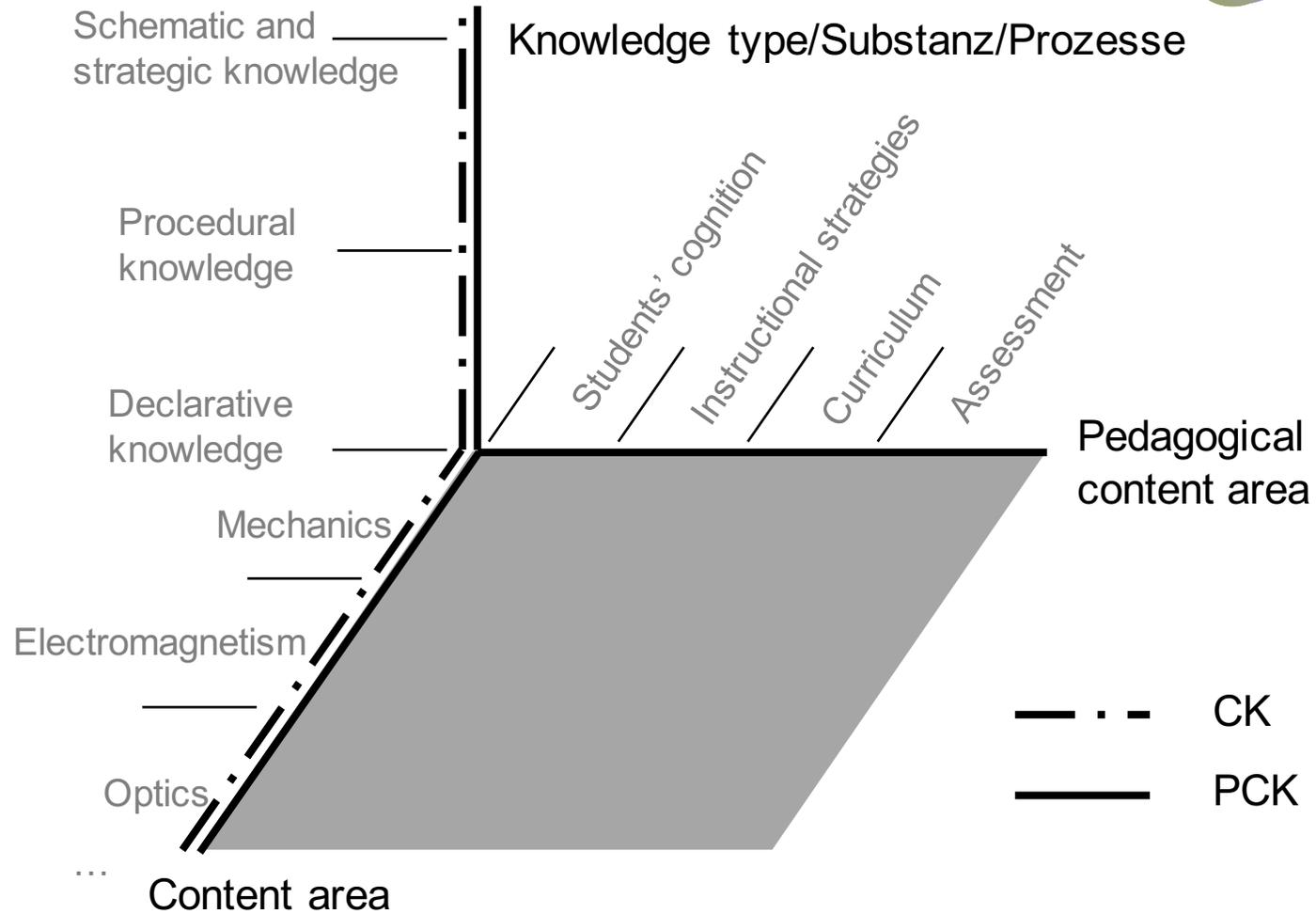
Christian-Albrechts-Universität zu Kiel

*KiL*

*Leibniz*  
Leibniz-Gemeinschaft



# KiL-Strukturmodell in Physik



(Shavelson, 2005; Grehn & Krause, 2007; Dorn & Bader, 2008; Magnusson, Krajcik & Borko, 1999; vgl. Tepner et al., 2012)



# Beispielaufgabe Physik – Prozedurales CK



Aus einer dünnen Sperrholzplatte wird eine quadratische Platte der Kantenlänge  $a$  gesägt. Ihr Trägheitsmoment bei Drehung um die zur Platte senkrechte Schwerpunktsachse betrage  $J$ .  
Wie groß wäre das Trägheitsmoment der Platte, wenn ihre Kantenlänge doppelt so groß wäre?

- A         $2 J$
- B         $4 J$
- C         $8 J$
- D         $16 J$**



# Implikationen für Aufgabenformate



- In der Regel werden wir in vielen Studiengängen mit den üblichen Antwortformaten MC, CMC, Kurzantwort und ausführliche Antwort auskommen
- MC- und CMC-Items sind viel besser als ihr Ruf (wenn sie gut konstruiert sind; vgl. Lindner, Strobel & Köller, ZfPP, 2015)
- Vignetten werden überschätzt
- Ebenso Computersimulationen, wie sie aus dem Paradigma des komplexen Problemlösens abgeleitet wurden (sehen schön aus, haben aber mehr psychometrische Probleme als Nutzen)
- Will man berufsnah testen (z. B. Ingenieurwissenschaften), so dürften realitätsnahe Simulationen die Methode der Wahl sein (aber: aufwendig in der Entwicklung und oft mit mäßigen psychometrischen Kennwerten)



# Worin ich vielleicht langfristig investieren würde

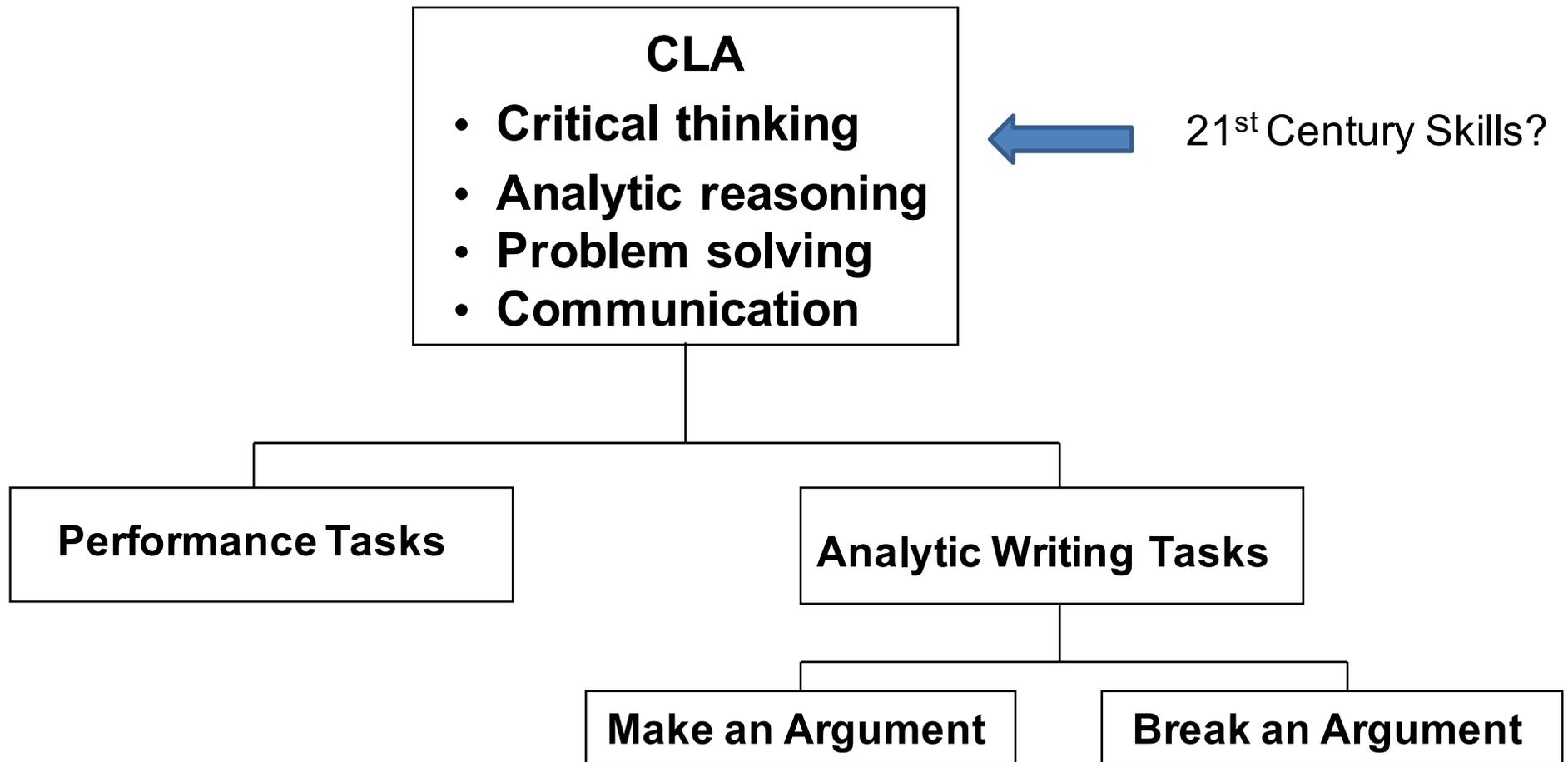


Richard J. Shavelson, Stanford University

THE COLLEGIATE  
LEARNING ASSESSMENT



# Worin ich vielleicht investieren würde (Shavelson, 2008)





# CLA (Shavelson, 2008)



## Task Format

- Real-world problem
- Holistic, complex problem
- Information that may:
  - Be relevant or irrelevant to problem
  - Be reliable or unreliable
  - Lead to judgmental errors (e.g., correlation not causality, representativeness)

## Response Format

- Make recommendation or decision, reach a conclusion, or solve a problem
- Minimally structured to support line of argument
- Written and not selected
- Requires evaluation of possible alternatives



# CLA-Scoring (Shavelson, 2008)



- Analytic reasoning and evaluation
  - Identifies strengths and weaknesses of alternative arguments
  - Accurately judges quality of evidence avoiding unreliable, invalid, and erroneous information
- Problem solving
  - Provides decision and solid rationale based on credible evidence
  - Acknowledges uncertainty and need for further information
- Writing effectiveness
  - Organizes "advice" in logically cohesive and easy-to-follow way
  - Provides valid and comprehensive details supporting each argument and information source on which based
- Writing mechanics
  - Writes well constructed complex sentences
  - Shows outstanding control of grammar conventions
  - Demonstrates adept use of vocabulary



# Herausforderungen von CLA



- Computer-basiertes automatisches Kodieren von Texten
- Entwicklung von fachspezifischen Aufgaben, die in vertretbarer Zeit zu bearbeiten sind und gemeinsame (und nicht nur aufgabenspezifische) Varianz erzeugen
- Validierung dieser Aufgaben
- Gibt es überhaupt systematische Lerngelegenheiten für solche Aufgaben im Studium?
- Welchen Mehrwert haben sie gegenüber den traditionellen Formaten?



# Letzter Punkt: Validierung (Messick, 1995)



Validierung bedeutet die kontinuierliche, empirische Evaluation von Messwertinterpretationen sowie der Konsequenzen, die sich aus diesen Interpretationen ergeben





# Ein paar steile Thesen zur Validierung



- Viele Projekte zur Testkonstruktion in den Kontexten Schule und Hochschule kommen kaum über die inhaltliche und strukturelle Validierung hinaus
- Die Passung eines IRT-Modells sagt nichts über die Validität eines Tests aus, eine schöne Wright Map im übrigen auch nicht
- Mehrdimensionale Modelle passen (meistens) besser als eindimensionale, sagen aber auch nicht viel über die Validität eines Tests bzw. seiner Testwerte aus
- Testbearbeitungsprozesse werden weitgehend außer Acht gelassen, auch fehlen in der Regel gute Modelle schwierigkeitsgenerierender Merkmale
- Breite Validierungen im Sinne der Analyse nomologischer Netzwerke sind seltene Ereignisse
- Externe Validierungen und Validierungen der Konsequenzen von Testergebnissen (Consequential Validity) finden im Schul- und Hochschulbereich viel zu selten statt



# Ein Plädoyer für Testbearbeitungs-/ Prozessanalysen



- Einfache technische Möglichkeiten beim CBA
- Log-File-Analysen
- Blickbewegungsanalysen, die Auskunft über Informationsverarbeitungsprozesse bei der Testbearbeitung geben

▪



# Nutzung von Eyetracking bei der Testvalidierung\*



## Beispielaufgabe

Das Wetter kann an verschiedenen Orten sehr unterschiedlich sein. In Ort A sind Wolken am Himmel und die Temperatur liegt bei +5°C. In Ort B sind keine Wolken und die Temperatur ist -5°C. Ort C hat Wolken und eine Temperatur von -5°C. In Ort D sind keine Wolken am Himmel und die Temperatur ist 5°C.

Kurzer  
Aufgabenstamm

In welchem Ort wird es voraussichtlich schneien?

- (A) Ort A
- (B) Ort B
- (C) Ort C
- (D) Ort D

Vier  
Antwortoptionen

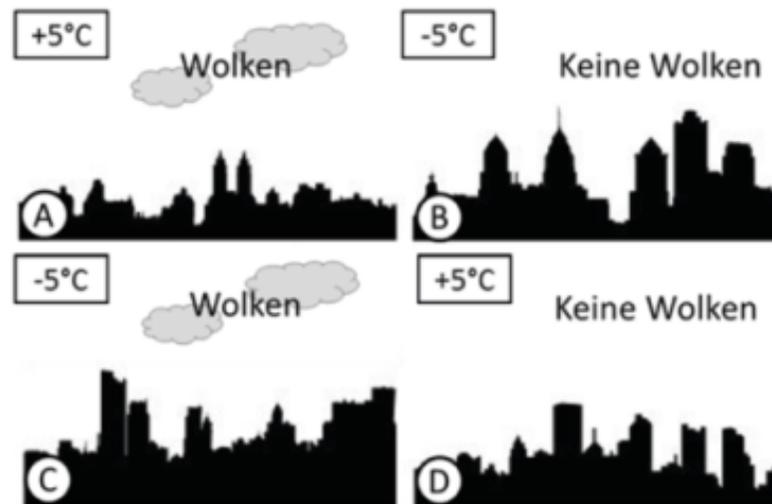
Frage

\*Lindner, M.A., Eitel, A., Strobel, B.; & Köller, O. (submitted). Identifying Processes Underlying the Multimedia-Effect in Testing: An Eye Movement Analysis.

# Beispielaufgabe

Das Wetter kann an verschiedenen Orten sehr unterschiedlich sein. In Ort A sind Wolken am Himmel und die Temperatur liegt bei +5°C. In Ort B sind keine Wolken und die Temperatur ist -5°C. Ort C hat Wolken und eine Temperatur von -5°C. In Ort D sind keine Wolken am Himmel und die Temperatur ist 5°C.

Kurzer  
Aufgabenstamm



Veranschaulichendes  
Bild

In welchem Ort wird es voraussichtlich schneien?

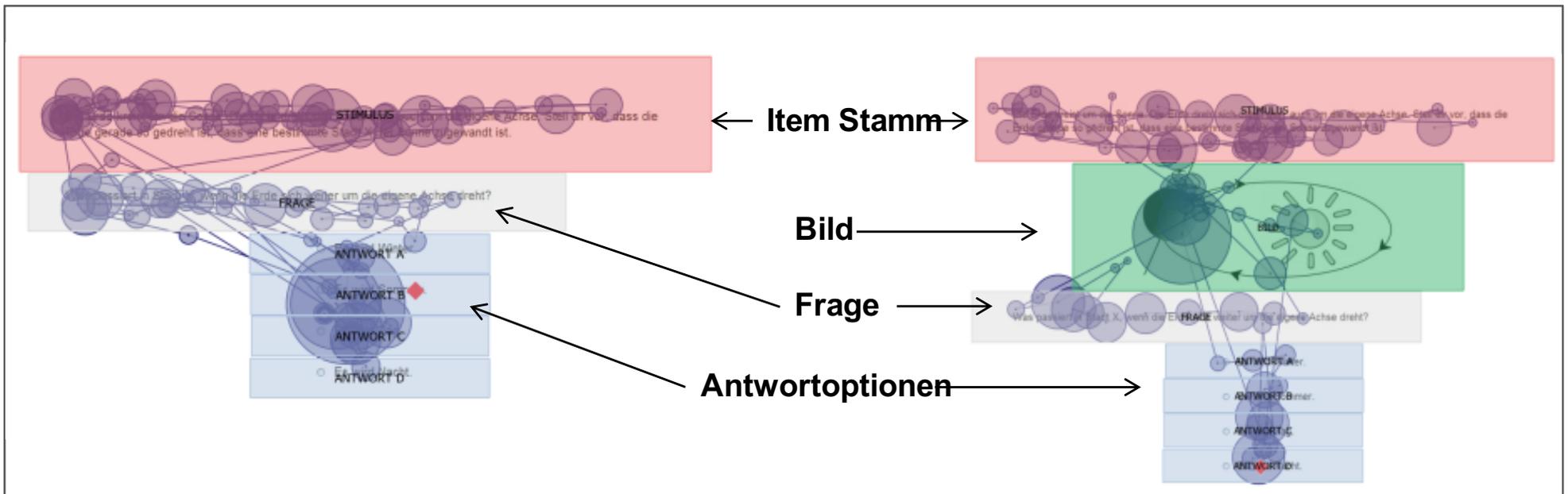
Frage

- (A) Ort A
- (B) Ort B
- (C) Ort C
- (D) Ort D

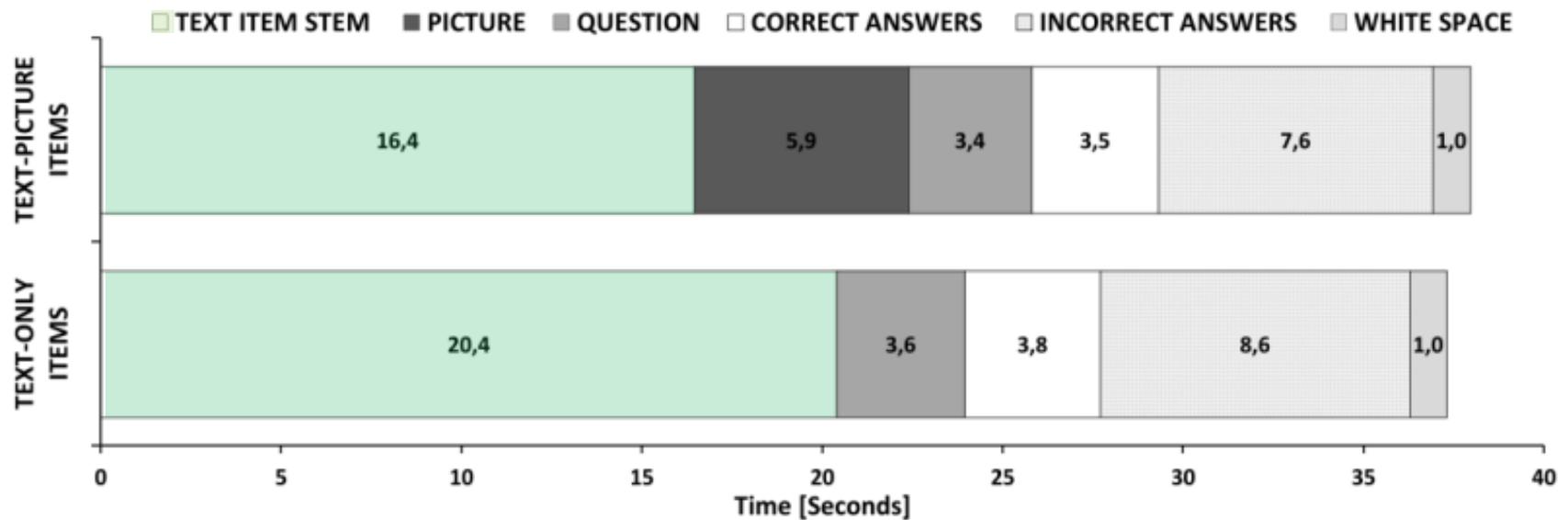
Vier  
Antwortoptionen

Itemstamm: nur Text

Itemstamm: Text und  
veranschaulichendes Bild



SMI RED 120Hz  
(Remote)



Das Wetter kann an verschiedenen Orten sehr unterschiedlich sein. In Ort A sind Wolken am Himmel und die Temperatur liegt bei +5°C. In Ort B sind keine Wolken und die Temperatur ist -5°C. Ort C hat Wolken und eine Temperatur von -5°C. In Ort D sind keine Wolken am Himmel und die Temperatur ist 5°C.

In welchem Ort wird es voraussichtlich schneien?

Ort A  
 Ort B  
 Ort C  
 Ort D

$$M_{(\text{TEXT})} = 20.39$$

$$M_{(\text{T-BILD})} = 16.45 \quad M_{\text{Diff}} = -3.94; SE_{\text{Diff}} = 0.31$$

$p < .001, d = .44$



# Nomologische Netzwerke



- Eigentlich das typische Vorgehen bei der Testvalidierung
- Test sollte hoch mit konstruktverwandten anderen Tests korrelieren (konvergente Validität) und gering mit konstruktfernen Tests (diskriminante Validität)
- Auswahl der konstrukt nahen und –fernen Tests sollte theoriebasiert geschehen

▪



# Beispiel: Ackermans PPIK (1996)



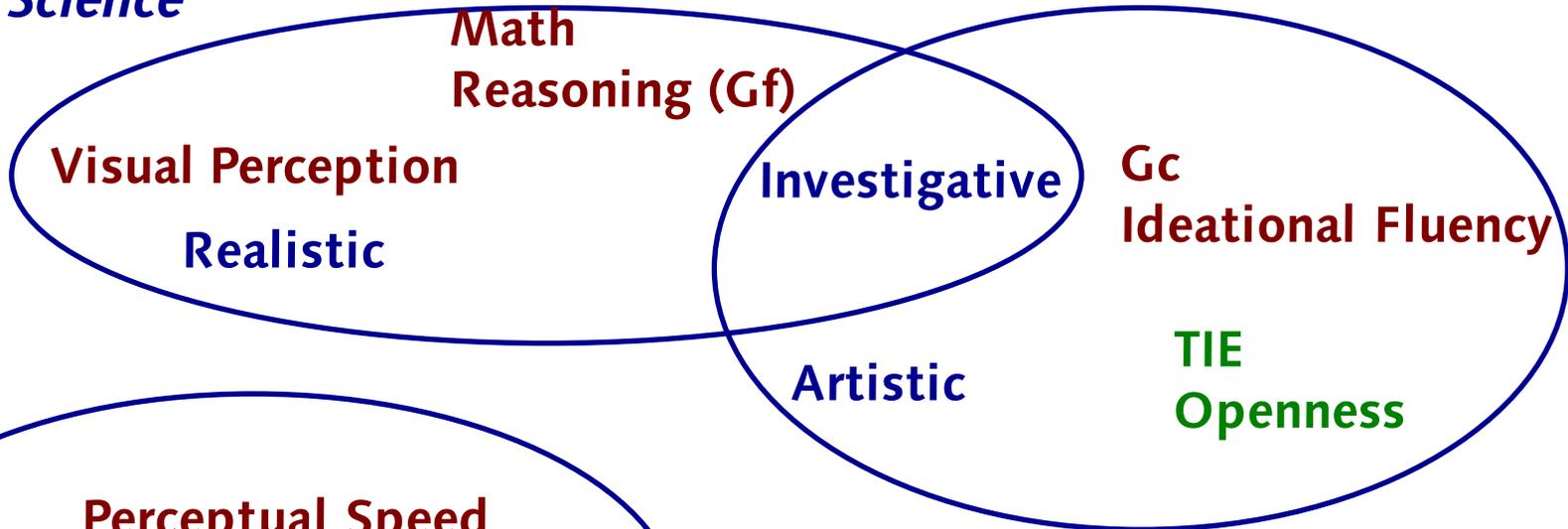
- PPIK-Theorie: Intelligence-as-**P**rocess, **P**ersonality, **I**nterests, and intelligence-as-**K**nowledge
- Modell zum Zusammenspiel von Kognitiven Fähigkeiten, beruflichen Interessen (RIASEC) und Persönlichkeitsmerkmalen (Big Five)
- Beschreibt wie Wissen aus der Interaktion von Intelligenz, Persönlichkeitsfaktoren und beruflichen Interessen entsteht



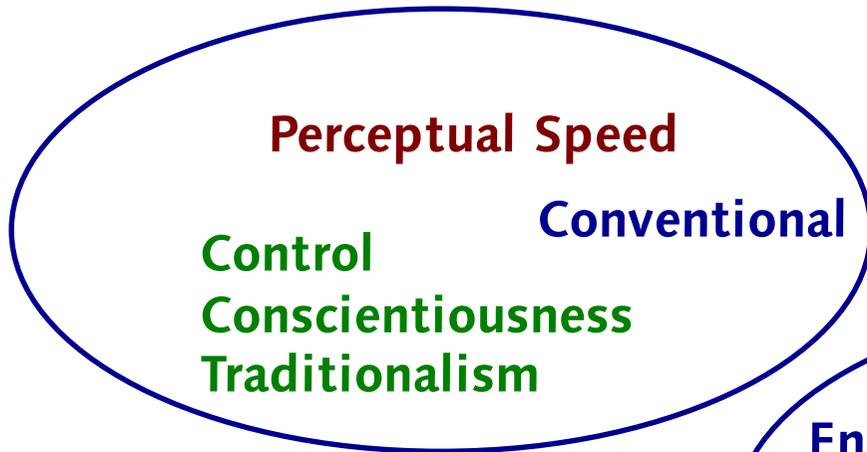
# Trait Complexes (Ackerman, 1996)



**Math/Science**



**Intellectual/Cultural**

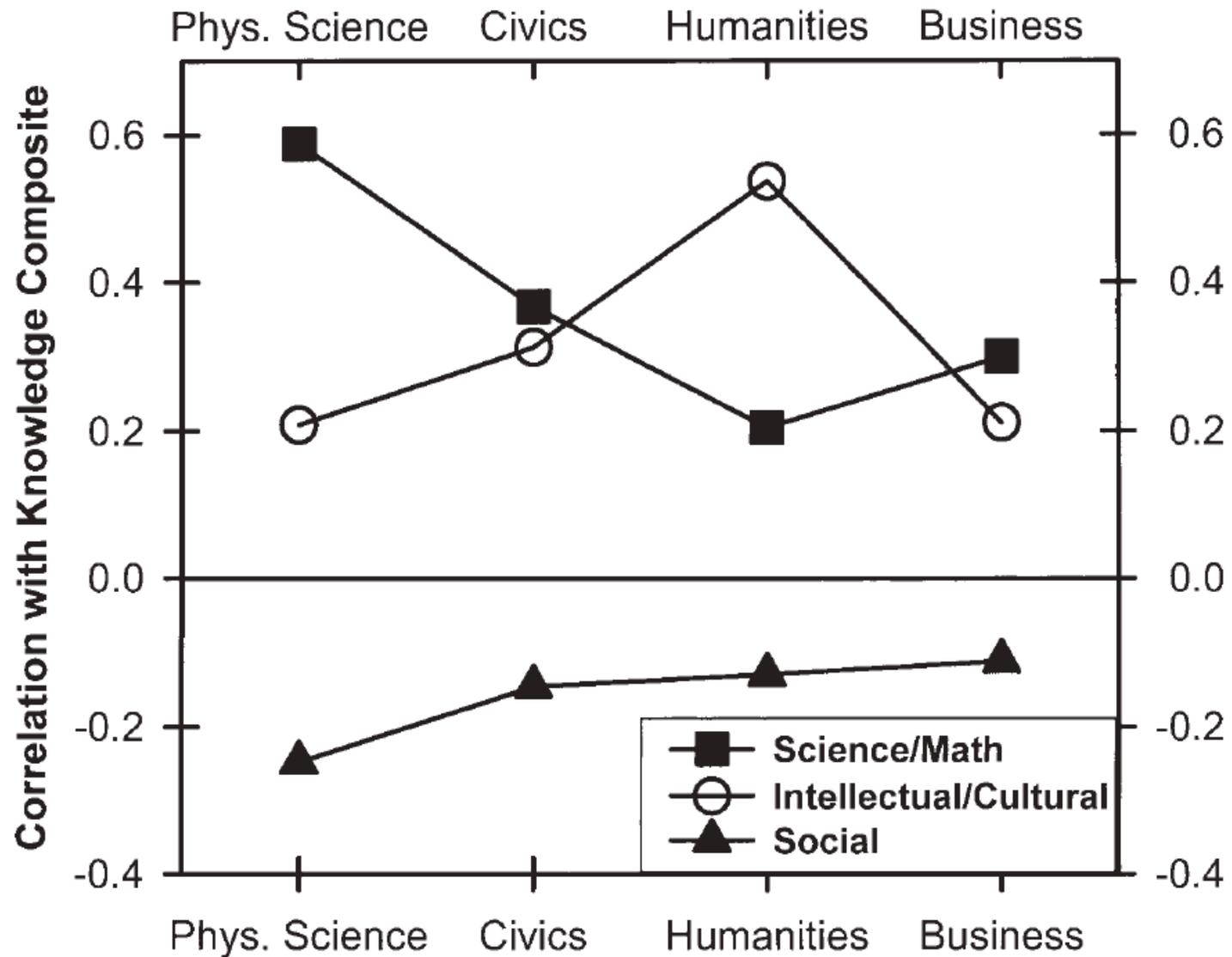


**Clerical/Conventional**



**Social**  
26

# Trait Complexes und Leistung (Ackerman & Beier, 2003)





# Schluss: Consequential Validity



- Positive und negative Folgen der Testentwicklung im Hochschulbereich
- Tests nur als Forschungsinstrumente oder auch als Motoren der Individualtestung?
- Outputorientierung im Hochschulbereich? Vergleiche zwischen Hochschulen?
- Vergleiche zwischen Studiengängen?



*There are no limits, only horizons!*

***Herzlichen Dank für Ihre Aufmerksamkeit!***

# Trait Complexes und University Majors (Ackerman & Beier, 2003)

