

Schwierigkeitserzeugende Merkmale physikalischer Leistungstestaufgaben

Alexander Kauertz

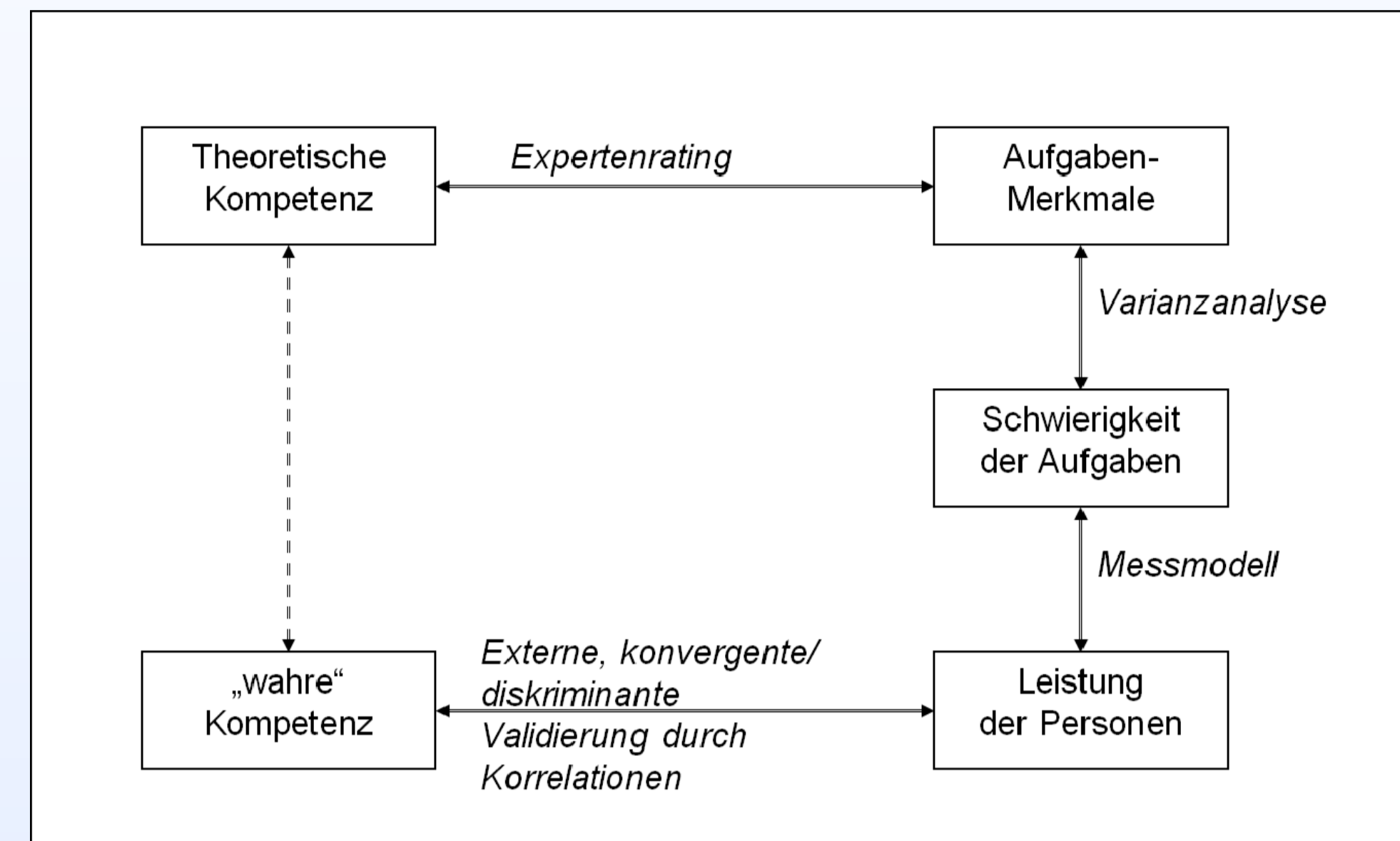
Gefördert durch: Deutsche Forschungsgemeinschaft DFG

Fragestellung

Inhalte sind bei Kompetenzmessung von zentraler Bedeutung. Es ist davon auszugehen, dass sie Merkmale besitzen, die schwierigkeitsgenerierend sind. Es wird z.B. angenommen, dass dadurch Kompetenz inhaltsbezogen gestuft werden kann. Neben dieser Struktur des Inhalts erzeugen allerdings auch Merkmale des Aufgabendesigns Schwierigkeit. In der Studie werden daher zwei Modelle verglichen, in denen inhaltsstrukturelle Merkmale und Aufgabendesign-Merkmale in ihrem Einfluss auf Schwierigkeit verglichen werden. Die folgenden Probleme werden in diesem Projekt bearbeitet:

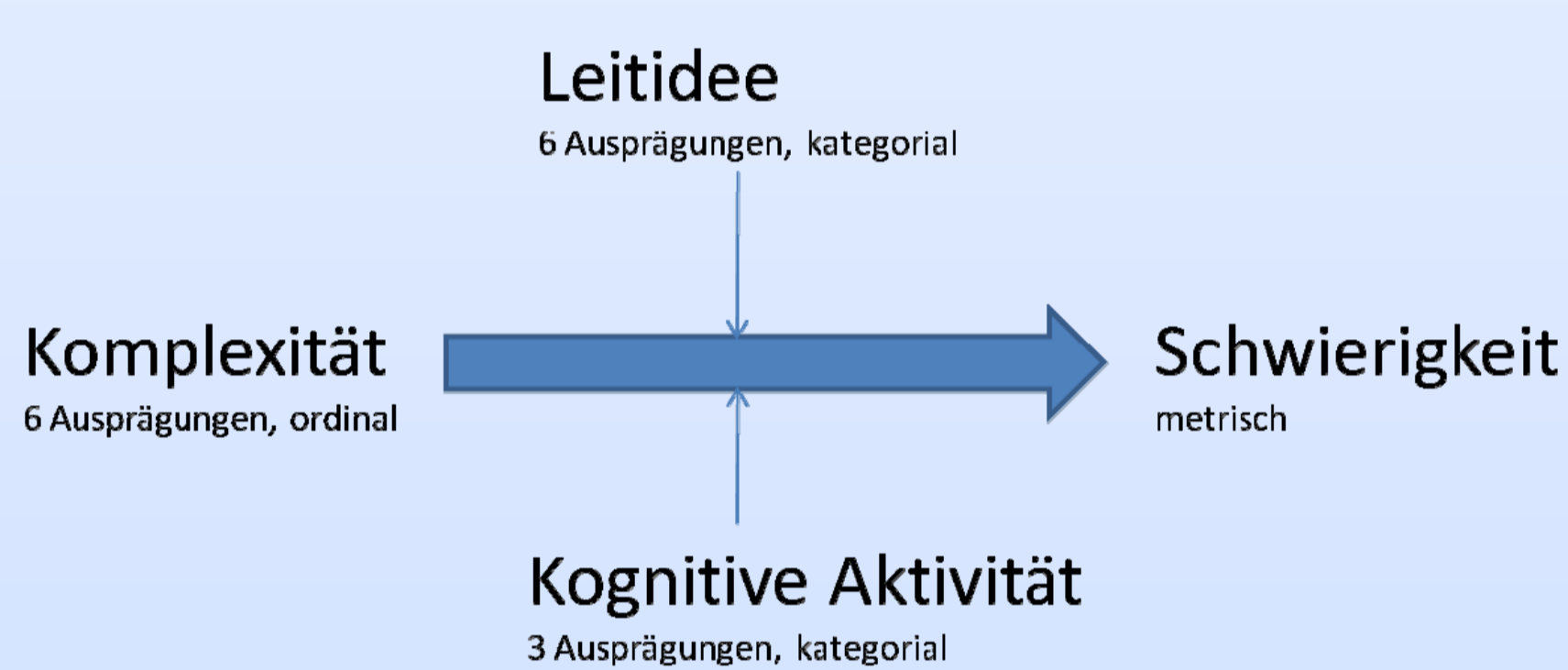
- (1) Entwickeln eines validen Kompetenztests, dessen Aufgaben Inhaltsstruktur- und Aufgabendesignmerkmale tragen.
- (2) Vergleich der Varianzaufklärung der Aufgabenschwierigkeit durch die Inhaltsstruktur und die Aufgabendesignmerkmale
- (3) Untersuchung der Dimensionierung des Inhaltsstrukturmodell mittels Modellvergleichen

Rahmenmodell



Arbeitsprogramm und verwendete Methoden der Studie

Inhaltsstruktur-Modell



Das Inhaltsstrukturmodell hat drei Dimensionen:

Komplexität: Fakt, Fakten, Zusammenhang, unverbundene Zusammenhänge, verbundene Zusammenhänge, übergeordnetes Konzept

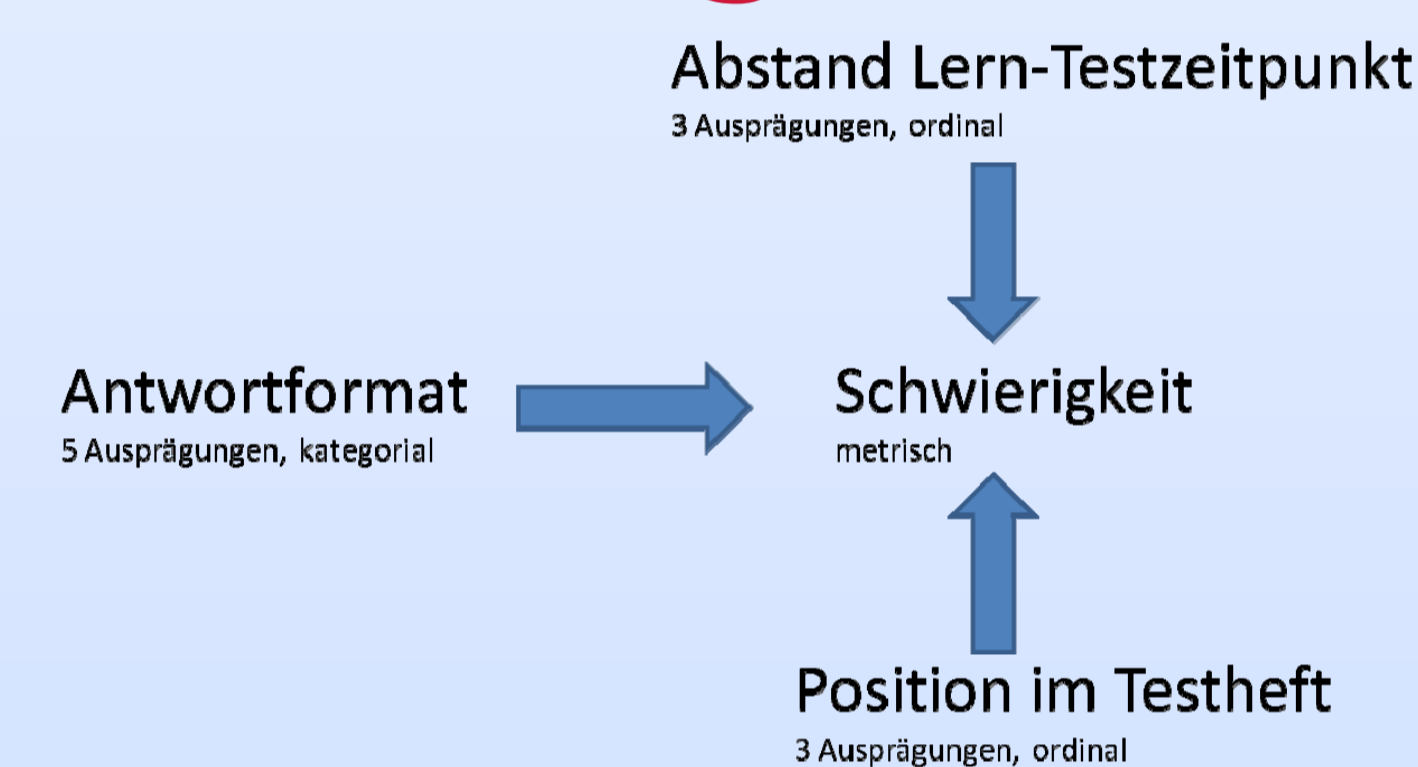
Kognitive Aktivität: Erinnern, Strukturieren, Explorieren

Leitidee: Energie, Materie, System, Wechselwirkung, naturwissenschaftliche Arbeitsweise, Formalismus

Komplexität wirkt direkt auf die Schwierigkeit, Kognitive Aktivitäten und Leitideen sind jeweils Moderatorvariablen.

	Kennwerte Inhaltsstruktur-Modell				
	F	df	p	R ²	QS
...vollständig	1,41	47	0,11	0,52	135,520
...Komplexität und Leitidee	4,32	10	<0,001	0,31	196,987

Aufgabendesign-Modell



Das Aufgabendesignmodell hat drei Dimensionen, die jeweils direkt auf die Schwierigkeit wirken:

Antwortformat: offen, halboffen, grafisch, Multiple Choice, Single Choice

Abstand Lern-Testzeitpunkt: entspricht der curricularen Zuordnung des Inhalts zu einer Jahrgangsstufe (6, 8, 9/10)

Position im Testheft: Vorne, Mitte, Hinten

	Kennwerte Aufgabendesign-Modell				
	F	df	p	R ²	QS
...vollständig	2,785	8	0,008	0,184	232,443
...ohne Antwortformat	4,033	4	0,004	0,135	246,191
...nur Antwortformat	2,626	4	0,039	0,093	258,402

Zu (2): Der Vergleich der beiden Modelle erfolgt über die aufgeklärte Varianz der Schwierigkeit:

Je größer R² (oder kleiner die Fehlervarianz QS), desto besser ist das Modell. Je weniger Freiheitsgrade (df) benötigt werden, desto besser ist das Modell. Zwar benötigt das Modell aus Komplexität und Leitidee 6 Freiheitsgrade mehr, klärt aber deutlich mehr Varianz auf und ist daher das zu bevorzugende. **Das Inhaltsstrukturmodell eignet sich zur Modellierung von Kompetenz.**

Zu (3): Es werden drei Varianten des Inhaltsstrukturmodells untersucht:

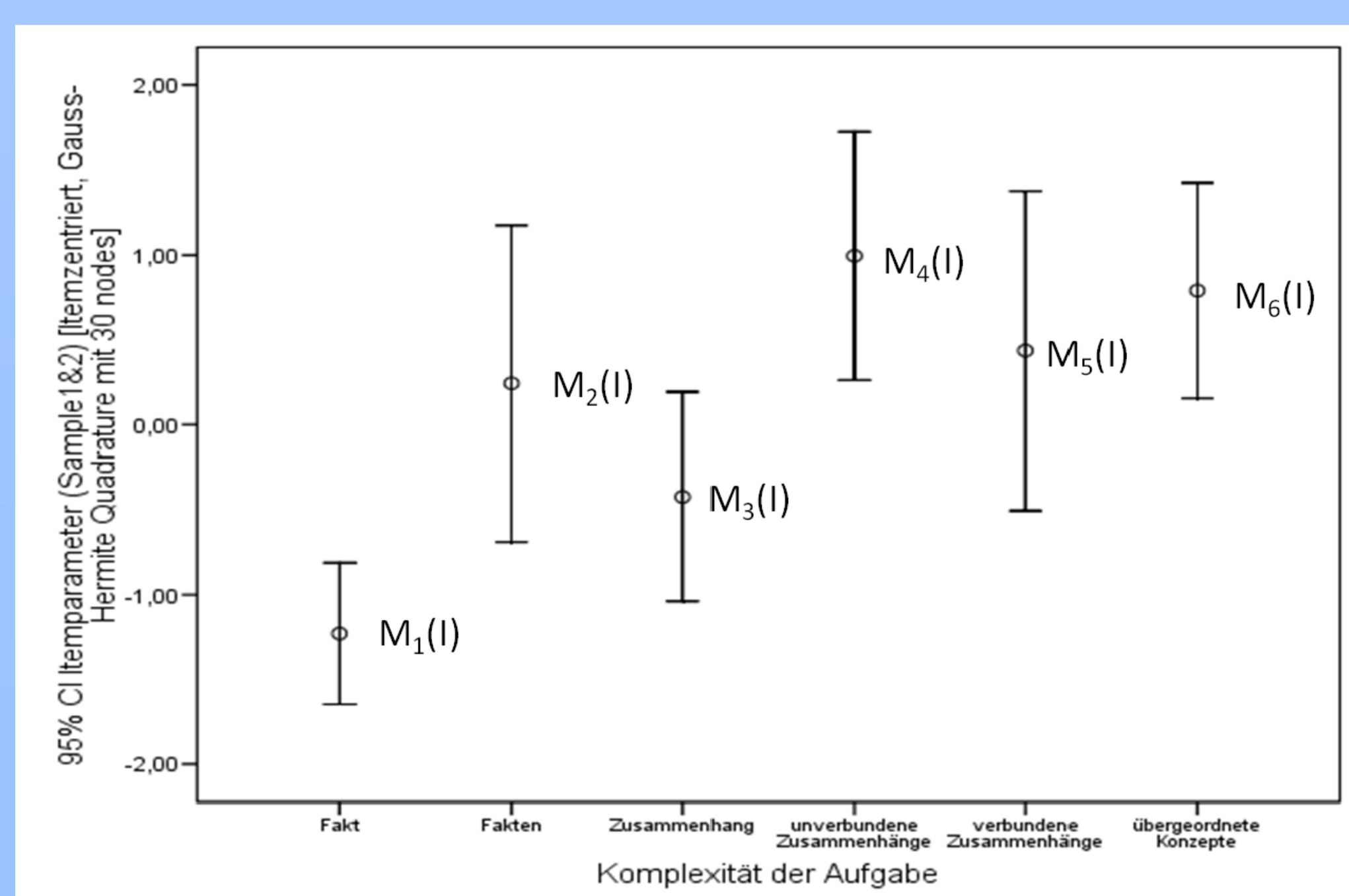
Leitideen und kognitive Aktivitäten werden nicht berücksichtigt (1. Modell mit 1 Skala), kognitive Aktivitäten werden berücksichtigt (2. Modell mit 3 Skalen) und Leitideen werden berücksichtigt (3. Modell mit 6 Skalen).

	Deviance	df	BIC
1. Modell	18346,67 3	133	18709,544 1
2. Modell (Kognitive Aktivitäten)	18340,32 5	138	18716,83 78
3. Modell (Leitideen)	18288,27 4	153	18705,71 21

Wenn ein Modell auf die Daten passt, sind die darin beschriebenen Skalen unabhängig voneinander und repräsentiert je eine Teilkompetenz.

Bayes Information Criterion dient als Vergleichsmaß für die Modelle. Je geringer der Wert, desto besser die Passung auf die Daten (siehe Tabelle links).

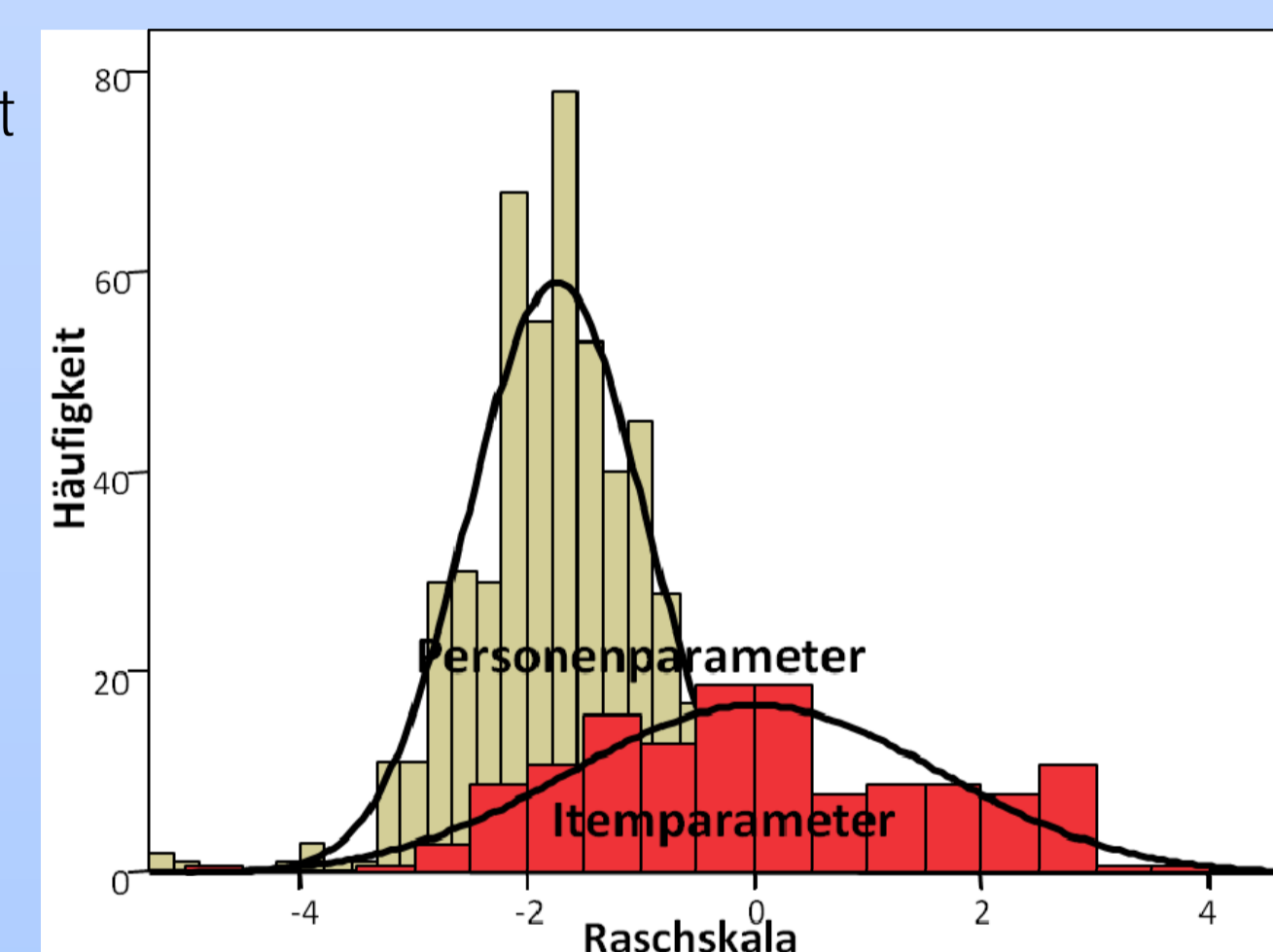
Daraus ergeben sich 6 Teilkompetenzen für Physik, die den 6 Leitideen entsprechen.



Die sechs Komplexitätsniveaus: ein Fakt, mehrere Fakten, ein Zusammenhang, mehrere Zusammenhänge und übergeordnetes Konzept

korrelieren mit der Schwierigkeit von Aufgaben (r = 0,38). **Aus der Komplexität ergeben sich somit Kompetenzstufen.** Die Konfidenzintervalle sind jedoch so groß, dass sich die Stufen empirisch nicht trennen lassen.

Testgüte



Reliabilität: $\alpha = 0,718$

$M_{\text{Personenparameter}} = -1,669$ $SD = 0,821$

Der Kompetenztest ist zufriedenstellend reliabel und konstrukt- sowie curricular valide. Die Items des Tests sind jedoch zu schwierig, die Passung zwischen Aufgabenschwierigkeit und Schülerfähigkeit ist also nicht gegeben. Die hohe Aufgabenschwierigkeit, die geringe Itemanzahl pro Skala, die relativ geringe Probandenzahl und das selektive Sample begrenzen die Generalisierbarkeit der Befunde.

Validierung:	Personenparameter im Leistungstest
Personenparameter im TIMSS-Test (WLE)	0,604(**)
Roh-Summenwert im kognitiven Fähigkeitstest nach Heller et al.	0,431(**)
Motivation (Gesamtskalenwert)	0,318(**)
Physiknote letztes Zeugnis	-0,479(**)
Wissenszuwachs (Residuen)	0,510(**)

Sample:

- N = 594 Schülerinnen und Schüler
- 10. Klasse, Gymnasium, NRW (22 Klassen)
- 120 Aufgaben in 6 Testheften (Multi-Matrix-Sample mit Ankeritems)
- Für N = 277 Schülerinnen und Schüler liegen zudem Daten vor zu:
 - Kognitiver Fähigkeit (KFT nach Heller und Perleth, 2000)
 - Motivation (nach Weblar & Pijahn, 2001)
 - Vorwissen (Test aus TIMSS-Physik-Items, Noten)
 - Wissenszuwachs (Prä-Post-Test über ein Halbjahr mit TIMSS-Items)

Kontakt

Universität Duisburg-Essen
Forscherguppe und Graduiertenkolleg
„Naturwissenschaftlicher Unterricht“
Schützenbahn 70, 45127 Essen
alexander.kauertz@uni-due.de
hans.fischer@uni-due.de

