

# **Standardbezogene Unterrichtsevaluation**

Möglichkeiten und Grenzen einer entwicklungs- bzw.  
rechenschaftsorientierten Evaluation schulischer Leistungen

**Andreas Helmke, Landau/Pfalz**

Baden, 02.11. 2004

© Helmke 2004

# Übersicht

- **Leistungsevaluation: Vielfalt von Facetten**
- **Ein Angebot-Nutzungs-Modell des Unterrichts**
- **Zwei Sichtweisen der Unterrichtsqualität**
- **Bedingungen der Nutzung externer Evaluation für die Unterrichtsentwicklung**
- **Das Projekt VERA (Vergleichsarbeiten in der Grundschule)**
- **Resümee: Bedingungen und Grenzen wirkungsorientierter Unterrichtsentwicklung**

# Evaluation schulischer Leistungen: Vielfalt von Komponenten und Facetten

- Benchmark-Orientierung: Landesmittel oder äquivalente Gruppe ("fairer Vergleich")?
- Hierarchische Ebenen: System - Schule – Klasse – Schüler(gruppe)
- Auflösungsgrad: Punktwert vs. Kompetenzverteilung, global vs. partikular, %correct vs. Falschlösungen
- Parameter: Mittelwert, Streuung, Zusammenhangsmaße (z.B. Koppelung von Sozialschicht und Sprachherkunft mit Kompetenz)
- Zeitfaktor: punktuell vs. Dauerbeobachtung

# Drei Bezugssysteme: am Beispiel der "Leistung" als Unterrichtsprodukt

## 1. **verteilungsbezogen**

(auch: komparativ, normbezogen)

Verortung durch Vergleich mit der Leistung Anderer: mit einer Norm, mit dem Durchschnitt, mit den Besten ("benchmark")

## 2. **sachbezogen**

(auch: kriterial)

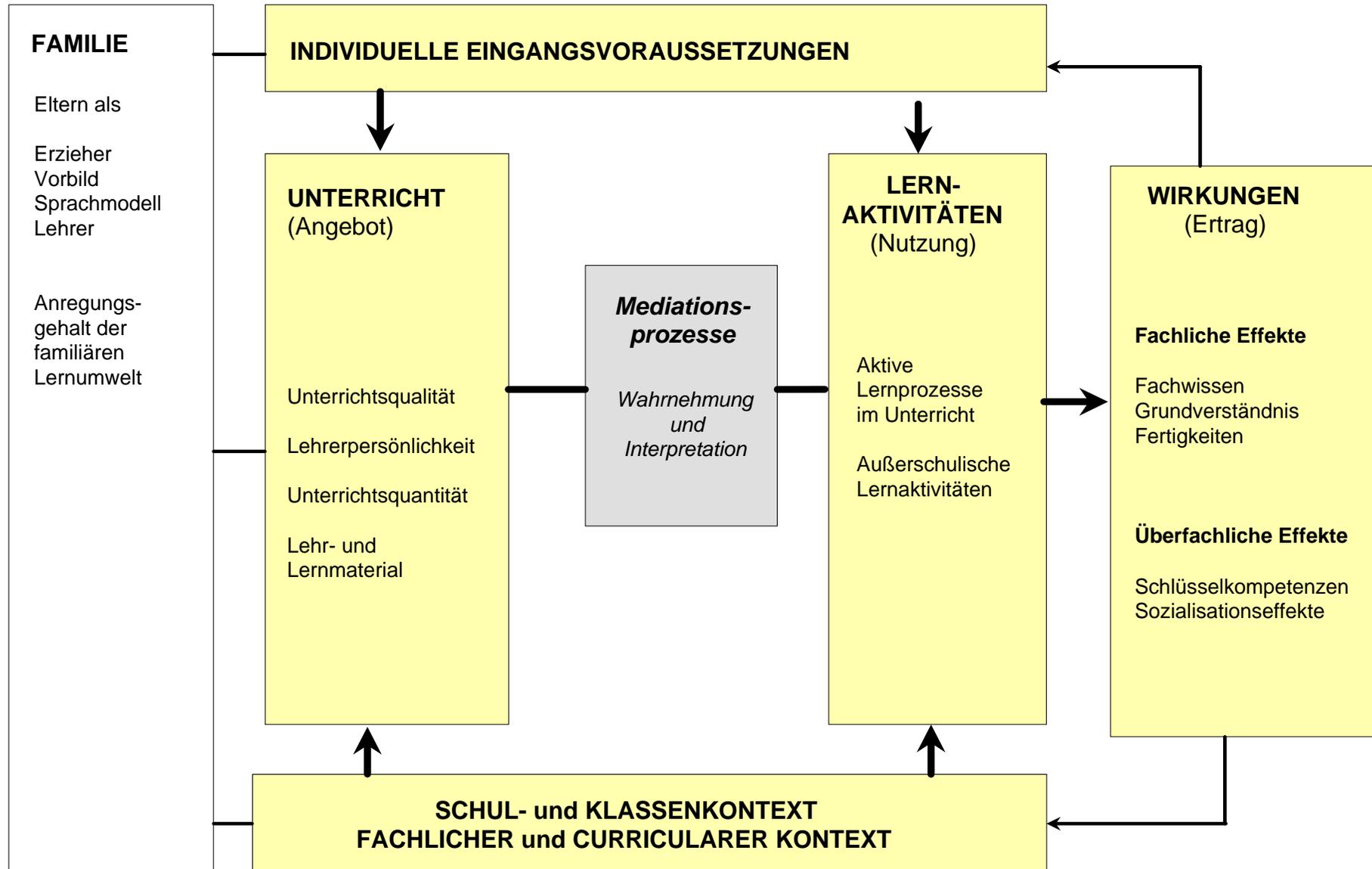
Verortung durch Vergleiche mit einem inhaltlich definierten Kriterium (einem Standard, einer Kompetenzstufe)

## 3. veränderungsbezogen

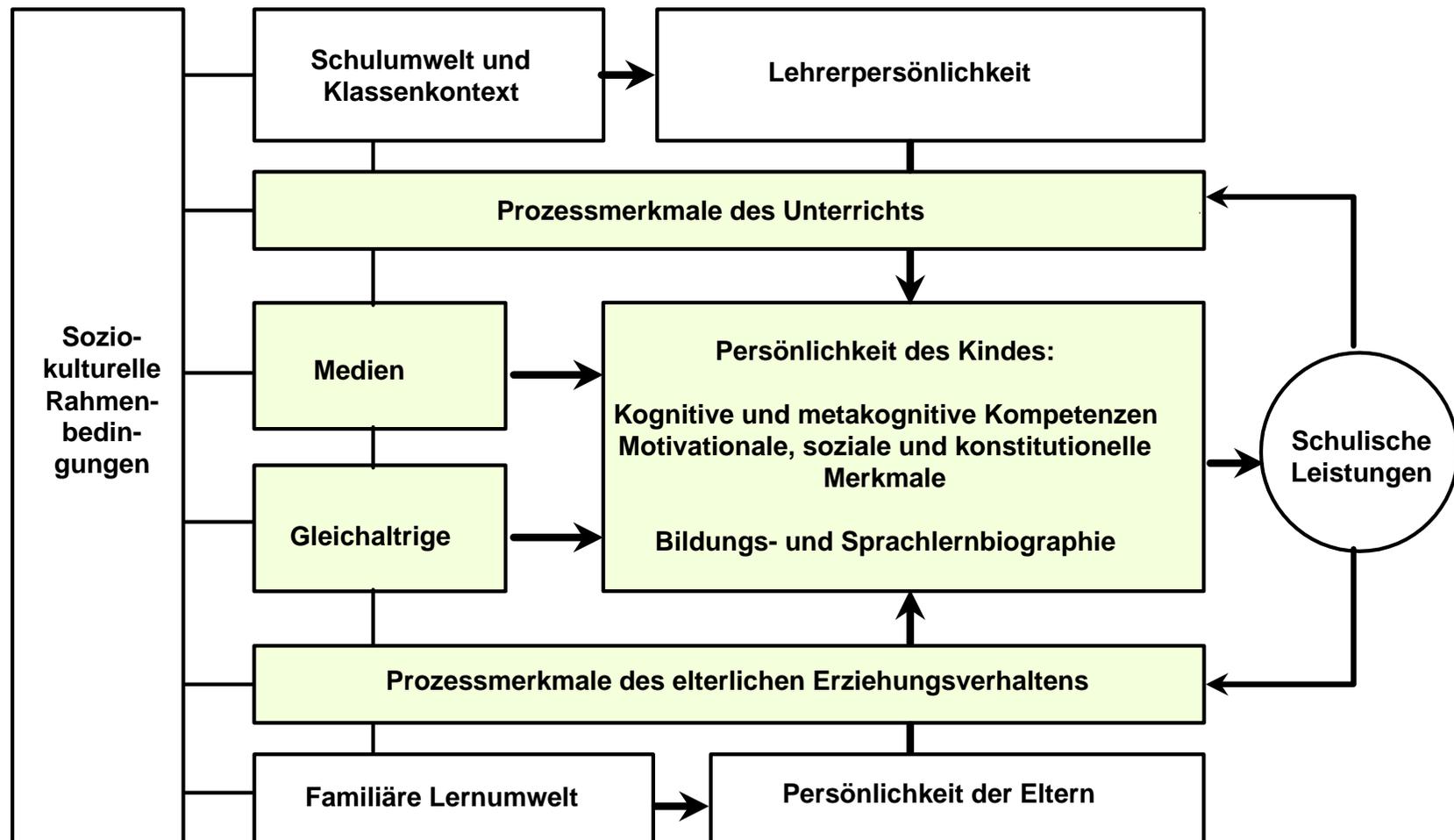
(auch: temporal, ipsativ)

- Vergleiche der gleichen Personen (Klassen, Schulen) über die Zeit: Zuwachs, Verlust o. Stagnation z.B. innerhalb eines Schuljahres (Pre-Post-Design; Längsschnitt)
- Vergleiche äquivalenter Altersgruppen zu verschiedenen Zeitpunkten (Zyklus, Zeitwandel)

# Ein Modell der Wirkungsweise des Unterrichts



# Ein Makro-Modell der Bedingungs-faktoren schulischer Leistungen



# Warum es „die“ optimale Lehrmethode nicht gibt - und nicht geben kann

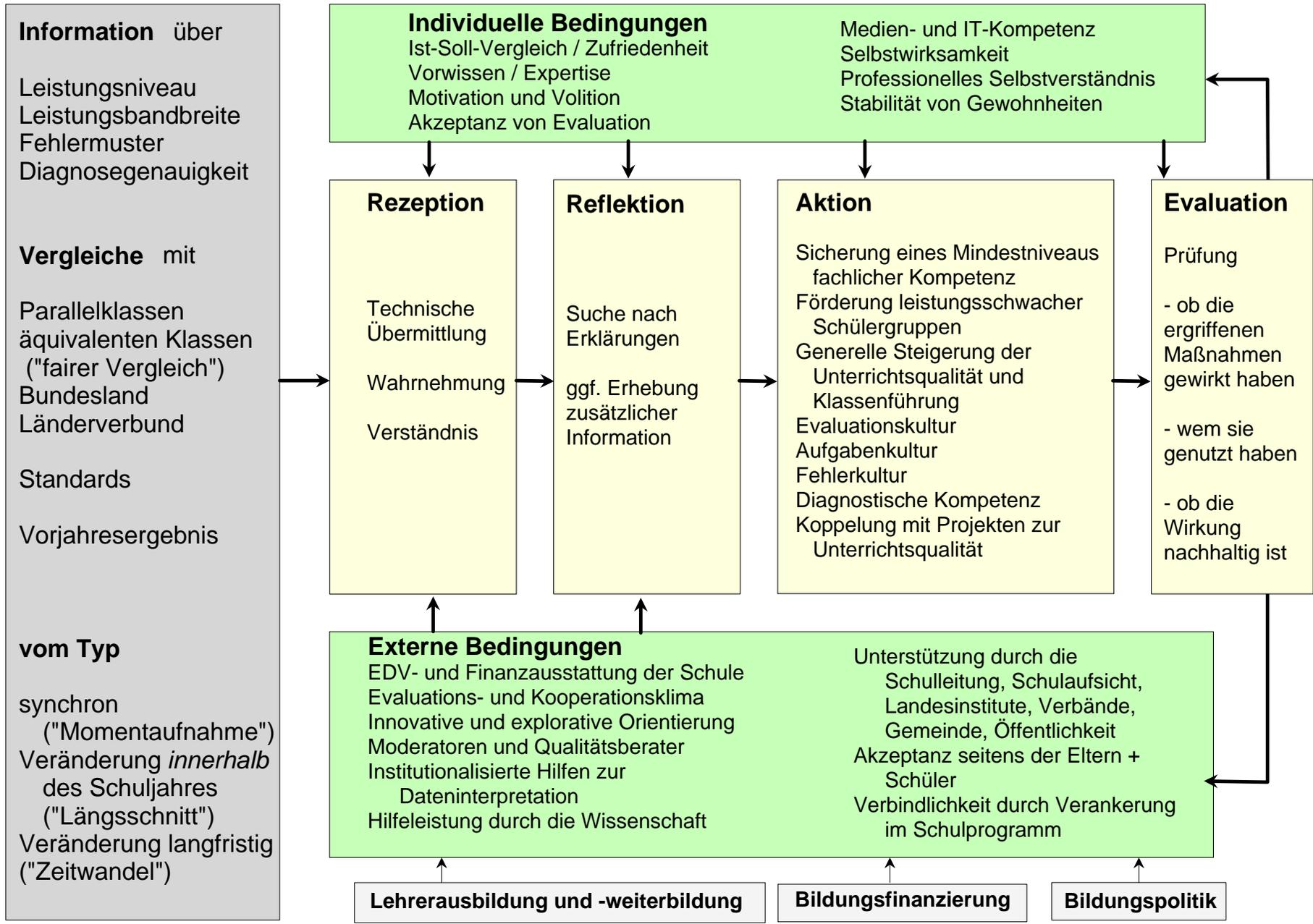
- Gut wofür? (für welche Bildungsziele?)
- Gut für wen? (Wechselwirkungen)
- Gut gemessen an welchen Startbedingungen (Klassenkontext)?
- Gut aus wessen Sicht? (Beurteilerperspektive)
- Gut für wann? (Zeitdimension)

# Ausgewählte Facetten des Modells

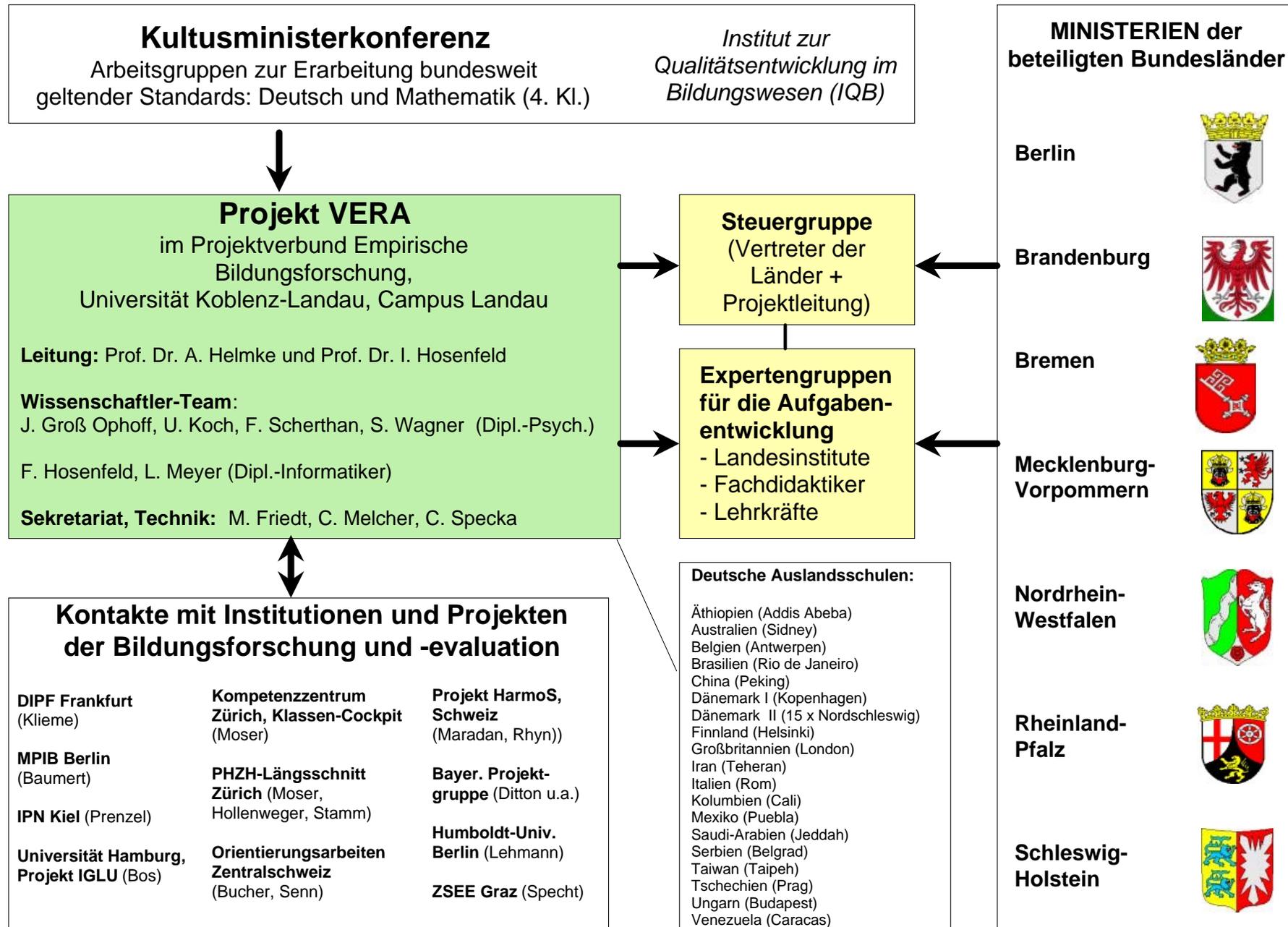
- Unterricht ist nur ein Angebot
- Unterrichtsqualität ist kontextabhängig
- Vielfalt von Zielen schulischen Unterrichts
- Beziehungen zwischen Unterrichtsmerkmalen und Leistungen sind weder direkt noch linear
- Lehrerkompetenzen sind teilweise wechselseitig kompensierbar
- Schülerleistungen sind vielfach bedingt; Unterrichtsqualität ist nur ein Wirkfaktor
- Eine Konsequenz: Die "Evaluation" der Qualität einer einzigen Unterrichtsstunde, ohne Kontextwissen, und mit akzeptablem Aufwand, ist nicht seriös.

# Bewertung der Unterrichtsproduktes

		negativ	positiv
Bewertung des Unter- richts- <u>prozesses</u>	negativ	"schlechter" und wirkungsloser Unterricht	<b>"schlechter", aber wirkungsvoller Unterricht</b>
	positiv	<b>"guter", aber wirkungsloser Unterricht</b>	"guter" und wirkungsvoller Unterricht



VERA



# Vier Ziele von Vergleichsarbeiten

- Bestandsaufnahme und Sicherung von Leistungsstandards auf System-, Schul- und Klassenebene
- Verbesserung der Professionalität der Lehrertätigkeit, insbesondere fachdidaktischer und diagnostischer Kompetenzen, und damit: der Unterrichtsqualität
- Ergänzende Information zur Schullaufbahnberatung der Eltern
- Beschleunigung der Umsetzung neuer Lehr- und Rahmenpläne
- Positive Nebenwirkung: Verbesserung informationstechnischer Kompetenzen der Lehrkräfte

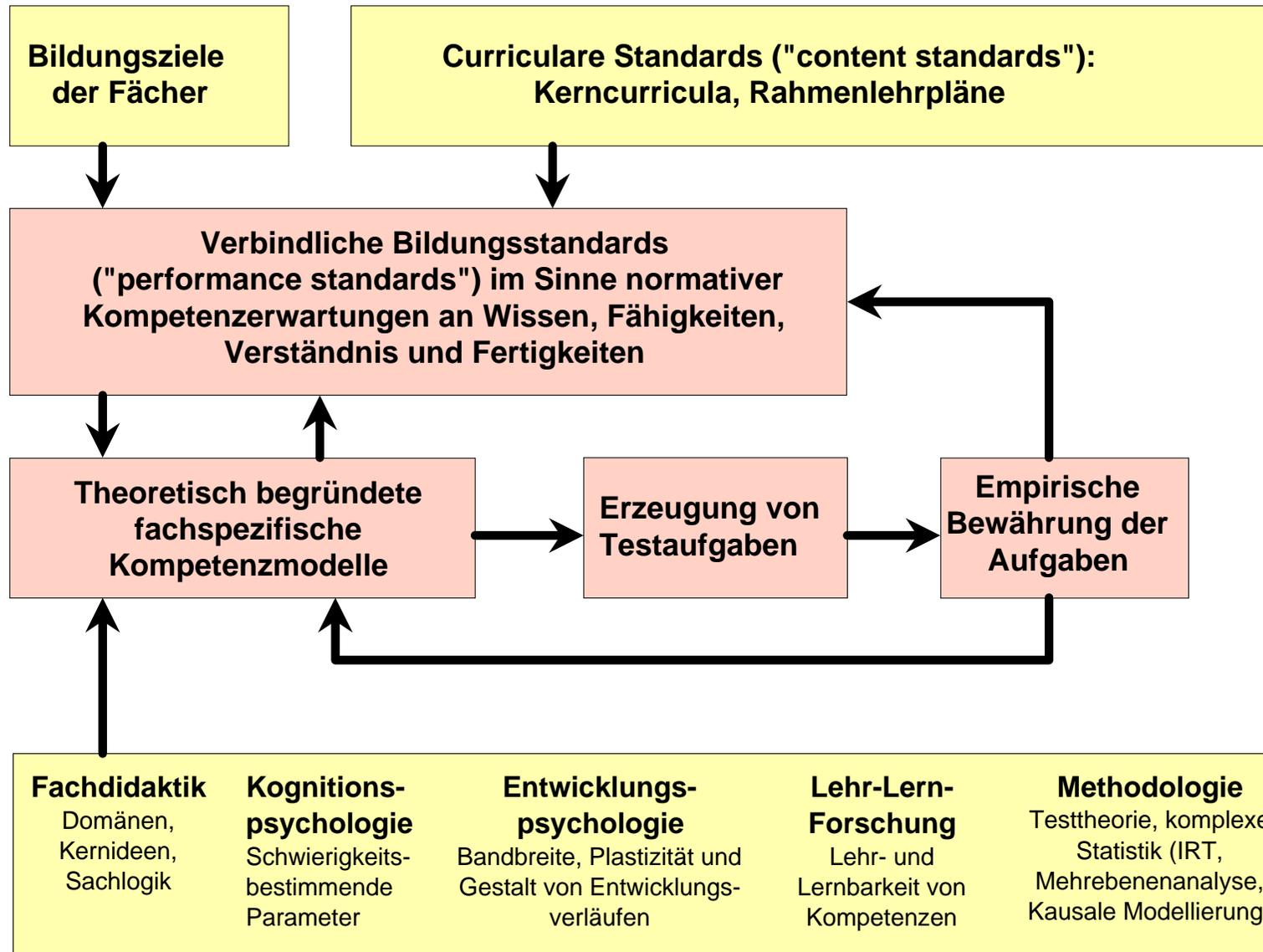
# Mathematik

- **Inhaltsbereiche**
  - Arithmetik
  - Geometrie
  - Sachrechnen
- **Tätigkeitsanforderungen**
  - Skizzieren
  - Schätzen
  - Argumentieren
- **Curriculum**
  - Repräsentation der zentralen Bereiche in hinreichender Breite und in Kombination mit verschiedenen Anforderungen

# Deutsch

- **Leseverständnis**
  - Kontinuierliche Texte (z.B. literarischer Text)
  - Diskontinuierliche Texte (Sachtexte, z.B. Fahrplan)
- **Schreiben**
- **Orthographie**
  - Satzdiktate
  - Alphabetische Sortierungen
  - Fehlerkorrekturen
  - Groß-Klein-Schreibung
- **Sprachbetrachtung**
  - Wortschatz
  - Syntax
- *Hörverständnis (geplant ab 2005)*

# Bildungspolitik



## Wissenschaft

# Vergleichsarbeiten als Rohstoff für die Diagnose von ....

- Aufgaben: Unterschieden zwischen Aufgaben (Aufgabenschwierigkeiten)
- Personen: Unterschiede zwischen der Leistung / der Rangposition / unterschiedlicher Schüler oder Schülergruppen
- Klassen: Unterschiede im Leistungsniveau / der Leistungsstreuung zwischen Klassen einer Schule
- Lernfortschritten (von Schülern und Klassen)
- Standortbestimmung der Klasse / der Schule im Vergleich mit Landes- oder nationalem Durchschnitt

# Pädagogische Optionen von VERA

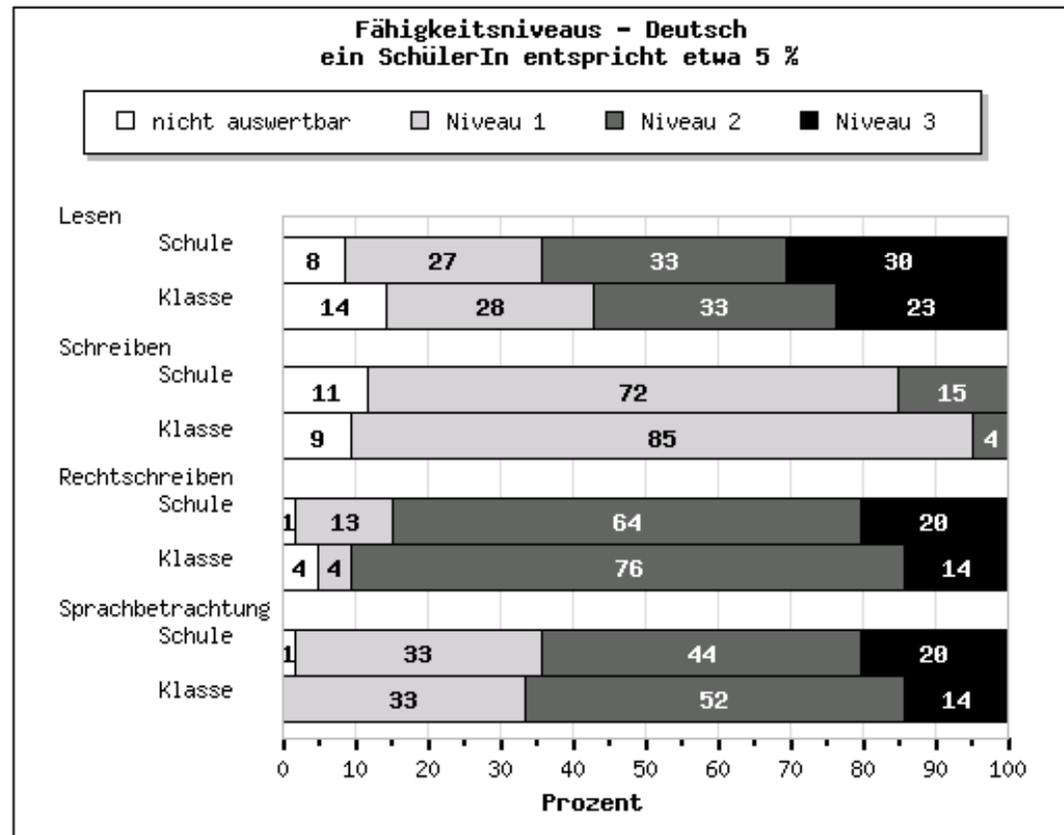
- Schaffung von Anlässen für kollegiale **Kooperation** im didaktischen Bereich (Auswahl, Beurteilung, Entwicklung von Aufgaben)
- **Fachdidaktische** Impulse (z.B. Analyse von Fehlern)
- **Pädagogische** Impulse: z.B. Suche **nach** Gründen für erwartungswidrige Ergebnisse
- Empirische Basis für gezielte Maßnahmen der **Förderung** (einzelner Schüler / defizitärer Bereiche)
- **Evaluation** der Effekte v. Förderungsmaßnahmen
- Erfassung + Training **diagnostischer Kompetenzen**

# **Ergebnisrückmeldung bei VERA 2004:**

## **Beispiele**

## Ergebnisse Deutsch: Fähigkeitsniveaus - Vergleich Klasse/Schule

Die einzelnen Niveaus werden in der Handreichung zur [Beschreibung der Fähigkeitsniveaus für Deutsch](#) detailliert erläutert! Hier wird die prozentuale Verteilung der Fähigkeitsniveaus Ihrer Klasse im Vergleich mit Ihrer Schule dargestellt.

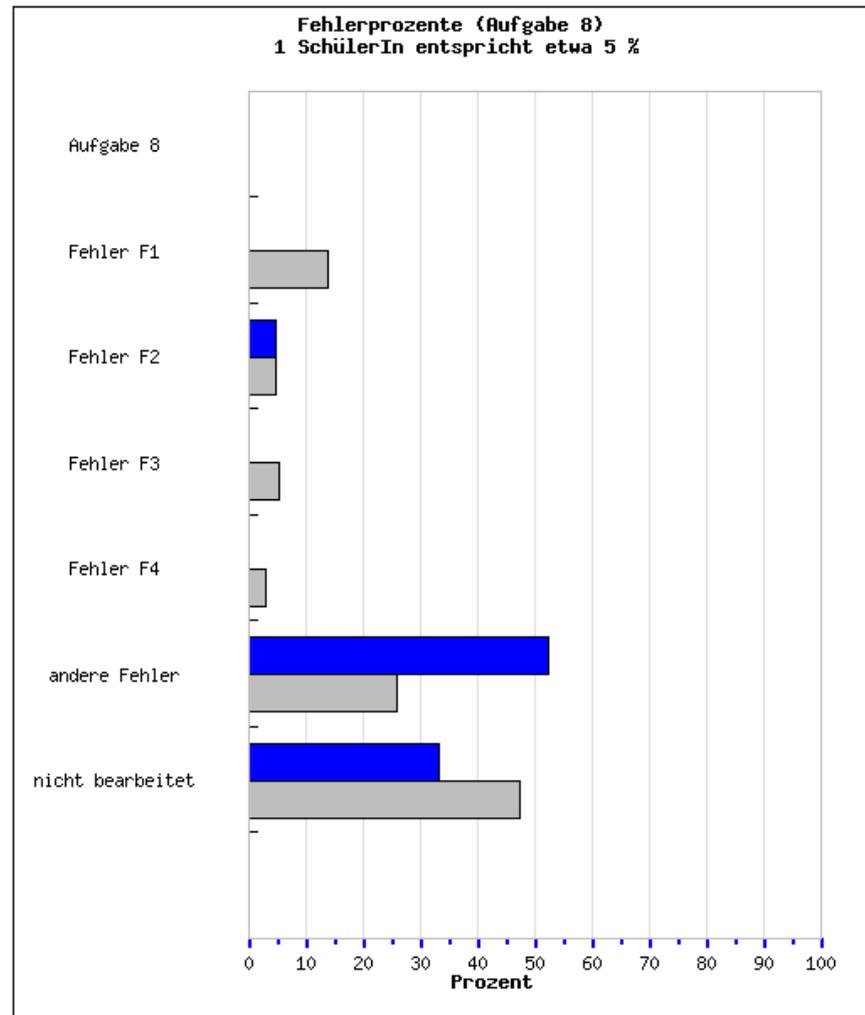


## Ergebnisse Mathematik: Fähigkeitsniveaus aller Schüler

Klasse: 4			
SchülerIn	Arithmetik	Geometrie	Sachrechnen
1	3	1	2
2	1	2	1
3	2	2	2
4	2	2	2
5	n.a.L.*	1	2
6	1	2	1
7	2	2	1
8	2	1	1
9	2	2	2
10	1	2	1
11	2	2	1
12	2	2	2
13	2	2	3
14	3	2	3
15	n.a.L.*	2	1
16	1	n.a.L.*	1

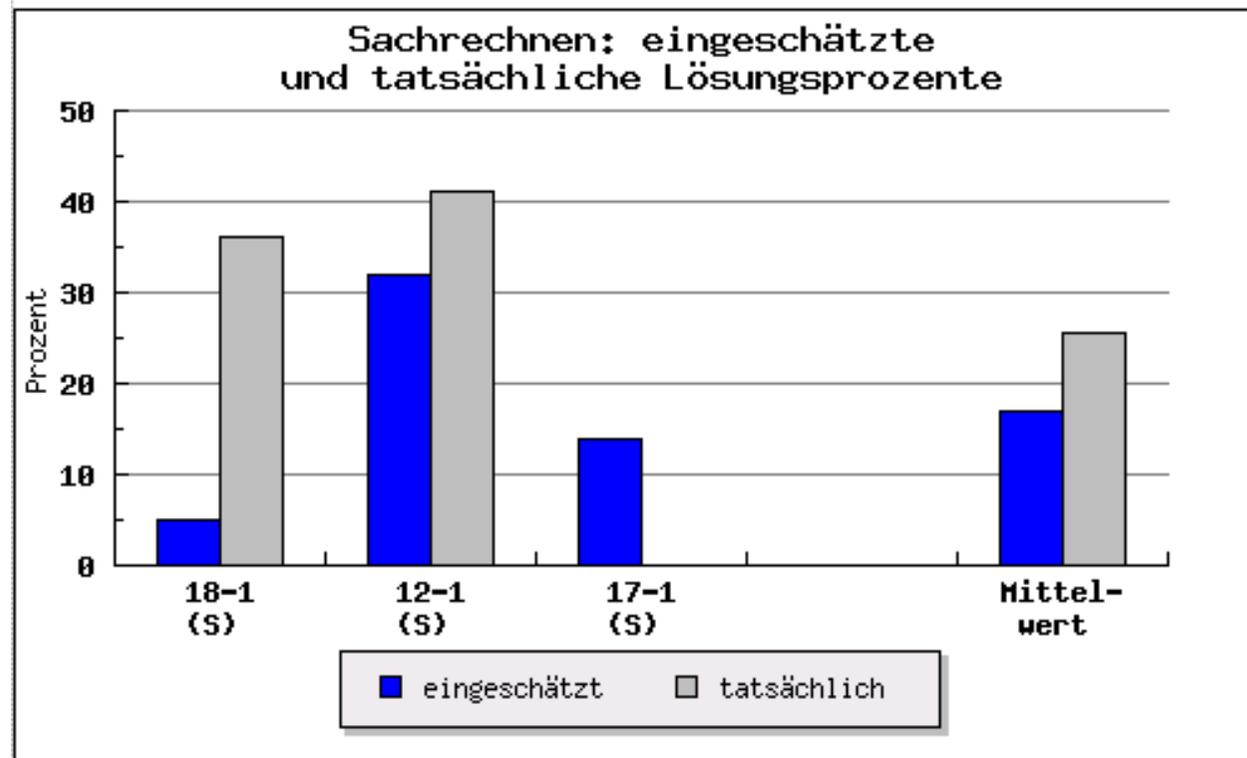
## Fehlerhäufigkeiten im Detail: Aufgabe 8

Die Grafik beinhaltet detaillierte Informationen auf der Ebene der einzelnen Aufgabe.  
Der obere Balken der Balkenpaare steht immer für Ihre Klasse, der untere für die Normierung.  
Die folgenden Balkenpaare zeigen Gegenüberstellungen für einzelne Fehlertypen.



## Sachrechnen: Geschätzte und tatsächliche Lösungshäufigkeiten

Die Abbildung zeigt den Vergleich Ihrer Einschätzungen mit den tatsächlichen Lösungshäufigkeiten, hier für den Bereich Sachrechnen. Zusätzlich zeigt das rechte Balkenpaar den jeweiligen Mittelwert an.



Ein Schüler entspricht etwa 5 Prozent

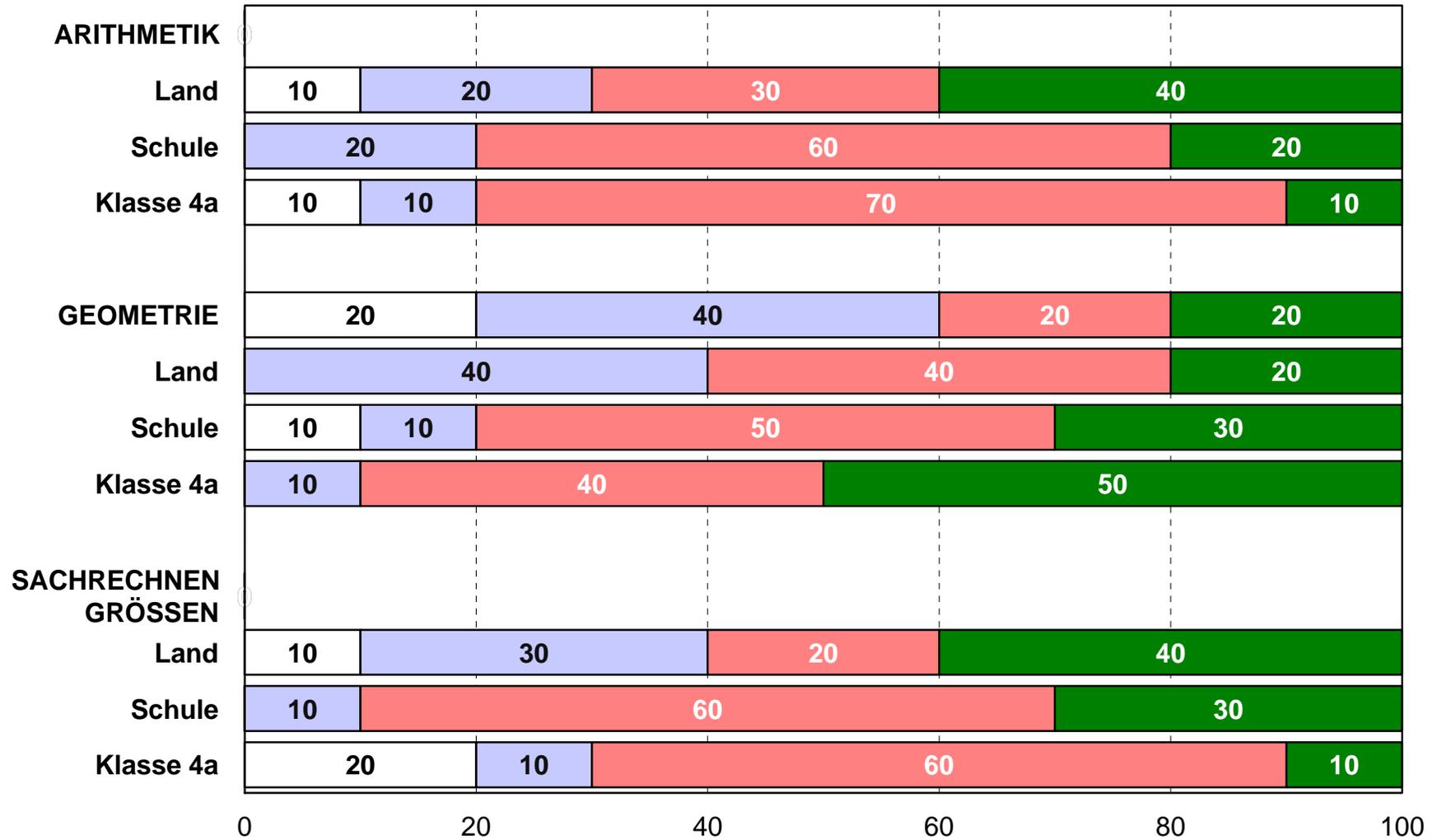
## Vergleich der geschätzten und tatsächlichen Rangfolge der Aufgabenschwierigkeit

Die Tabelle stellt die Rangfolge Ihrer Schätzung der Aufgabenschwierigkeit der Rangfolge der tatsächlichen Lösungshäufigkeit gegenüber. Die Aufgaben sind nach Ihren Schwierigkeitsschätzungen aufsteigend geordnet. Für weitere Informationen klicken Sie auf die jeweiligen Tabellenüberschriften.

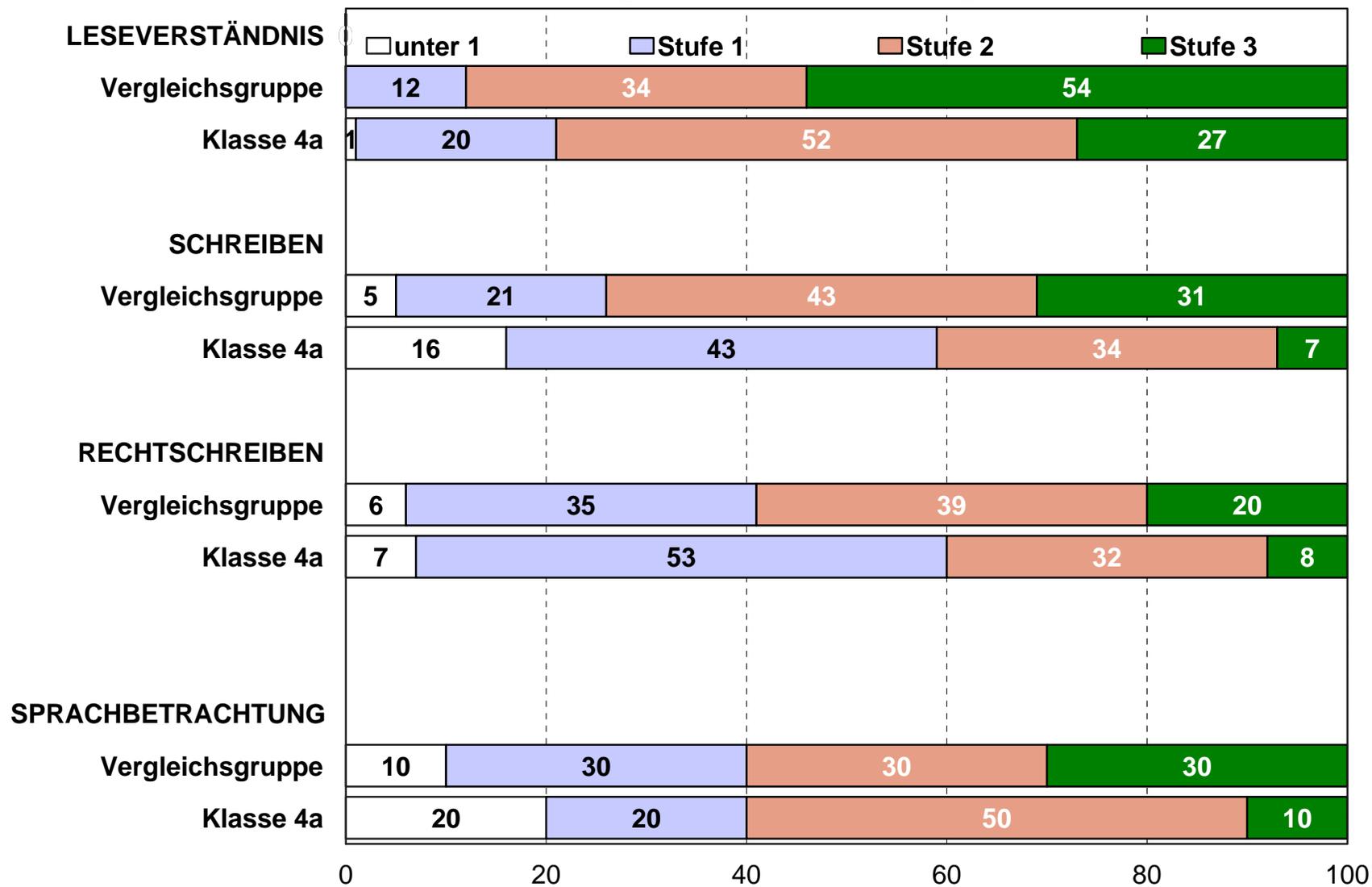
<u>Ihre</u> <u>Schätzung</u>	<u>tatsächliche</u> <u>Rangfolge</u>	<u>Inhalts</u> <u>bereich</u>	<u>Aufgaben</u> <u>nummer</u>
1	1	A	16-1
1	2	A	16-2
1	6	A	16-3
4	13	A	15-1
5	10	S	12-1
6	6	A	19-1
6	3	A	19-3
6	3	A	19-4
6	3	A	19-2
10	6	G	11-1
11	16	S	17-1
11	16	G	13-1
11	9	G	14-1
14	14	A	20-2
14	11	A	20-1
16	14	A	20-3
16	11	S	18-1

unter 1    
 Stufe 1    
 Stufe 2    
 Stufe 3

### Prozentuale Fähigkeitsstufenverteilung

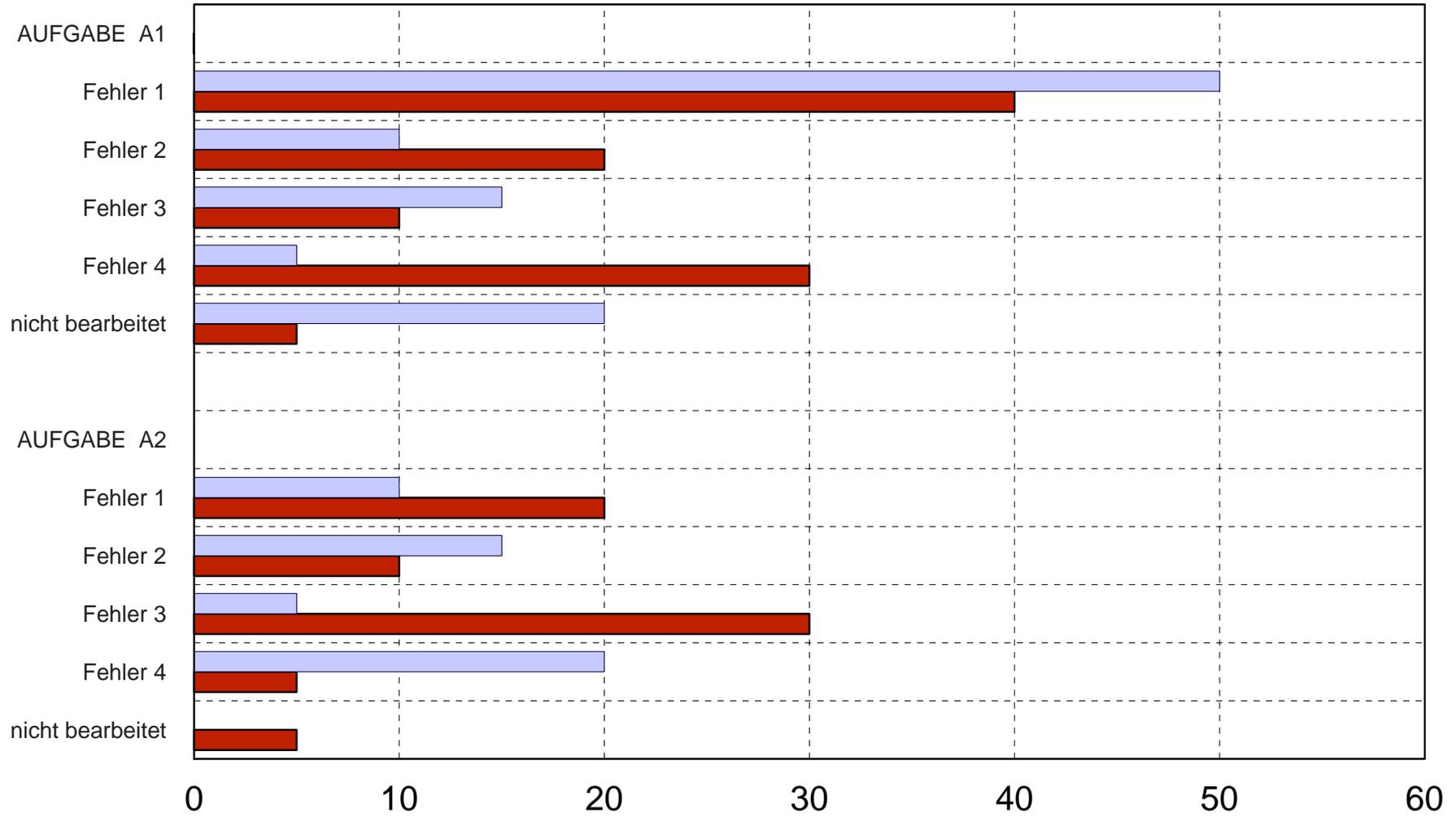


## Prozentuale Fähigkeitsstufenverteilung



■ Klasse X

■ insgesamt (Normierungsstudie)



# Lehrer als Diagnostiker

# Diagnostische Kompetenzen von Lehrkräften

- Diagnostische Kompetenz: eine Schlüsselkompetenz für guten Unterricht
- Alarmierendes PISA-Ergebnis: Magere Trefferquote (10%) bei der Frage „Welche Schüler sind schwache Leser?“
- Lehrkräfte wissen zu wenig über die individuellen Defizite ihrer Schüler.
- Sie wissen auch kaum, wie ihr Unterricht bei den Schülern „ankommt“.

## Diagnostische Expertise umfasst neben der Diagnosegenauigkeit ...

- Orientierung über Gütekriterien und Bezugsnormen der Beurteilung schulischer Leistungen
- Kenntnis typischer Fehler und Verzerrungen des Lehrerurteils
- Fähigkeit, im Unterricht selbst Tests ad hoc zu entwickeln, einzusetzen, auszuwerten und die Ergebnisse zurückzumelden
- Orientierung über wichtige Test- und Fragebogenverfahren für schulische Zwecke und Wissen, wie man sich diese beschaffen kann

Status

Eingabe der  
Lösungs-  
häufigkeiten

Navigation

Angemeldet: B 007008



SITZUNG  
UNTERBRECHEN

## Eingabe der Lösungshäufigkeiten für die ausgewählten Aufgaben

Bitte tragen Sie in die Spalte der jeweiligen Klasse Ihre Schätzungen ein, wie viele Schülerinnen und Schüler Ihrer Klasse die jeweilige Aufgabe lösen werden.

(Bitte nur die Schülerinnen und Schüler berücksichtigen, die in die Wertung eingehen.)

Aufgaben- nummer	Auswahl- schlüssel	Schule:	Schule:	Schule:
		007008	007008	007008
		Klasse:	Klasse:	Klasse:
		4a	4b	4c
		Anzahl:	Anzahl:	Anzahl:
		23	17	26
<u>1</u>	<u>Symmetrie (Geometrie)</u>	16	14	
<u>2</u>	<u>Umrechnungen mit Zeitmaßen</u> <u>(Sachrechnen/Größen)</u>			
<u>3</u>	<u>Stellenwertsystem (Arithmetik)</u>			
<u>4</u>	<u>Bruchteile (Sachrechnen/Größen)</u>			
<u>5</u>	<u>Einmaleins (Arithmetik)</u>			
<u>6</u>	<u>Schätzen (Standardverfahren)</u>			
<u>7</u>	<u>Daten gewinnen, darstellen,</u> <u>interpretieren</u>			
<u>8</u>	<u>Freie Auswahl</u>			
<u>9</u>	<u>Freie Auswahl</u>			
<u>10</u>	<u>Freie Auswahl</u>			



ABBRECHEN

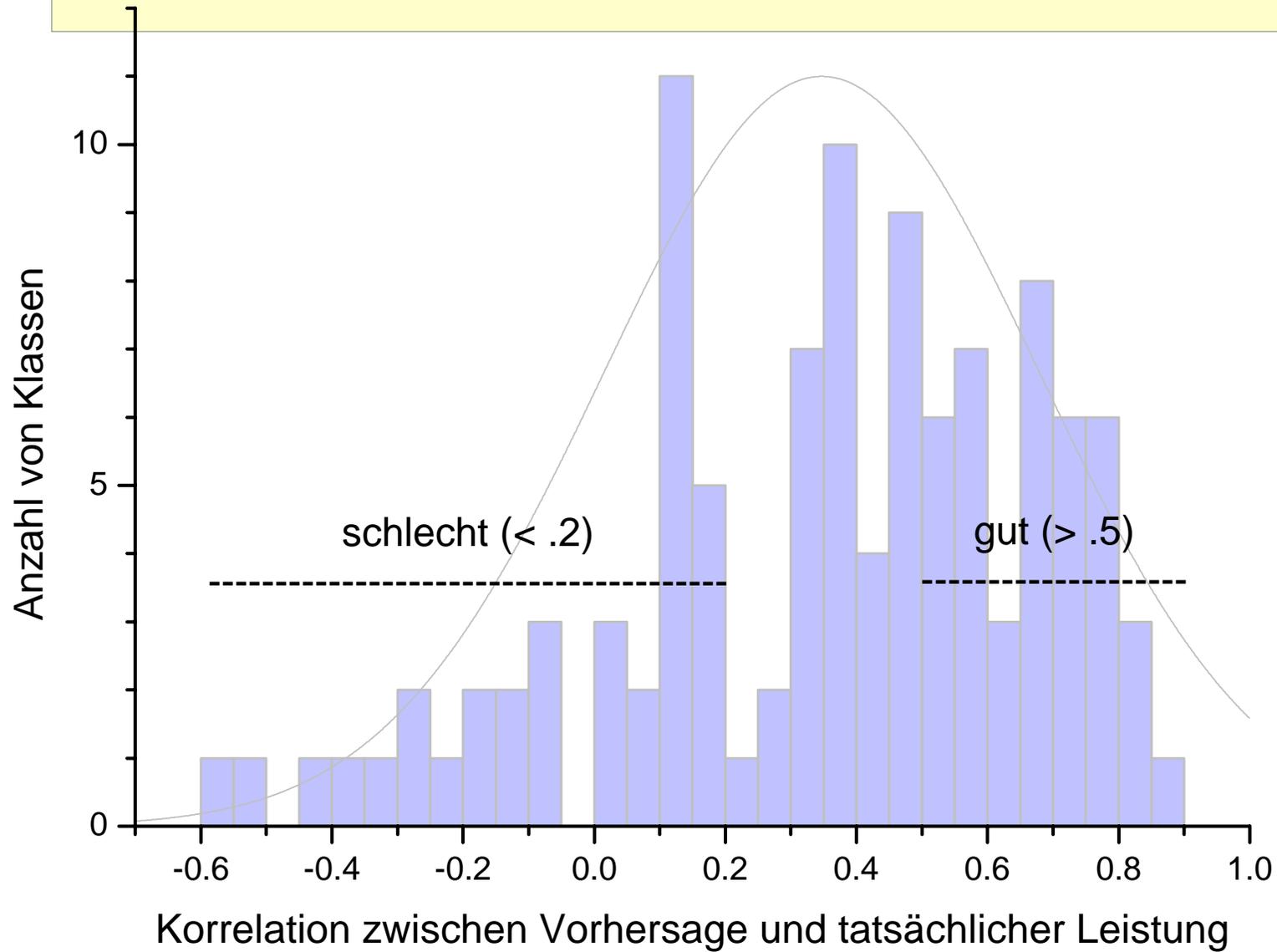


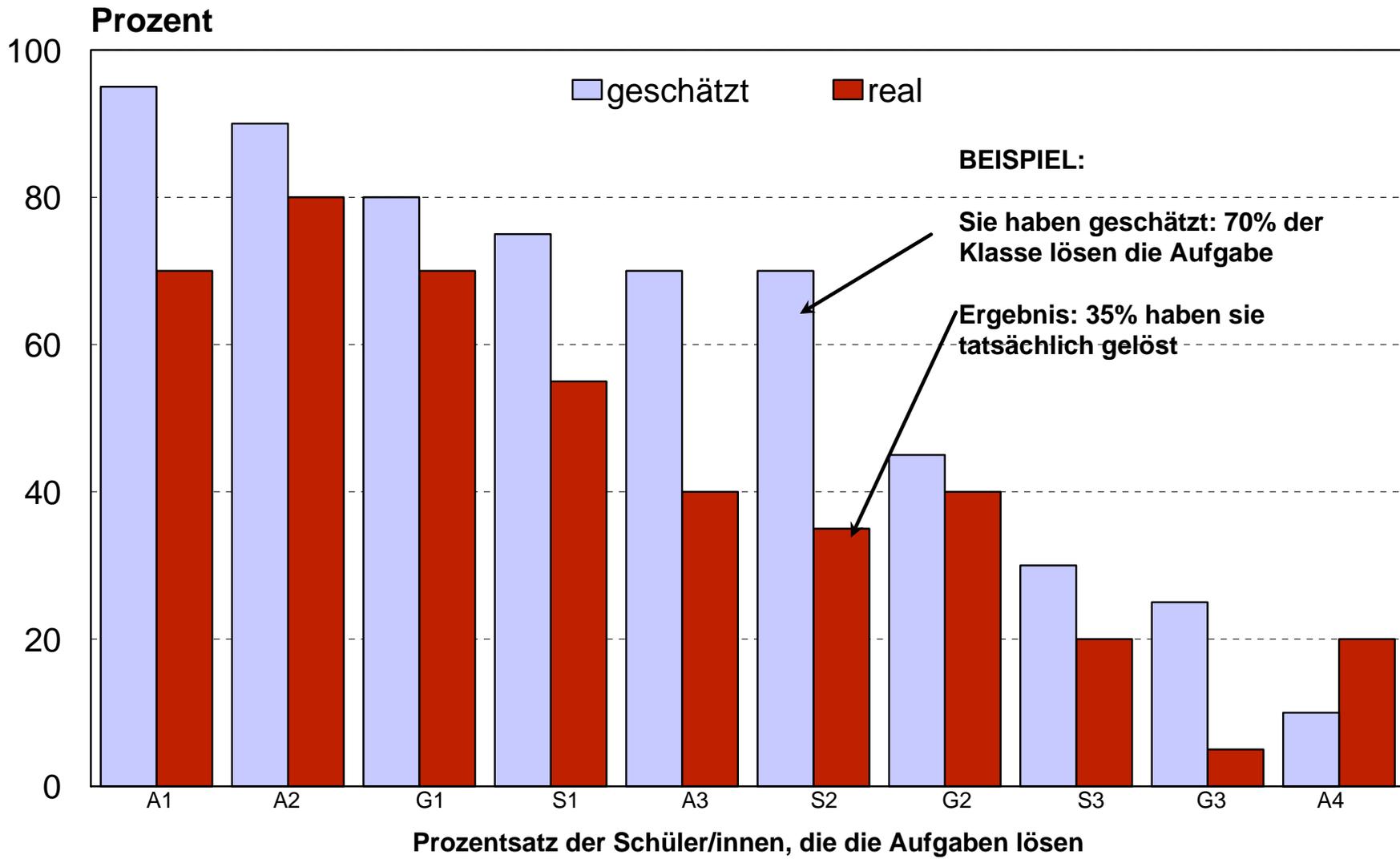
SEITE DRUCKEN



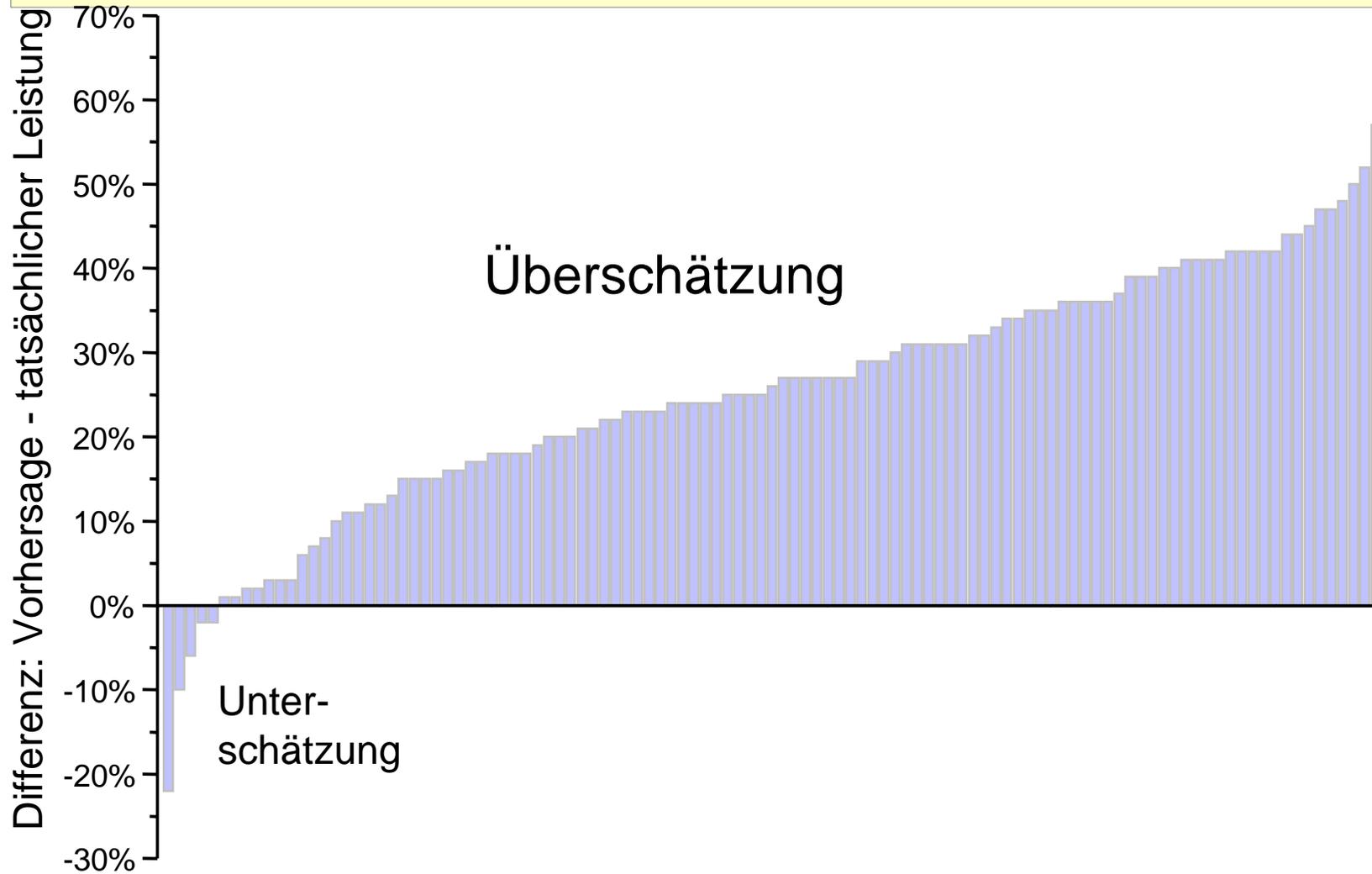
EINGABEN SPEICHERN

# Diagnosegenauigkeit I





# Diagnosegenauigkeit II



## Neue lerndiagnostische Aufgaben angesichts der Bildungsstandards

- Beschaffung, Einsatz und Auswertung standardsorientierter Tests (rechtzeitig – d.h. vor der eigentlichen Standardsüberprüfung)
- Verständnis für Kompetenzmodelle entwickeln (z.B. Merkmale, die Aufgaben leicht oder schwer machen)
- Selbst kompetenzorientierte Aufgaben entwickeln
- Nutzung von Mehrfachkorrekturen
- Fehlertypen und –muster analysieren
- Tests wiederholen

# Fragen von ARGEV

# ARGEV-Fragen

Primäre *Ziele* einer wirkungsorientierten und standard-basierten Systemsteuerung?  
Welche *Bedingungen* müssen erfüllt werden, damit diese Ziele erreicht werden können?

- Sicherung eines Sockelniveaus an Kompetenzen für möglichst alle Schüler; Bildungsgerechtigkeit; Transparenz
- Abgestimmtes Vorgehen in der Lehreraus- und Fortbildung; rechtzeitige Schaffung effizienter Unterstützungssysteme
- Evaluation der Evaluation: Pilotierung, empirische Überprüfung verschiedener Steuerungsvarianten, wissenschaftliche Begleitung

# ARGEV-Fragen

Soll Leistungsüberprüfung primär als ergebnis-orientierte *Selbst- oder Fremdevaluation* konzipiert werden?

- Selbstevaluation ist unabdingbar, reicht jedoch nicht aus
- Wichtig: Berücksichtigung der von den Bildungsstandards nicht direkt erfassten Klassenstufen (als Basis für gezielte Förderung in Richtung Standards)
- Fremdevaluation: dosiert (nicht notwendigerweise flächendeckend, umfassend und jährlich), abgestuft, kontrolliert

# Resümee: Koppelung von Leistungsmessung & Unterrichtsentwicklung: Bedingungen und Grenzen

- Allgemein: Stärkung förderlicher, Abbau hemmender kontextueller und individueller Bedingungen (s. Modell)
- Fachdidaktisch gehaltvolle Ergebnisrückmeldung + Arbeitsmaterial für Anknüpfungen + Handreichungen
- Schulinterner Support: Führungsrolle der Schulleitung; verbindliche Zielvereinbarungen; regelmäßige Kooperation; Vernetzung mit unterrichtsbezogenen Projekten (wie Team Teaching, Unterrichtsvideografie, Schülerfeedback); Best Practice-Konzepte: Wie gehen nachweislich erfolgreiche Klassen mit Evaluation um?
- Schulexterner Support: Kompetente Unterstützung durch Moderatoren, Lehrerfortbildung; Anerkennung & Hilfe durch die Schulaufsicht; Anreizsysteme; ausreichende Rahmenbedingungen
- Gefahren, Risiken: Abwertung anderer Fächer, nicht-kognitiver und nicht ökonomisch testbarer Kompetenzen

# ARGEV-Fragen

Gibt es bezogen auf die Leistungsmessung eine *sinnvolle Koppelung* zwischen standard-basierter *Selbstevaluation* und *Formen der Fremdevaluation* (externe Schulevaluation, standardisierte Tests)?

- Externe Evaluation als Motor und als Hilfe zur Selbsthilfe. Aber: Methodische Gütekriterien der externen Leistungsmessung sind schulintern unerreichbar
- Entwicklung eines zentralen Pools fachdidaktisch geprüfter und normierter (geeichter) Testaufgaben, die verschiedene Kompetenzstufen repräsentieren
- Systematische Nutzung moderner Technologien (Internet, Datenbank, Transferprogramme) zur Ermöglichung kontinuierlicher Selbstevaluation: Aufgabe des Bildungsmonitoring (siehe VERA)

ANDREAS HELMKE

ANDREAS HELMKE

UNTERRICHTS  
**QUALITÄT**

ERFASSEN • BEWERTEN • VERBESSERN

UNTERRICHTSQUALITÄT ERFASSEN • BEWERTEN • VERBESSERN

Kallmeyer 