



Grundlagen der Psychologie des Lehrens und Lernens

Psychologische Grundlagen für motivierende Bewertung

Susanne Narciss
TU Dresden

Grob-Übersicht Vorlesungstermine

Datum	Inhalt
20.10.2015	Organisatorisches - Zielreflexion - Einführung
27.10.2015	Motivation – integratives Rahmenmodell
03.11.2015	Leistungsmotivation – Stolz auf die eigene Tüchtigkeit
10.11.2015	Anreizwerte in Lehr-Lernsituationen
17.11.2015	Erwartungen - Kompetenzeinschätzung
24.11.2015	Motivation durch Belohnung - Zensuren?
01.12.2015	Lehrerverhalten als Motivationsquelle - BNO
08.12.2015	Bedingungen für eine motivierende Leistungsbewertung
15.12.2015	Loben – Feedback – ITFL-Modell
05.01.2016	Feedback geben in Lehr-Lernsituationen
12.01.2016	Interaktion- und Kommunikation - Lehrerrolle
19.01.2016	Lehrererwartungen – Pygmalioneffekte im Unterricht
26.01.2016	Lehrerkognitionen und –handeln im Unterricht ==> Training 18:30
02.02.2014	Inhaltliche Fragen
09.02.2014	Klausur zu Wintersemester

- **Leistungsbewertung – Diagnose – Diagnostischer Prozess**
 - ▶ Begriffsbestimmung
 - ▶ Diagnostische Kompetenz von Lehrpersonen
 - ▶ Diagnosegenauigkeit in Lehr-Lernsituationen - Befunde

- **Fehlerquellen im Diagnoseprozess - Testgütekriterien**
 - ▶ Validität
 - ▶ Reliabilität
 - ▶ Objektivität

- **Bezugsnormen reflektiert vor dem Hintergrund der Gütekriterien**

- **Bewerten von Lernergebnissen bei Gruppenleistungen**
 - ▶ Determinanten für Gruppenleistung
 - ▶ Mess- und Bewertungsprozess für Gruppenleistungen

■ Diagnose

- ◆ **griechisch** διάγνωσις, *diagnosis* 'Unterscheidung, Entscheidung', aus διά-, *diá-*, 'durch-' und γνώσις, *gnósis*, 'Erkenntnis, Urteil'
- ◆ auf der Grundlage von Erkenntnis eine Unterscheidung oder Entscheidung vornehmen

■ Diagnostik (Schmitt & Gerstenberg, 2014)

- ▶ theoriegeleitet & systematisch hochwertige Informationen sammeln
- ▶ über ein diagnostisches Objekt (z.B. Schüler, Schulklassen, Schulen, Unterricht, Bildungskontext)
- ▶ Ziele:
 - ◆ relevante Merkmale des Objekts möglichst genau beschreiben
 - ◆ künftige Merkmale des Objekts möglichst genau vorhersagen
 - ◆ Maßnahmen (z.B. Instruktionen, Interventionen) bestimmen, um
 - erwünschte Merkmale des Objekts herbeiführen / bewahren
 - unerwünschte Merkmale des Objekts verhindern / beseitigen

Diagnose - Diagnostik – Diagnostiker

Begriffsbestimmung

Diagnostizieren

- **Diagnose-Fragen / Aufgaben**

- ▶ Beschreibung
- ▶ Prognose
- ▶ Selektion, Platzierung
- ▶ Intervention, Instruktion

- **Sammeln von Informationen**

- ▶ theoriegeleitet & systematisch
- ▶ relevante Merkmale
- ▶ Gütekriterien

- **Nutzung der Informationen, um Entscheidung zu fällen zu D-Fragen**

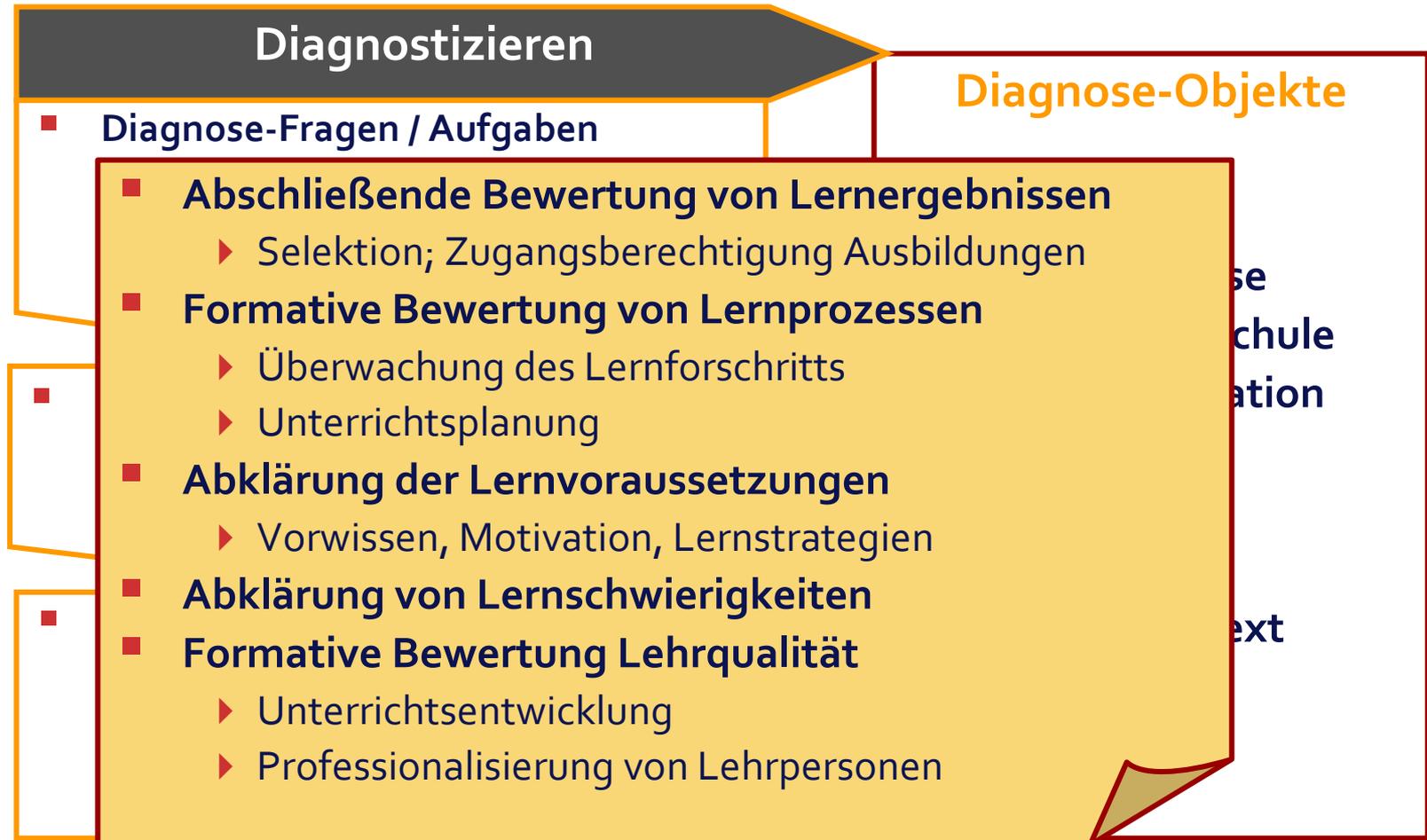
- ▶ reflektieren – Forschungsstand
- ▶ Kategorisieren & integrieren

Diagnose-Objekte

- **Person(en)**
- **Gruppe – Klasse**
- **Institution – Schule**
- **Lehr-Lernsituation**
- **Lernergebnis**
- **Lernprozess**
- **Lernstrategie**
- **Sozialer Kontext**
- **Lehrmaterial**
- **Lehrqualität**

Diagnostische Fragen in Lehr-Lernsituationen

Beispiele



se
chule
ation

ext

Diagnostische Kompetenz(en) Lehrpersonen

- **Diagnostische Kompetenz** (Schrader, 2013)
 - ▶ Fähigkeit einer Lehrperson, die im Lehrberuf anfallenden diagnostischen Aufgabenstellungen erfolgreich zu bewältigen,
 - ▶ Qualität der erbrachten Diagnoseleistungen
 - ▶ Fähigkeit, genaue diagnostische Urteile zu fällen

- **Diagnostische Kompetenz zentral für Lehrqualität**
 - ▶ Lehr-Lernziel erreicht? – Lehrer fällen etwa alle 2-3 min. ein Urteil zu dieser Frage (Shavelson & Stern, 1981)
 - ▶ Instruktionsbedarf? – Adaptivität von Unterricht zentraler Faktor für Lehrqualität (Helmke, 2010; 2012)
 - ▶ Feedback? – Formatives Feedback zentraler Faktor für Lehrqualität (z.B., Hattie, 2009)

- **Befunde zur Diagnosekompetenz im Sinne der Diagnosegenauigkeit**
 - ▶ Rangordnungen der Leistungen werden im Durchschnitt relativ genau geschätzt
 - ▶ Spannweite jedoch groß, d.h. zwischen einzelnen Lehrpersonen bzw. Klassen gibt es deutliche Unterschiede



Lehrperson Diagnostiker

Wissen/Kompetenzen

- fachlich
- fachdidaktisch
- psychologisch
- pädagogisch
- sozial

Einstellungen/Beliefs

- epistem. Überzeugungen
- pädagog. Orientierungen
- subjektive Theorien

Diagnostizieren

- Diagnose-Fragen / Aufgaben
 - ▶ Beschreibung
 - ▶ Prognose
 - ▶ Selektion, Platzierung
 - ▶ Intervention, Instruktion
- **Sammeln von Informationen über relevante Merkmale**
 - ▶ theoriegeleitet
 - ▶ systematisch
 - ▶ Gütekriterien
- Begründete Entscheidung fällen zu Diagnose-Fragen
 - ▶ Qualifikationsziel erreicht?
 - ▶ Instruktionsbedarf?

Diagnose-Objekte

- **Person(en)**
- Gruppe – Klasse
- Institution
- Lehr-Lernsituation
- **Lernergebnis**
- Lernprozess
- Lernstrategie
- Sozialer Kontext
- Lehrmaterial
- Lehrqualität

■ Testtheoretischer Hintergrund:

- ▶ Bei jeder Messung treten Messfehler auf
- ▶ die klassische Testtheorie unterscheidet daher
 - ◆ wahren Wert **T** (true value)
 - ◆ Messfehler - Fehleranteil **E** (error)
 - ◆ Beobachteter Messwert **X**, der sich zusammen setzt aus dem wahren Wert und Fehleranteilen

$$X = T + E + S$$

■ Unsystematische Messfehler – Zufallsfehler - E

- ▶ Fehler, die zu zufälligen Abweichungen vom Mittelwert führen

■ Systematische Messfehler - S

- ▶ Fehler, die zu systematischen Abweichungen vom Mittelwert führen

- **Unsystematische Messfehler – Zufallsfehler**
 - ▶ Fehler, die zu zufälligen Abweichungen vom Mittelwert führen
- **Systematische Messfehler**
 - ▶ Fehler, die zu systematischen Abweichungen vom Mittelwert führen
- **Frage zum selbst Denken:**
 - ▶ Welche Beispiele für systematische Messfehler gibt es in der Schule?

Fehlerquellen - Lehrpersonen

Beispiele für typische systematische Fehler - 1

- **"Halo" - engl. der "Hof" des Mondes**
Ein besonderes Merkmal beeinflusst die Beurteilung anderer Merkmale
 - ▶ Wer schön aussieht, wird auch als kompetent beurteilt
 - ▶ Schrift, Form
 - ▶ Primacy Effekt - Erster Eindruck dominiert (Jones et al. (1968))

- **Harari und Mc David (1973) :**
 - ▶ Aufsatzbeurteilung:
 - ◆ positiv beurteilte Namen (z.B. David, Michaela)
 - ◆ negativ beurteilte Namen
 - ▶ Noten im Durchschnitt um eine ganze Note unterschiedlich



Fehlerquellen - Lehrpersonen

Beispiele für typische systematische Fehler - 2

- **Urteilsfehler (z.B. Halo-Effekt)**
- **Erwartungs-, Pygmalioneffekte – Konsistenzeffekte**
 - ▶ Ignorieren erwartungsinkongruenter Informationen,
 - ◆ Fehler bei „leistungsstarken“ Schülern werden eher übersehen
 - ▶ Umdeuten diskrepanter Infos
 - ◆ Erwartungswidrige Leistungen => Kognitive Dissonanz
=> Attributionen so, dass Dissonanzreduktion stattfindet
 - ▶ Fokussieren auf erwartungskonforme Informationen

Systematische Fehler - Gütekriterien

Bedeutung systematischer Fehler für Validität

■ Systematische Messfehler

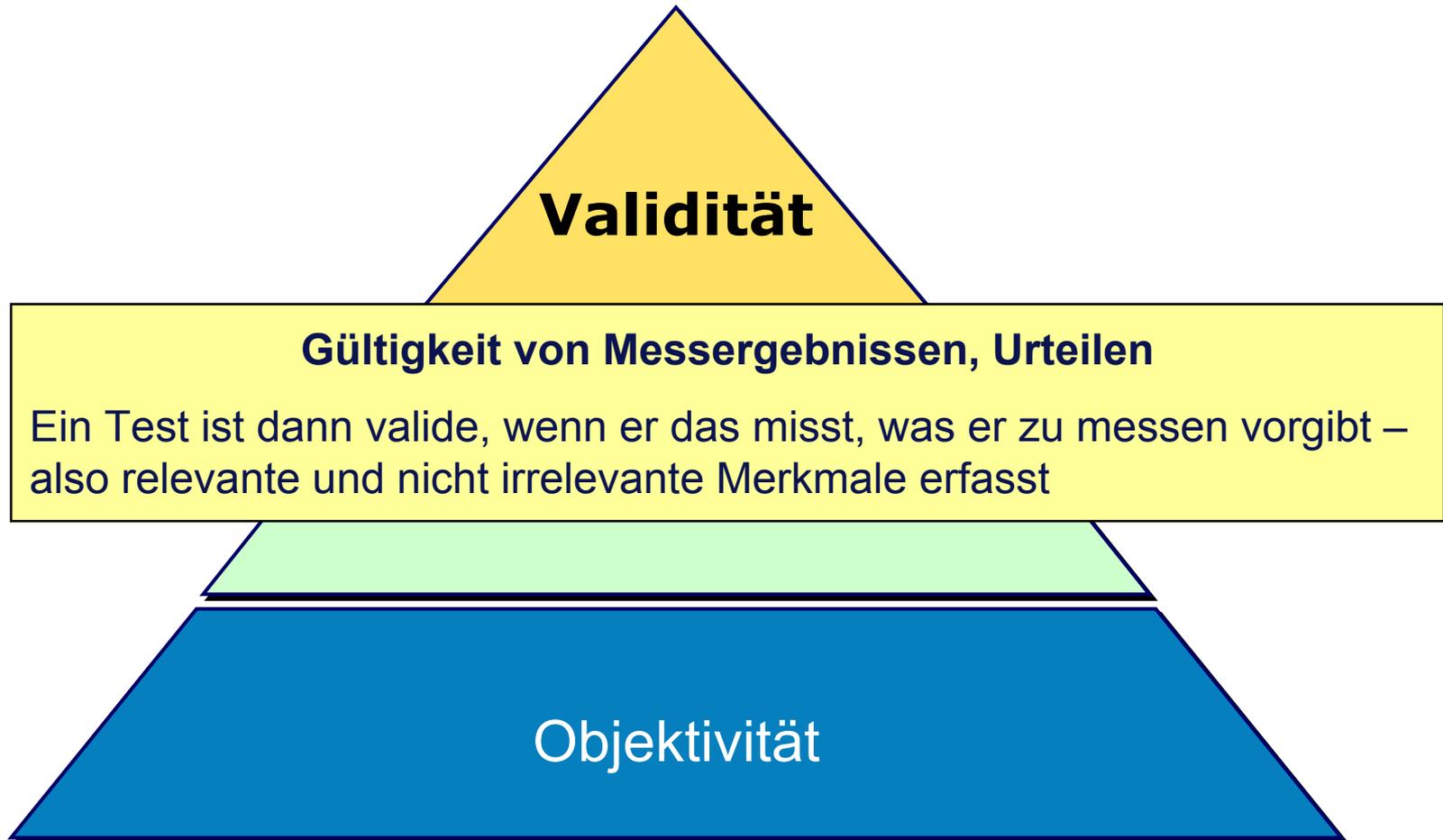
- ▶ neben relevanten Merkmalen werden auch irrelevante Merkmale beachtet
- ▶ Lassen sich nicht durch einen größeren Stichprobenumfang kontrollieren
- ▶ **Reduzieren die Validität der Messergebnisse**

■ Frage

- ▶ Was bedeutet Validität?

Hauptgütekriterien für Diagnostische Prozesse

Validität



Hauptgütekriterien für Diagnostische Prozesse

Validität

- Ein Test ist dann valide, wenn er das misst, was er zu messen vorgibt.
 - ▶ **Inhaltsvalidität – Kriteriumsvalidität**
 - ◆ Das Testergebnis erlaubt Schlussfolgerungen auf vorher festgelegte, inhaltlich begründete Kriterien.
 - ▶ **Übereinstimmungsvalidität**
 - ◆ Das Testergebnis stimmt mit einem Außenkriterium überein, z.B. mit dem Ergebnis eines anderen Tests, von dem man annimmt, dass er dasselbe Merkmal misst.
 - ▶ **Vorhersagevalidität**
 - ◆ Das Testergebnis lässt Vorhersagen auf künftige Ereignisse zu.
 - ▶ **Konstruktvalidität**
 - ◆ Das Testergebnis stimmt mit den theoretischen Vorhersagen zu dem zu messenden Konstrukt überein.

■ Systematische Messfehler

- ▶ neben relevanten Merkmalen werden auch irrelevante Merkmale beachtet
- ▶ **Reduzieren die Validität der Messergebnisse**

■ Hohe Validität einer Messung setzt voraus, dass

- ▶ Relevante Merkmale möglichst theoriegeleitet & fachlich begründet identifiziert wurden => Inhaltsvalidität – Konstruktvalidität
- ▶ **Identifizieren inhaltlich begründeter Merkmale mit Blick auf Lehrziele**
- ▶ möglichst wenig systematische Messfehler auftreten
- ▶ **Identifizieren von Quellen für systematische Fehler**

- **Zufallsfehler**

- ▶ Fehler, die zu unsystematischen Abweichungen vom Mittelwert führen

- **Frage zum selbst Denken:**

- ▶ Welche Fehlerquellen gibt es für zufällige Fehler?

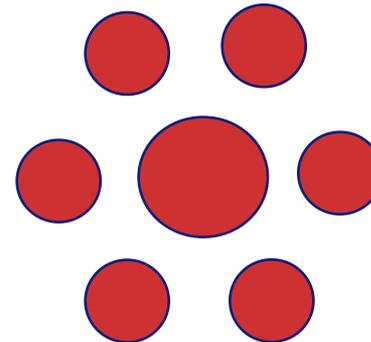
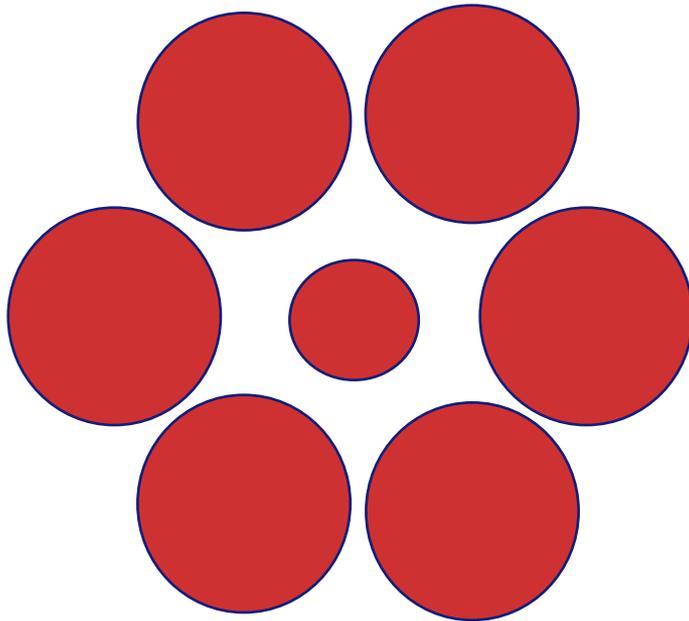
- **Lehrperson**

- ▶ Inkonsistentes Verhalten beim Auswerten und Bewerten
 - ◆ Z.B. Reihenfolge-Effekte

Fehlerquellen – Diagnostischer Prozess

Beispiele für typische systematische Fehler - 3

- **Lehrpersonen:**
 - ▶ Reihenfolge-Effekte: z.B. Kontrasteffekte



■ Reihenfolge-Effekte

- ▶ Kontrasteffekte
- ▶ Milde – Strenge – Milde
- ▶ Müdigkeitseffekte

Strategien entwickeln für das „Ausschalten“ dieser Quellen
z.B.

- „Quer“ Korrigieren, d.h. Aufgabe für Aufgabe
- Pausen beim Korrigieren machen

Zufallsfehler - Fehlerquellen

■ Lehrperson

- ▶ Reihenfolge-Effekte
- ▶ Flüchtigkeitsfehler

■ Diagnostisches Objekt - Schüler

- ▶ Schwankungen bei der Aufmerksamkeit der Schüler
- ▶ Flüchtigkeitsfehler
- ▶ Zufallstreffer – Raten
- ▶ nicht eindeutige Aufgabenformulierungen

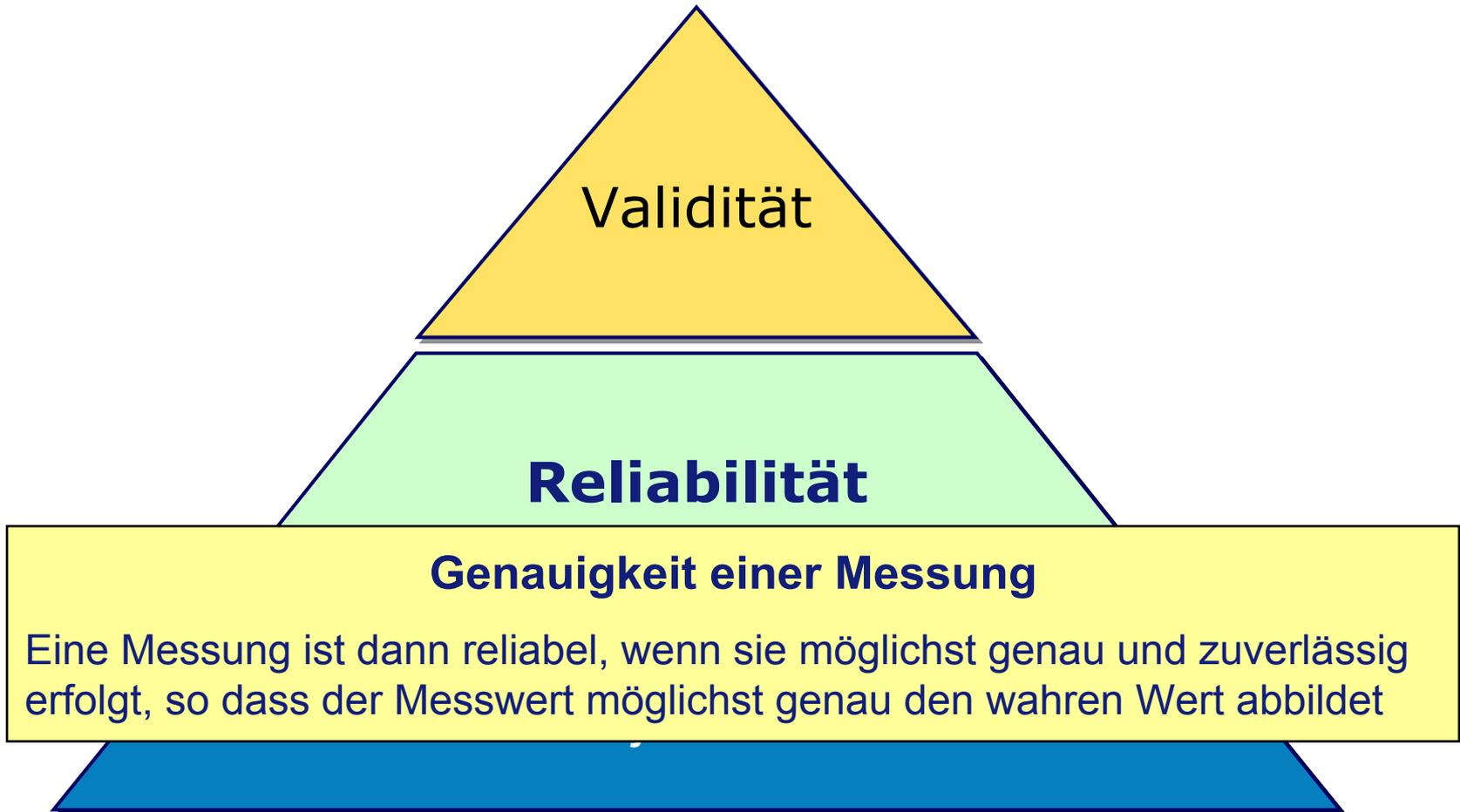
! **Zufallsfehler vermindern die Reliabilität einer Messung**

■ Frage:

- ▶ Was bedeutet Reliabilität?

Hauptgütekriterien für Diagnostische Prozesse

Reliabilität



Unsystematische Fehler - Reliabilität

Bedeutung unsystematischer Fehler - Implikationen



■ Testtheoretischer Hintergrund:

- ▶ bei jeder Messung treten Messfehler auf
- ▶ ein beobachteter Messwert X , sich zusammen aus dem wahren Wert T und einem Fehleranteil E
$$X = T + E + S$$
- ▶ wahrer Wert und Anteil unsystematischer Fehler sind unkorreliert
- ▶ unsystematische Fehler gleichen sich bei einer hohen Zahl wiederholter Messungen eines Merkmals aus
- ▶ je mehr Messdaten bzgl. eines Merkmals erfasst werden, desto kleiner ist der Mittelwert aller bei einer Messung gemachten unsystematischen Fehler

■ Hohe Reliabilität setzt voraus

- ▶ so viele Messdaten bei der Messung erhoben werden, dass sich Zufallsfehler „rausrechnen“

Je mehr klar gestellte Prüfungsaufgaben desto höher die Reliabilität

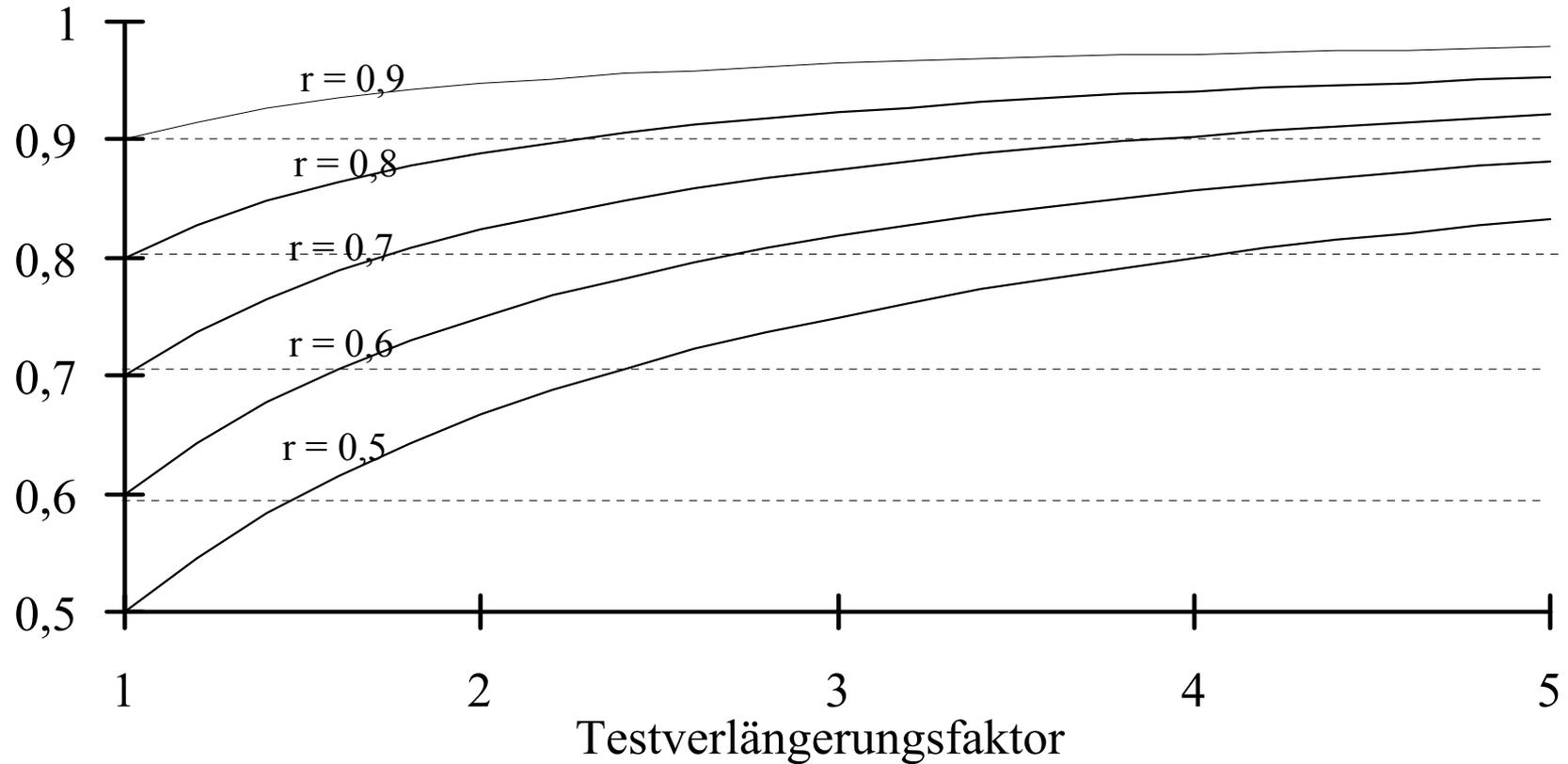
Zensuren sollten auf einer ausreichenden Zahl von Leistungsmessungen beruhen (mind. 5-6)

■ Wie kann man den Fehleranteil möglichst klein halten?

- ◆ Je größer die Anzahl der unabhängigen Messungen desselben Messobjekts desto geringer der Anteil zufälliger Fehler, denn es kommt zu einem Fehlerausgleich, da der Mittelwert des Fehleranteils umso kleiner wird, je mehr Messdaten vorliegen.
- ▶ Mehrere klar gestellte Prüfungsaufgaben sind besser als wenige Aufgaben
 - ◆ Kompensation von Zufallstreffern
 - ◆ Kompensation von Zufallsausrutschern

Max. Wert	Erreichter Wert	Wahrer Wert
10	7 [70%]	4 [40%]
20	14 [70 %]	11 [55%]
30	21 [70%]	18 [60%]
40	28 [70%]	25 [62,5%]
50	35 [70%]	32 [64,5%]

Reliabilität des verlängerten Tests



Aus: Rost, J. (1996) Lehrbuch Testtheorie Testkonstruktion. Bern: Huber

Effekte von Zufallsfehlern und systematischen Fehlern auf die Reliabilität - Validität

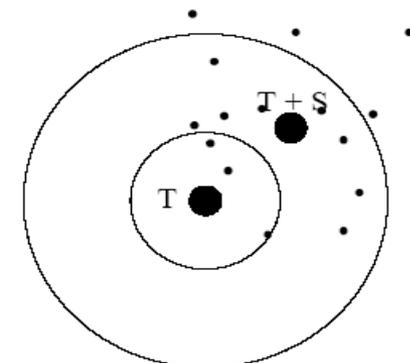
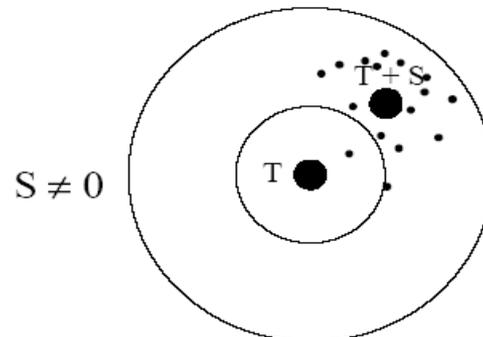
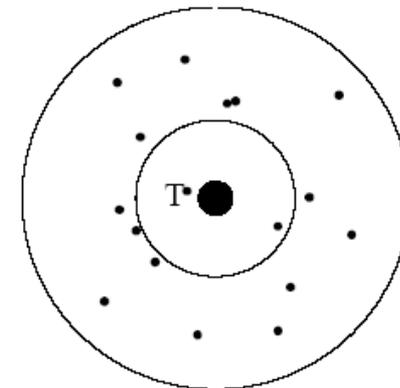
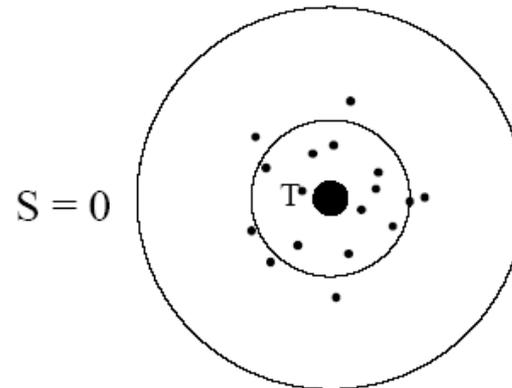
$$X = T + S + E$$

X = der beobachtete Wert •

T = der wahre Wert

S = systematischer Fehler

E = Zufallsfehler

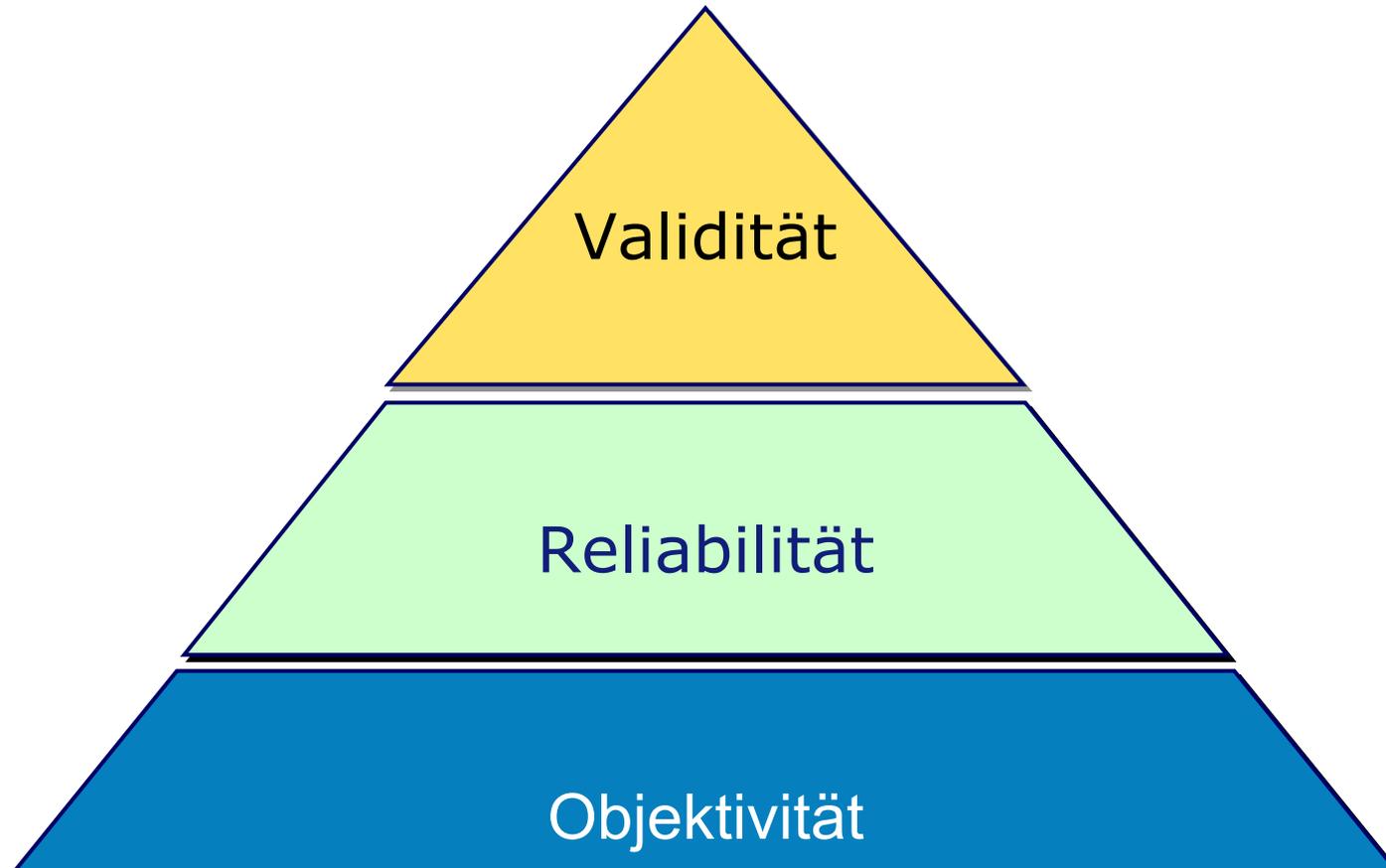


- **Fehler bei der Datenerhebung – Durchführung**
 - ▶ Lehrer gibt einzelnen Schülern Tipps
 - ▶ Bedingungen für den Einsatz unterschiedlicher Hilfsmittel sind nicht transparent
 - ▶ unterschiedliche Lehrer erlauben unterschiedliche Hilfsmittel
- **Fehler bei der Datenauswertung**
 - ▶ unscharfe Beurteilungskriterien
 - ▶ unerwartete Schülerantworten
 - ▶ inkonsistenter Umgang mit Zweifelsfällen
 - ▶ Unterschiedliche Gewichtung von Kriterien
- **Fehler bei der Interpretation der Messwerte**
 - ▶ unklare Gütemaßstäbe
 - ▶ Anwendung der sozialen Bezugsnorm

Fehler im diagnostischen Prozess vermindern die Objektivität

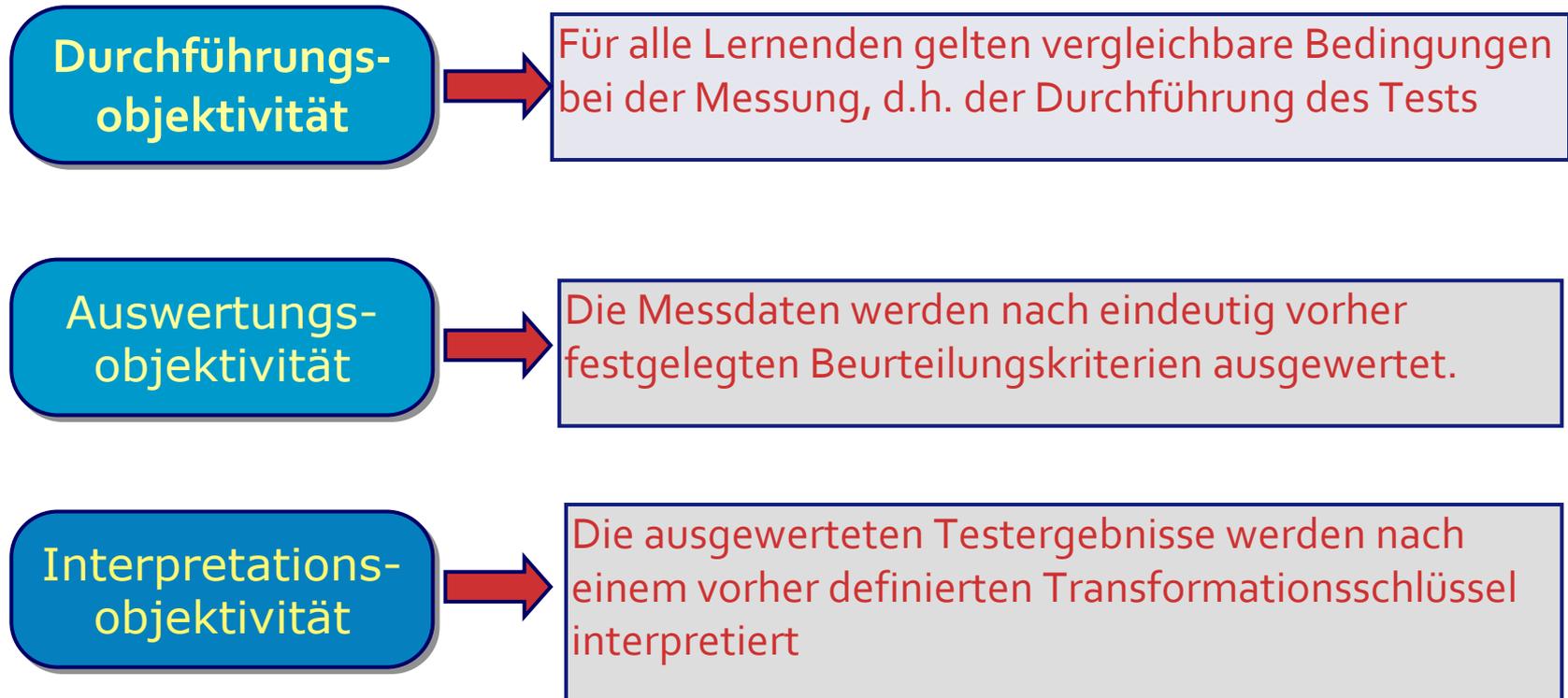
Hauptgütekriterien für Diagnostische Prozesse

Objektivität

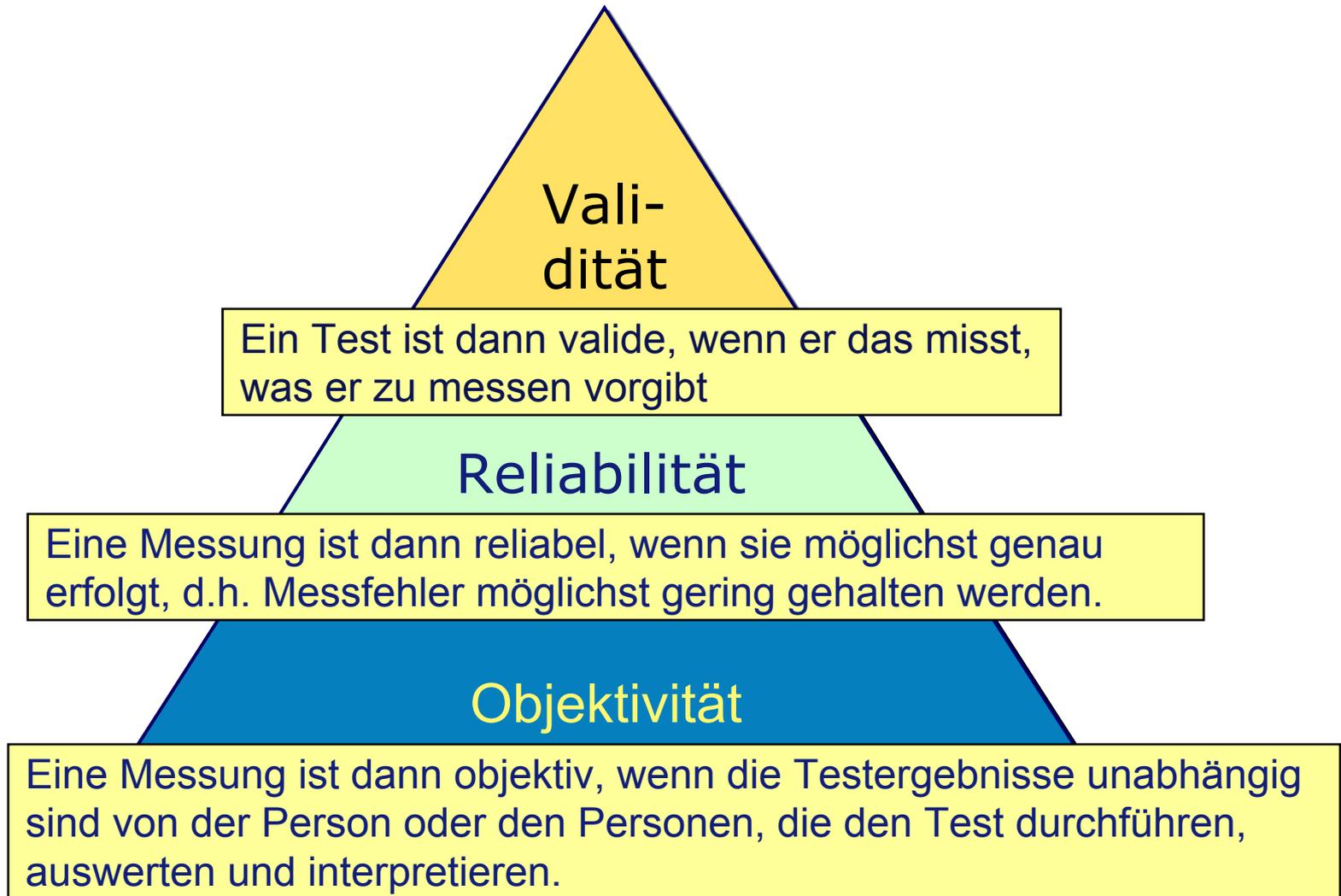


Eine Messung ist dann objektiv, wenn die Testergebnisse unabhängig sind von der Person oder den Personen, die den Test durchführen, auswerten und interpretieren.

Hohe Objektivität setzt voraus:

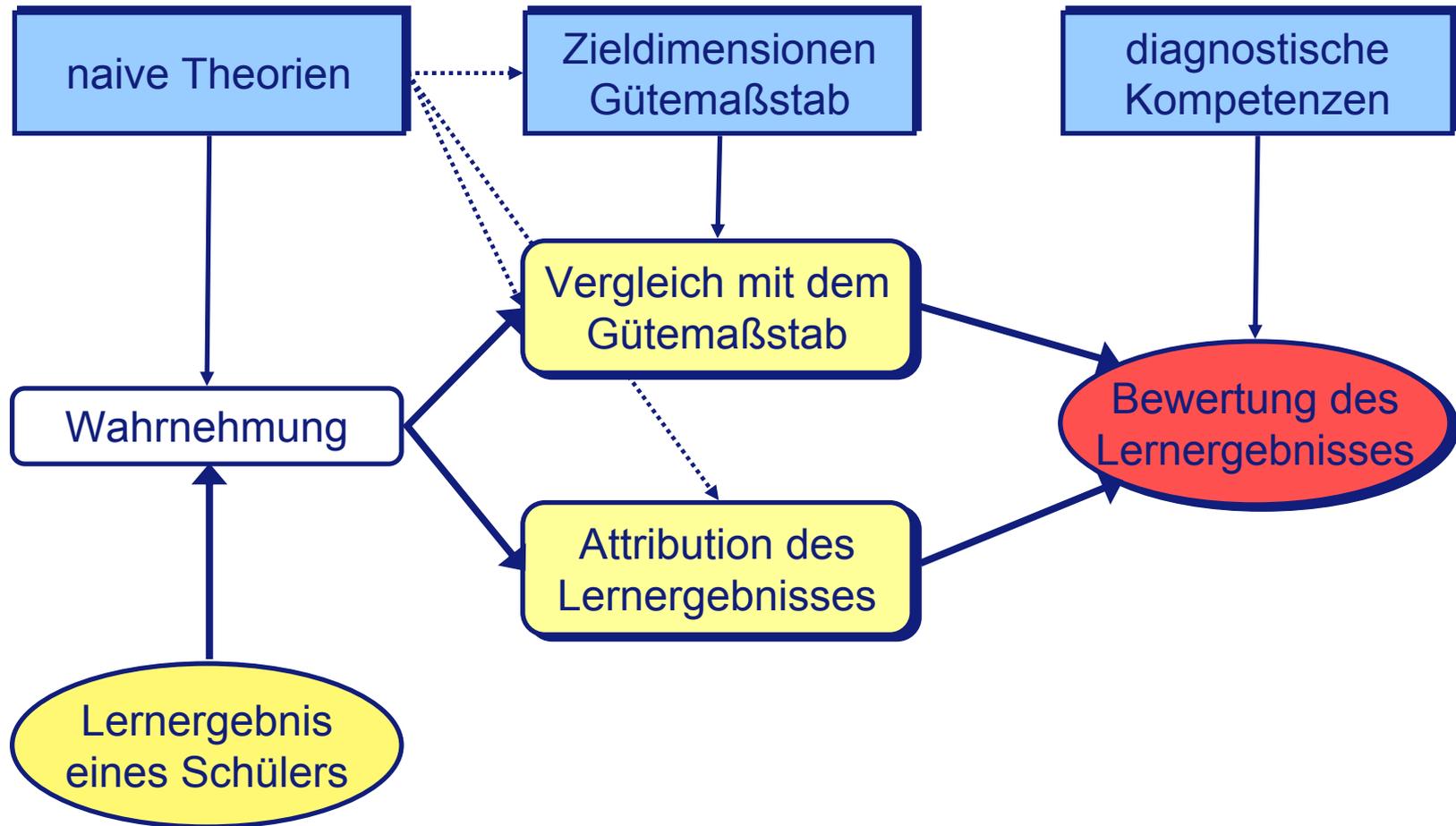


Hauptgütekriterien der Leistungsbewertung



Messfehlerquellen - Lehrperson

Einflussfaktoren - Bewertung von Lernergebnissen



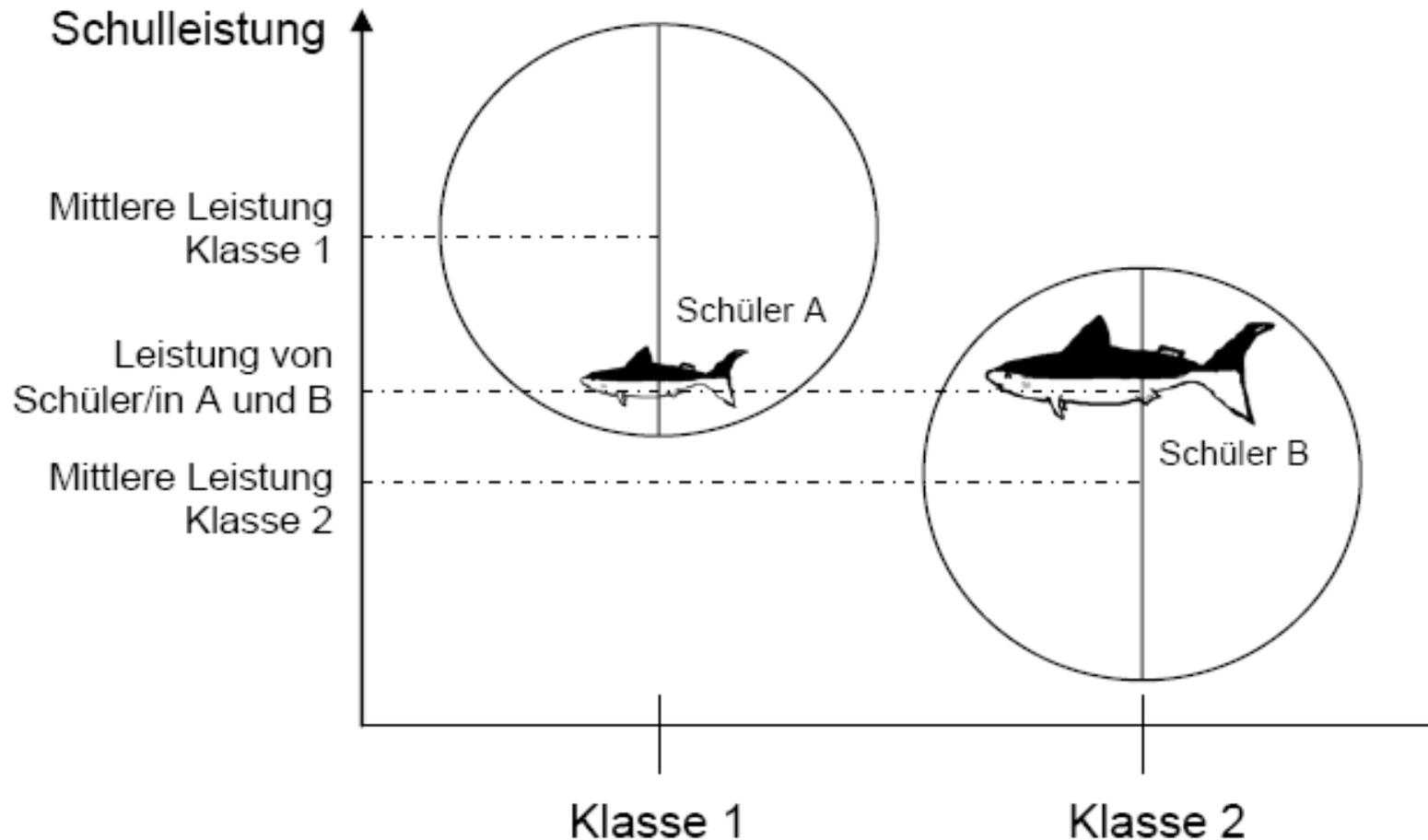
Leistungsbewertung und Motivation

Bezugsnormen – Gütekriterien

- Soziale Bezugsnorm:
- Frage zum Selbst denken:
 - ▶ Lassen sich mit der sozialen Bezugsnorm die Gütekriterien erfüllen?
 - ◆ Objektivität (Durchführung, Auswertung, Interpretation)
 - ◆ Reliabilität
 - ◆ Validität

Genese akademisches Selbstkonzept

Big fish little pond – effect (aus Köller, 2004)



Vergleichbarkeit von Zensuren

		Note I	Note II	Note III	Note IV	Note V	Note VI
Gruppe 1	Fehler	0 - 1	1,5 - 3,5	4 - 12	12,5 - 20	20,5 - 22,5	23 - 24
Gruppe 2	Fehler	0	1 - 2	2,5 - 6,5	7 - 9,5	10 - 11	11,5
Gruppe 3	Fehler	0	1 - 2	2,5 - 6,5	7 - 9,5	10 - 11	11,5
Gruppe 4	Fehler	0	1 - 2	2,5 - 8	8,5 - 14,5	15 - 16	16,5
Gruppe 5	Fehler	0	1 - 2	2,5 - 5	5,5 - 9	9,5 - 10,5	11
Gruppe 6	Fehler	0 - 1	1,5 - 3	3,5 - 6	6,5 - 9	9,5 - 11	11,5 - 12,5
Gruppe 7	Fehler	0	1,5 - 3	3,5 - 6	6,5 - 9	9,5 - 11	11,5 - 12,5
Gruppe 8	Fehler	0	1 - 3,5	4 - 9	9,5 - 13,5	14 - 16,5	17
Gruppe 9	Fehler	0	0	0 - 2	2,5 - 4	4,5 - 5	5,5

Schiefele schreibt dazu: „Das heißt: ein Schüler, der 5 Fehler machte, bekommt in einer Klasse die Note 5, in einer zweiten die Note 3. In einer Klasse reicht die Note 3 von 4 bis 12 Fehlern, in einer anderen beginnt sie schon bei 0 Fehlern. In Gruppe 8 bekommt man mit 3 Fehlern noch eine 2, während man in der Gruppe 9 für dieselbe Leistung bereits eine 5 einstecken muss.“

s. auch http://www.mpib-berlin.mpg.de/Pisa/PISA-E_Vertief_Zusammenfassung.pdf

Leistungsbewertung und Motivation

Bezugsnormen – Gütekriterien - Motivation

■ Soziale Bezugsnorm:

- ▶ Verstößt gegen Interpretations-Objektivität, also auch wenig reliabel und valide
- ▶ Motivational ungünstig sowohl für Leistungsstarke, als auch für Leistungsschwache

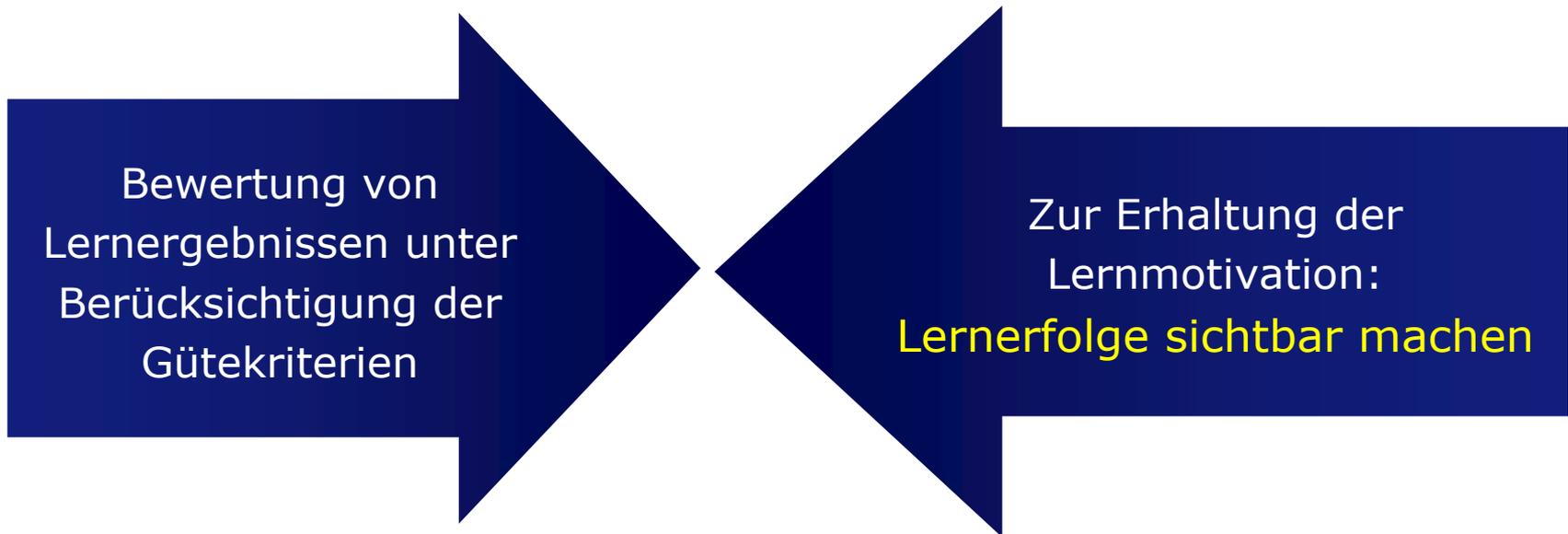
■ Individuelle Bezugsnorm:

- ▶ Gütekriterien können dann erfüllt werden, wenn transparent ist, auf welche Weise die Veränderungen in die Bewertung einbezogen werden
- ▶ Motivational am Günstigsten, da Lernfortschritte sichtbar werden

■ Sachliche Bezugsnorm:

- ▶ Gütekriterien können erfüllt werden
- ▶ Transparenz der Bewertungskriterien und Transformationschlüssel
- ▶ Motivational fast genauso günstig wie die individuelle Bezugsnorm

- Ein Dilemma?



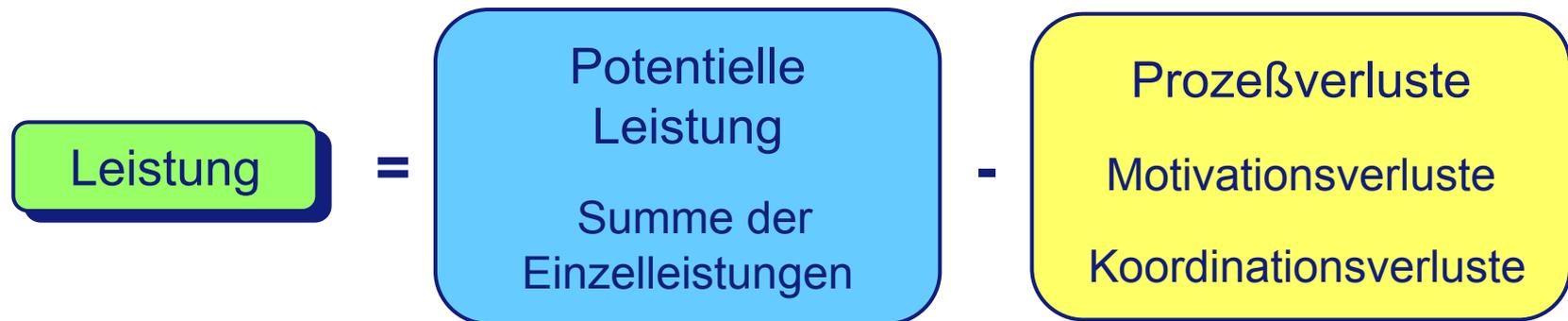
- Kombination kriteriumsorientierter & die individueller BN

- **Voraussetzungen für eine motivierende Leistungsbeurteilung**
 - ▶ Testgütekriterien
 - ▶ Bezugsnormen reflektiert vor dem Hintergrund der Gütekriterien

- **Bewerten von Lernergebnissen bei Gruppenleistungen**
 - ▶ Begriffsbestimmung Gruppenleistung
 - ▶ Determinanten für Gruppenleistung
 - ▶ Mess- und Bewertungsprozess für Gruppenleistungen
 - ▶ Koordinationsverluste
 - ▶ Motivationsverluste

Bewerten von Gruppenleistungen

Gruppenleistung

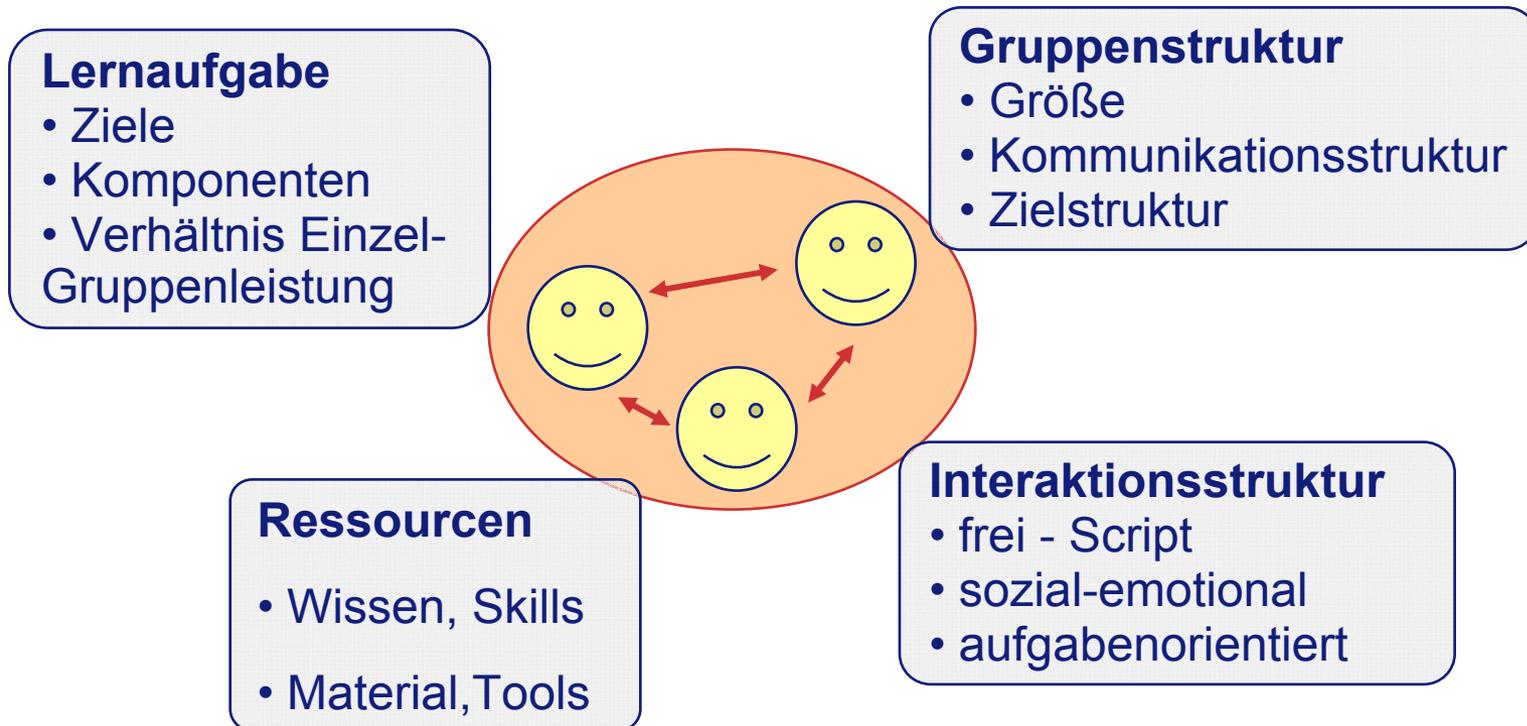


Bewerten von Gruppenleistungen

Determinanten für Gruppenproduktivität

Zum selbst Denken:

Wann gibt es negative Einflüsse der Faktoren?

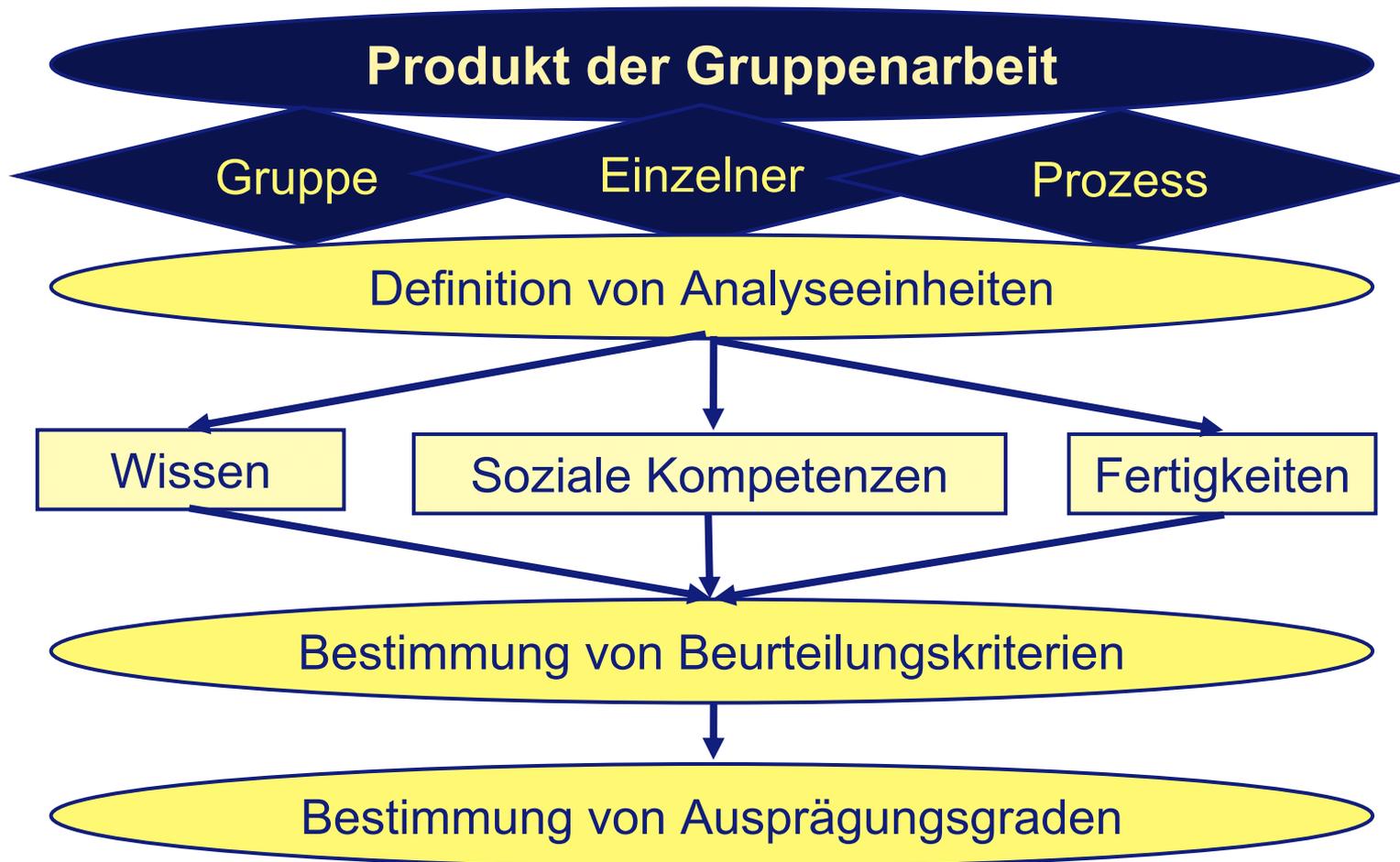


Bewerten von Gruppenleistungen

Bedingungen für effizientes Lernen in Kleingruppen

- **Gruppen-Aufgaben**
 - ▶ teilbar, so dass die spezifischen Fähigkeiten und Fertigkeiten der Gruppenmitglieder genutzt werden können
 - ▶ verschiedene Ressourcen einbeziehen
 - ▶ Instruktionen etc. schriftlich
- **Interaktionsstrukturierung**
 - ▶ Gruppengröße
 - ▶ Kommunikationsstruktur
 - ▶ evtl. Kommunikationsregeln
 - ▶ Regeln für Artikulation von Denkprozessen
- **Transparente Leistungsbewertung**
 - ▶ klare Kriterien
 - ▶ nicht nur Produkt, sondern auch Prozess

Differenzierte Bewertung von Gruppenleistungen



Bewertung von Gruppenleistungen

Illustration der Problematik anhand eines Beispiels

Aufgabe: Finden von Variationsmöglichkeiten

Gruppe: heterogen – 3 maximal

Vergleichen Sie die Produktbewertung mit den Bewertungen der Einzelleistungen.
Was fällt auf?

Noten der Experten- und Prozessbeurteilung:

Produktbewertung	Bewertung der Einzelleistungen
G1: sehr gut	S1: sehr gut S2: befriedigend S3: mangelhaft
G2: gut	S1: gut - befriedigend S2: gut S3: ausreichend
G3: sehr gut – gut	S1: befriedigend S2: sehr gut – gut S3: mangelhaft

Bewertung von Gruppenleistungen

Gruppen- und Einzelleistungen bewerten

Aufgabe: Finden von Variationen zu einem musikalischen Thema

Gruppe: heterogen – 3 max. 4 Personen

Noten der Experten- und Prozessbeurteilung:

Produktbewertung	Bewertung der Einzelleistungen	Bewertung von Gruppen- und Einzelleistung
G1: sehr gut	S1: sehr gut S2: befriedigend S3: mangelhaft	S1: sehr gut 1 S2: gut 2,3 S3: befriedigend 4,6
G2: gut	S1: gut – befriedigend S2: gut S3: ausreichend	S1: gut – 2,3 S2: gut 2 S3: befriedigend 3,3
G3: sehr gut – gut	S1: befriedigend S2: sehr gut – gut S3: mangelhaft	S1: gut - 2,8 S2: sehr gut – gut 1,5 S3: befriedigend - 3,8

Literaturempfehlungen

Grundlagen

- ▶ **Hany, E. (2008).** Pädagogisch-psychologische Diagnostik. In A. Renkl (Hrsg.) *Lehrbuch Pädagogische Psychologie* (S. 389-468). Bern: Huber.
- ▶ Rheinberg, F. & Krug, S. (2005). *Motivationsförderung im Schulalltag*. Göttingen: Hogrefe.

Originalpublikationen zum Vertiefen:

Krampen, G. (1987). Differential effects of teacher comments. *Journal of Educational Psychology, 79*, 137-146.

Interessante Links:

<http://www.prueferportal.org/html/765.php>

<http://www.foerdern-individuell.de/index.php?Seite=863&>

<http://www.guterunterricht.de/>

Vielen Dank für Ihre Aufmerksamkeit!