

# Gütekriterien: Validität (15.5.)

- Besprechung der Hausaufgaben
- Validität: Arten und Schätzmethoden
  - Inhaltsvalidität
  - Empirische Validität
  - Konstruktvalidität
  - Testfairness
- Normierung
- Nebengütekriterien
- Zusammenfassung

# Validität (Gültigkeit)

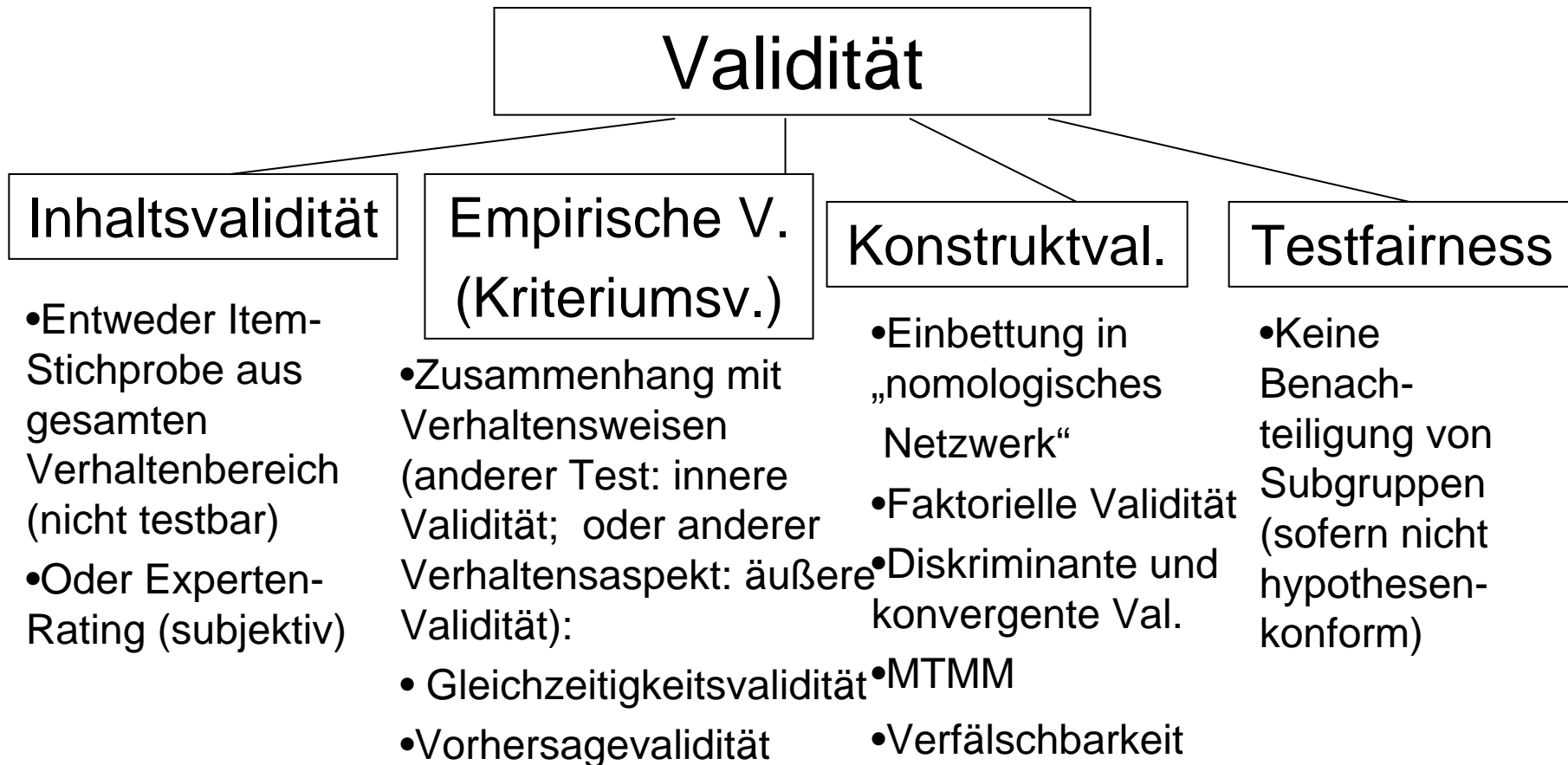
- Grad in dem ein Test das misst, was er messen soll

- Beispiel: Waage zur „Größenmessung“



# Validität (Gültigkeit)

- Grad, in dem ein Test das misst, was er messen soll



# Zur empirischen Validität: innere und äußere Validität, Gleichzeitigkeits- und Vorhersagevalidität

	gleichzeitig	zukünftig (prognostisch)
innere Validität		
äußere Validität		

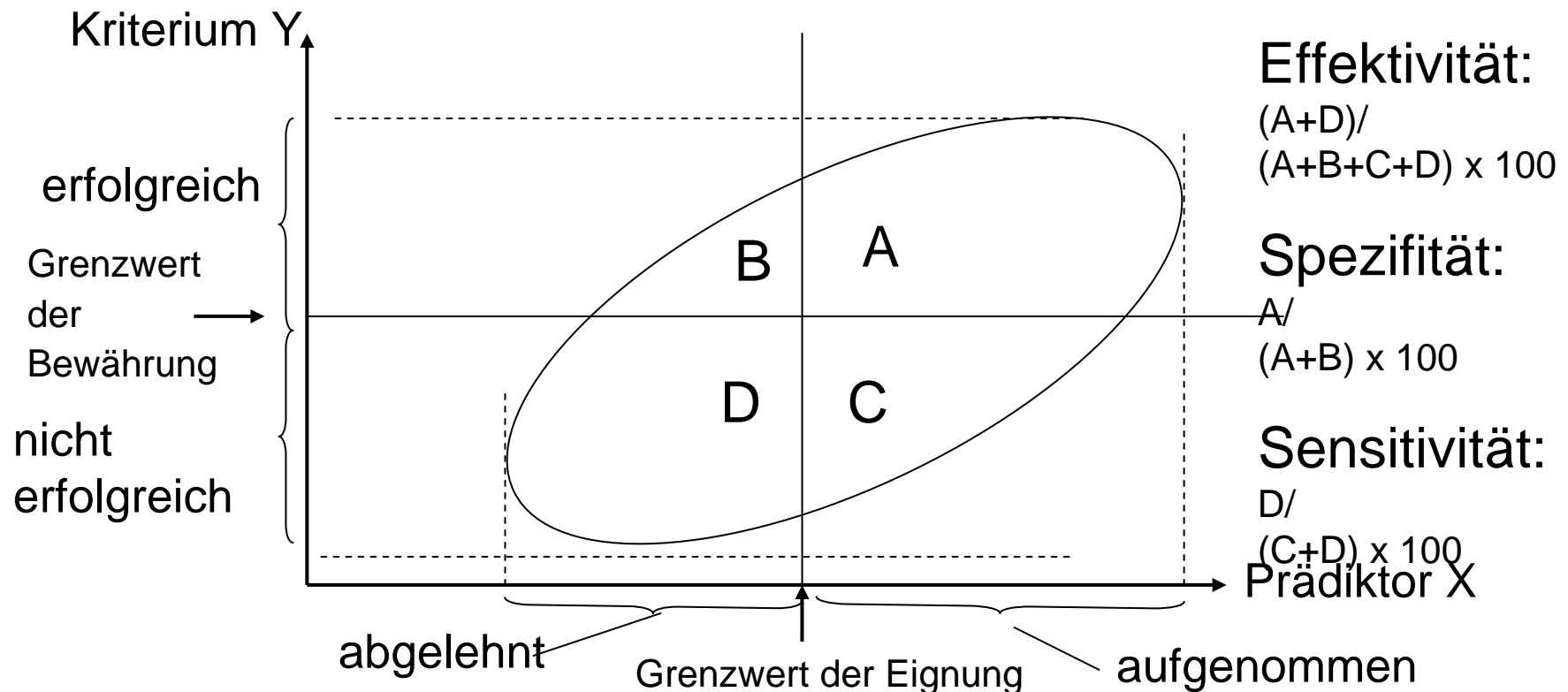
# Zur empirischen Validität: innere und äußere Validität, Gleichzeitigkeits- und Vorhersagevalidität

	gleichzeitig	zukünftig (prognostisch)
innere Validität	z.B. Korrelation der Skalen induktives Denken (Thurstone) und fluide Intelligenz (Cattell) beide gemessen zu t1	z.B. Korrelation der Skalen induktives Denken (Thurstone) gemessen zu t1 und fluide Intelligenz (Cattell) gemessen zu t2
äußere Validität	z.B. Korrelation der Skala induktives Denken (Thurstone) und Schulnoten beide gemessen zu t1	z.B. Korrelation der Skala induktives Denken (Thurstone) gemessen zu t1 und Examensnote gemessen zu t2

# Zur Vorhersage-Validität bei Selektionsentscheidungen

A = richtig Positive    B = falsch Negative

C = falsch Positive    D = richtig Negative



# Zur Konstruktvalidität: konvergente/ diskriminante Validität, multitrait- multimethod-Matrix (MTMM); Beispiel aus Fisseni

*Tabelle 4-12: Multitrait-Multimethod-Matrix*

I, II, III bezeichnen drei *Methoden*  
(Test, Exploration, Verhaltensbeobachtung)  
A, B, C bezeichnen drei *Merkmale*,  
deren jedes mit allen drei Methoden erfasst wird

		<i>I</i> TEST			<i>II</i> EXPLOR.			<i>III</i> VERHALT.		
		<i>A</i>	<i>B</i>	<i>C</i>	<i>A</i>	<i>B</i>	<i>C</i>	<i>A</i>	<i>B</i>	<i>C</i>
<i>I</i>	<i>A</i>	79								
	<i>B</i>	14	80							
	<i>C</i>	06	34	77						
<i>II</i>	<i>A</i>	48			73					
	<i>B</i>	12	46		01	77				
	<i>C</i>	07	19	51	09	12	81			
<i>III</i>	<i>A</i>	48			55			83		
	<i>B</i>	09	49		12	65		10	81	
	<i>C</i>	11	11	56	04	28	59	13	11	85

# Zur Konstruktvalidität: konvergente/ diskriminante Validität, multitrait- multimethod-Matrix (MTMM); Beispiel aus Fischni

*Tabelle 4-12: Multitrait-Multimethod-Matrix*

I, II, III bezeichnen drei *Methoden*  
(Test, Exploration, Verhaltensbeobachtung)  
A, B, C bezeichnen drei *Merkmale*,  
deren jedes mit allen drei Methoden erfasst wird

		I TEST			II EXPLOR.			III VERHALT.		
		A	B	C	A	B	C	A	B	C
I	A	79								
	B	14	80							
	C	06	34	77						
II	A				73					
	B	12	46		01	77				
	C	07	19	51	09	12	81			
III	A							83		
	B	09	49		12	65		10	81	
	C	11	11	56	04	28	59	13	11	85

Reliabilitäten



# Zur Konstruktvalidität: konvergente/ diskriminante Validität, multitrait- multimethod-Matrix (MTMM); Beispiel aus Fischni

Tabelle 4-12: Multitrait-Multimethod-Matrix

I, II, III bezeichnen drei *Methoden*  
(Test, Exploration, Verhaltensbeobachtung)  
A, B, C bezeichnen drei *Merkmale*,  
deren jedes mit allen drei Methoden erfasst wird

		I TEST			II EXPLOR.			III VERHALT.		
		A	B	C	A	B	C	A	B	C
I	A	79								
	B	14	80							
	C	06	34	77						
II	A	48			73					
	B	12	46		01	77				
	C	07	19	51	09	12	81			
III	A	48			55			83		
	B	09	49		12	65		10	81	
	C	11	11	56	04	28	59	13	11	85

konvergente  
Validitäten

# Zur Konstruktvalidität: konvergente/ diskriminante Validität, multitrait- multimethod-Matrix (MTMM); Beispiel aus Fiebert

Tabelle 4-12: Multitrait-Multimethod-Matrix

I, II, III bezeichnen drei *Methoden*  
(Test, Exploration, Verhaltensbeobachtung)  
A, B, C bezeichnen drei *Merkmale*,  
deren jedes mit allen drei Methoden erfasst wird

		I TEST			II EXPLOR.			III VERHALT.		
		A	B	C	A	B	C	A	B	C
I	A	79								
	B	14	80							
	C	06	34	77						
II	A	48			73					
	B	12	46		01	77				
	C	07	19	51	09	12	81			
III	A	48			55			83		
	B	09	49		12	65		10	81	
	C	11	11	56	04	28	59	13	11	85

diskriminante  
Validitäten  
bei gleicher  
Methode

# Zur Konstruktvalidität: konvergente/ diskriminante Validität, multitrait-multimethod- Matrix (MTMM); Beispiel aus Fisseni

Tabelle 4-12: Multitrait-Multimethod-Matrix

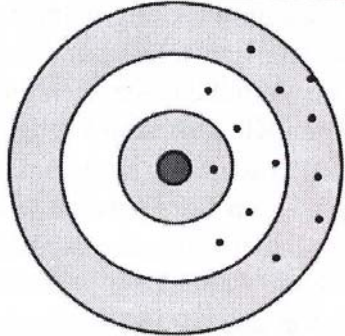
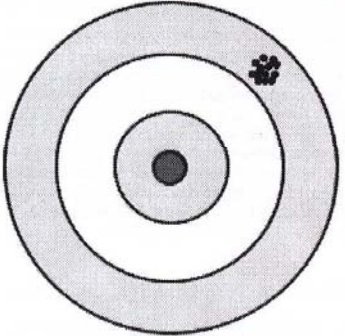
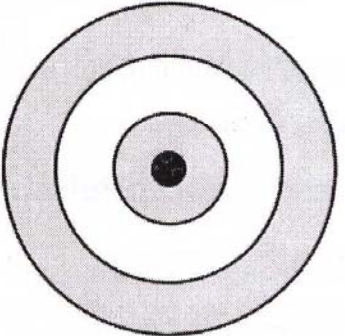
I, II, III bezeichnen drei *Methoden*  
(Test, Exploration, Verhaltensbeobachtung)  
A, B, C bezeichnen drei *Merkmale*,  
deren jedes mit allen drei Methoden erfasst wird

		I TEST			II EXPLOR.			III VERHALT.		
		A	B	C	A	B	C	A	B	C
I	A	79								
	B	14	80							
	C	06	34	77						
II	A	48			73					
	B	12	46		01	77				
	C	07	19	51	09	12	81			
III	A	48			55			83		
	B	09	49		12	65		10	81	
	C	11	11	56	04	28	59	13	11	85

diskriminante  
Validitäten bei  
verschiedenen  
Methoden

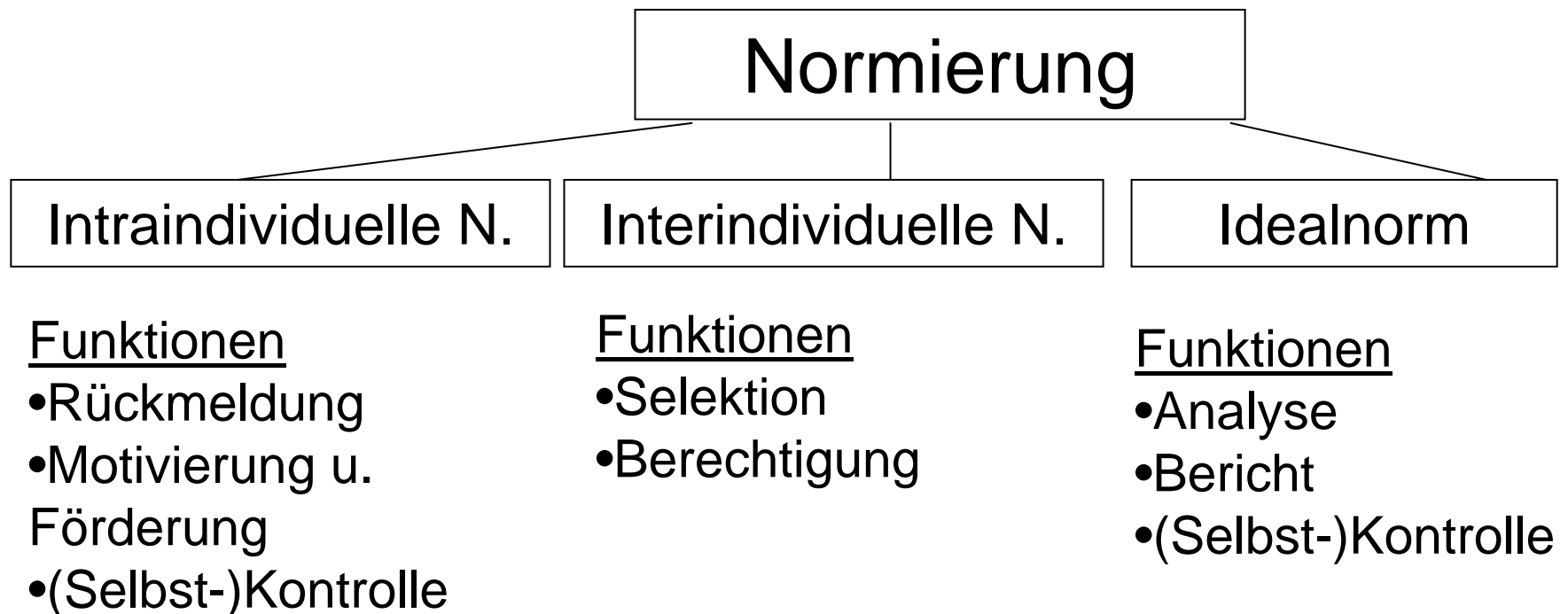
# Verhältnis von Reliabilität und Validität

- Reliabilität ist eine notwendige aber nicht hinreichende Bedingung für Validität!
- Ein Test kann nicht valider sein, als er reliabel ist

	niedrige Validität	hohe Validität
niedrige Reliabilität	 <p>Die Treffer liegen nicht im Zentrum und nicht dicht beieinander.</p>	
hohe Reliabilität	 <p>Die Treffer liegen dicht beieinander aber nicht in der Mitte.</p>	 <p>Die Treffer liegen dicht beieinander und in der Mitte.</p>

# Testnormierung

- Einordnung des Testergebnisses in Bezugssystem





# Zur Normierung: Einige Normierungen bei interindividueller Norm

- Rohwerte aus verschiedenen Tests lassen sich transformieren (und damit vergleichen)
  - Äquivalentnormen: z.B. bei Berechnung des IA (Binet)
  - Abweichungsnormen:  
z.B.  $IQ = 100 + 15 (x-M)/s$
  - Prozentränge:

*Tabelle 4-16: Berechnung von Prozenträngen*

Vorgegeben seien

- Rohwerte von 2 bis 12 (X),
- Häufigkeit der Probanden je Rohwertklasse (f),
- kumulierte Häufigkeiten (cumf).

Berechnet werden Prozentränge (PR).

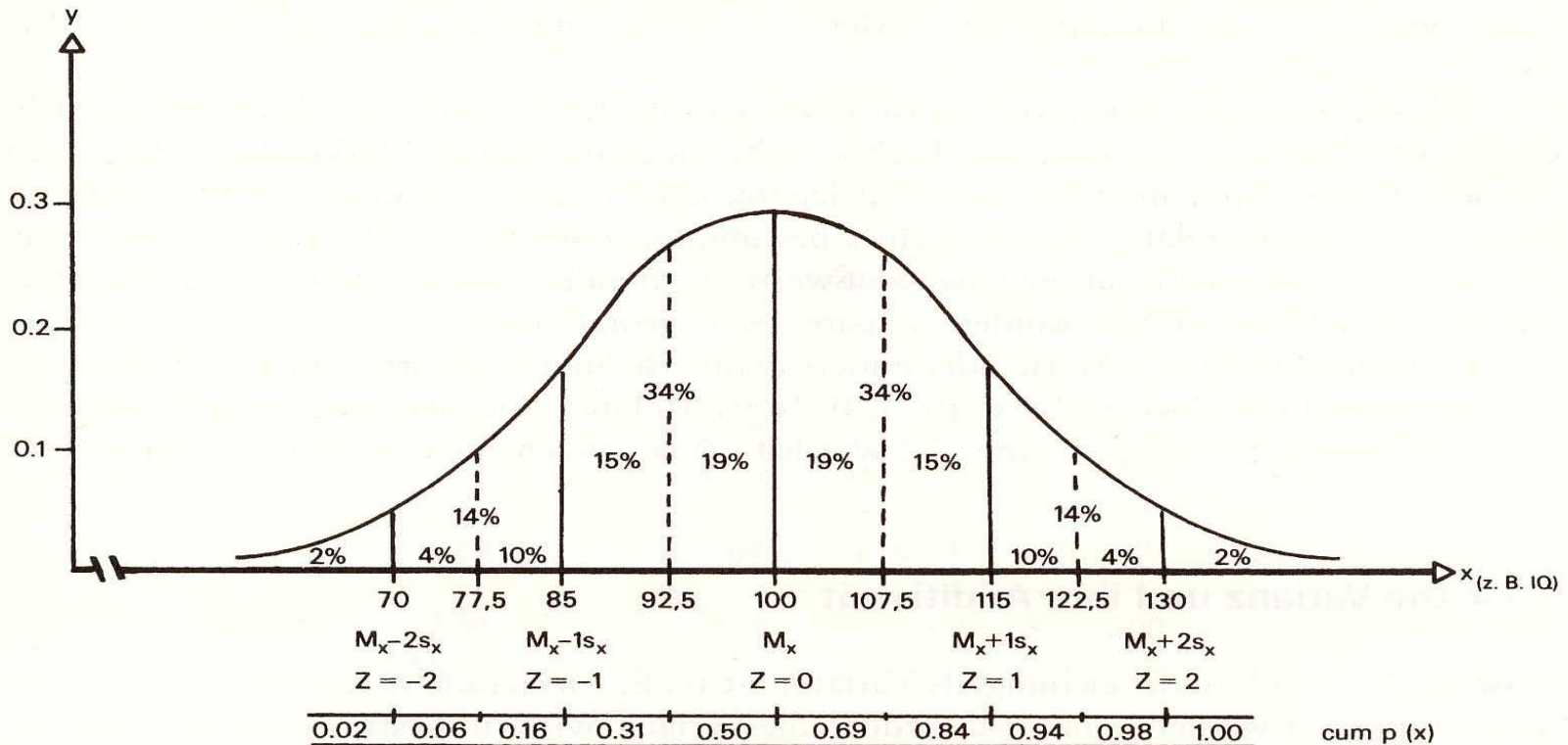
X	2	3	4	5	6	7	8	9	10	11	12
f	11	21	26	32	38	44	40	33	27	19	12
cumf	11	32	58	90	128	172	212	245	272	291	303
PR	4	11	19	30	42	57	70	81	90	96	100

Für die Rohwertklasse X = 4 berechnet sich der Prozentrang wie folgt:

$$PR_4 = (58 / 303) \times 100 = 19.14$$

Der Maßzahlklasse 4 kommt der PR 19 zu, der besagt: Rund 19 Prozent der 303 Probanden erreichen 4 Punkte oder weniger, rund 80 Prozent erreichen mehr.

# Beispiel: Theoretische Verteilung der Intelligenz



**Abb. 7.2:** Flächenproportionen unter der Normalverteilungskurve für eine Variable mit dem arithmetischen Mittel  $M_x = 100$  Einheiten und einer Standardabweichung von  $s_y = 15$  Einheiten. In der unteren Hälfte der Graphik sind die zugehörigen kumulierten Wahrscheinlichkeiten  $\text{cum } p(x)$  eingetragen, die angeben, wie groß die Wahrscheinlichkeit für den entsprechenden X-Wert *oder* einen kleineren Wert ist.

# Nebengütekriterien

- Ökonomie
- Nützlichkeit und Zumutbarkeit
- Vergleichbarkeit
- Akzeptanz



# Zusammenfassung der Gütekriterien

- Objektivität, Reliabilität (Zuverlässigkeit) und Validität (Gültigkeit) als wichtigste Gütekriterien
- Objektivität ist die Voraussetzung für Reliabilität und Validität
- Ein Messverfahren ist dann reliabel, wenn der Messfehler-Anteil gering ist (-> kleineres Vertrauensintervall für den wahren Wert)
- Validität als wichtigstes Kriterium: Ein Messverfahren muss das messen, was es messen soll, und nichts anderes!
- Ein Test kann nicht valider sein, als er reliabel ist!