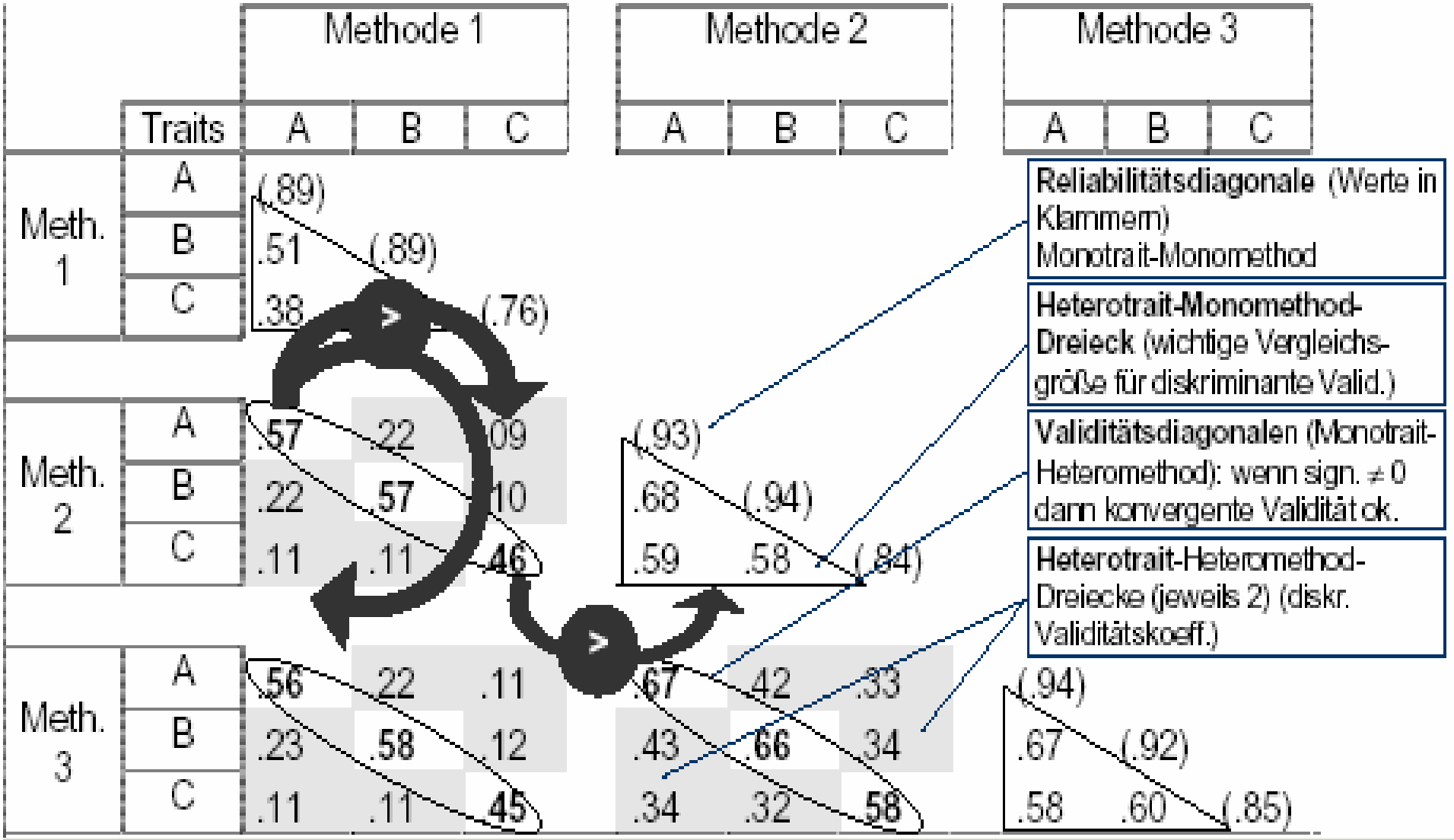
LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

Vorlesung Testtheorien

Dr. Tobias Constantin Haupt, MBA

Sommersemester 2007

Konstruktvalidität: MTMM



Konstruktvalidität: MTMM

MTMM-Matrix

Besteht aus den spaltenweise miteinander korrelierten Meßwerten und läßt sich in vier Teilmatrizen zergliedern:

❶ ***Monotrait-Monomethod-Diagonale (Reliabilitätsdiagonale)***: Jedes Einzelkonstrukt (mit einer Methode gemessen) wird mit sich selbst korreliert (Autokorrelation). Hier stehen aber meist die Reliabilitäten der Messungen (Konvention), möglicherweise um die gemeinsame Methodenvarianz abschätzen zu können, da gleiche Reliabilitäten der Verfahren eigentlich Anwendungsvoraussetzungen für die MTMM sind.

❷ ***Monotrait-Heteromethod-Diagonalen (konvergente Validitätskoeffizienten)***: Ein Konstrukt wird mit mehreren Methoden gemessen. Sign. Korrelationen [die zudem über 0.3 (mittlerer Effekt), bzw. über 0.5 (großer Effekt) liegen sollten] zwischen den Methoden zeigen konvergente Validität an.

❸ ***Heterotrait-Monomethod-Blöcke (bzw. Dreiecke)***: Hier werden die Konstrukte jeweils untereinander korreliert (jedes mit jedem), allerdings immer nur bei gleicher Methode (3 Korrelationen mit jeder der drei Methoden im obigen Beispiel). Die Konstrukte sollten, wenn diskriminante Validität vorliegt, möglichst gering miteinander korrelieren (ansonsten gibt's Redundanzen).

❹ ***Heterotrait-Heteromethod-Blöcke (bzw. Dreiecke)***: Korrelationen zwischen unterschiedlichen Konstrukten, die mit unterschiedlichen Methoden gemessen worden sind. Hier werden die geringsten Korrelationen erwartet, da weder methodische noch inhaltliche Übereinstimmungen vorliegen.

Konstruktvalidität: Probleme der MTMM

- Die Analysekriterien der MTMM erlauben aufgrund ihres informellen Charakters keine objektive, quantitative Bestimmung der konvergenten und diskriminativen Validität.

- Dies führt dazu, daß der Auswerter z.B. selbst entscheiden muß, wie viele Verletzungen eines Kriteriums (die bei zunehmender Anzahl von Traits und Methoden aufgrund von Stichprobenfluktuationen und anderen Artefakten sehr wahrscheinlich werden) er tolerieren soll, ohne damit etwa die diskriminante Validität in Frage zu stellen.

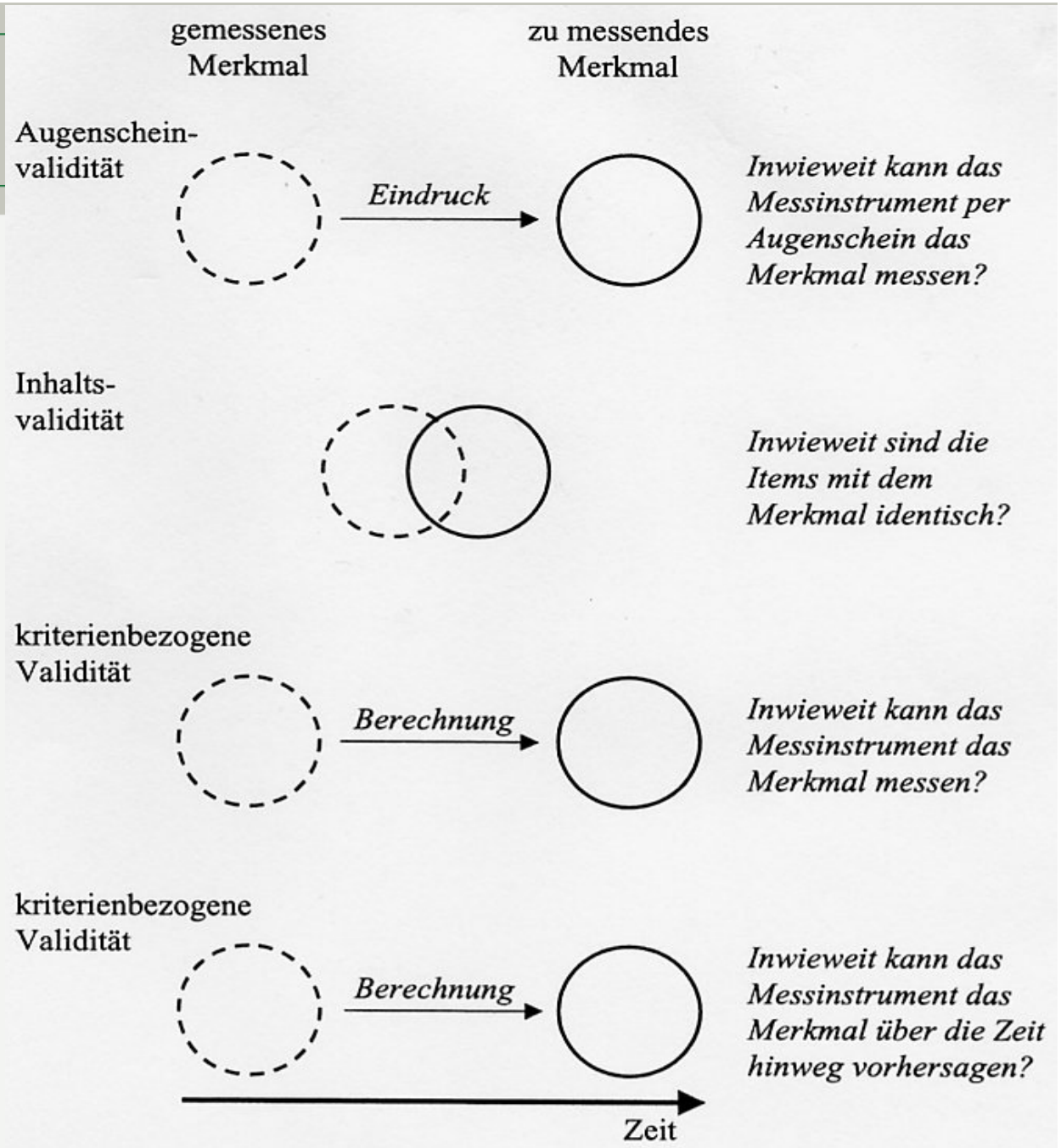
Konstruktvalidität: Probleme der MTMM

Das **wichtigste Problem** stellt die implizite Voraussetzung des Auswertungsverfahrens dar, daß alle Merkmale mit gleicher Zuverlässigkeit gemessen werden müssten, was in der Praxis nicht der Fall ist und so zu Fehleinschätzungen führt. (Ostendorf, Angleitner & Ruch, 1986).

Zusätzliche Literatur zur MTMM:

- Campbell, D.T. & Fiske, D.W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56*, 81-105.

Die **Grundformen der Validität** nochmals kurz und bündig in mit graphischer Veranschaulichung

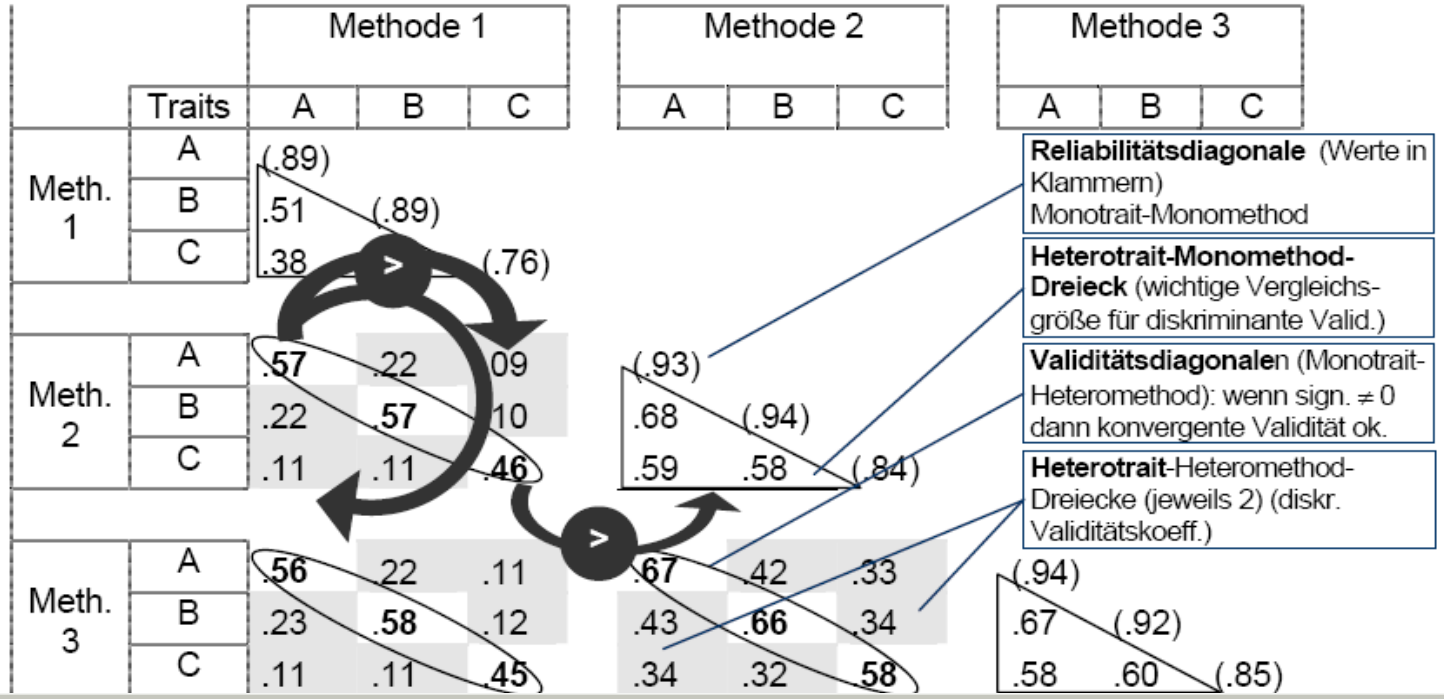


MTMM (Campbell & Fiske, 1959)[1]

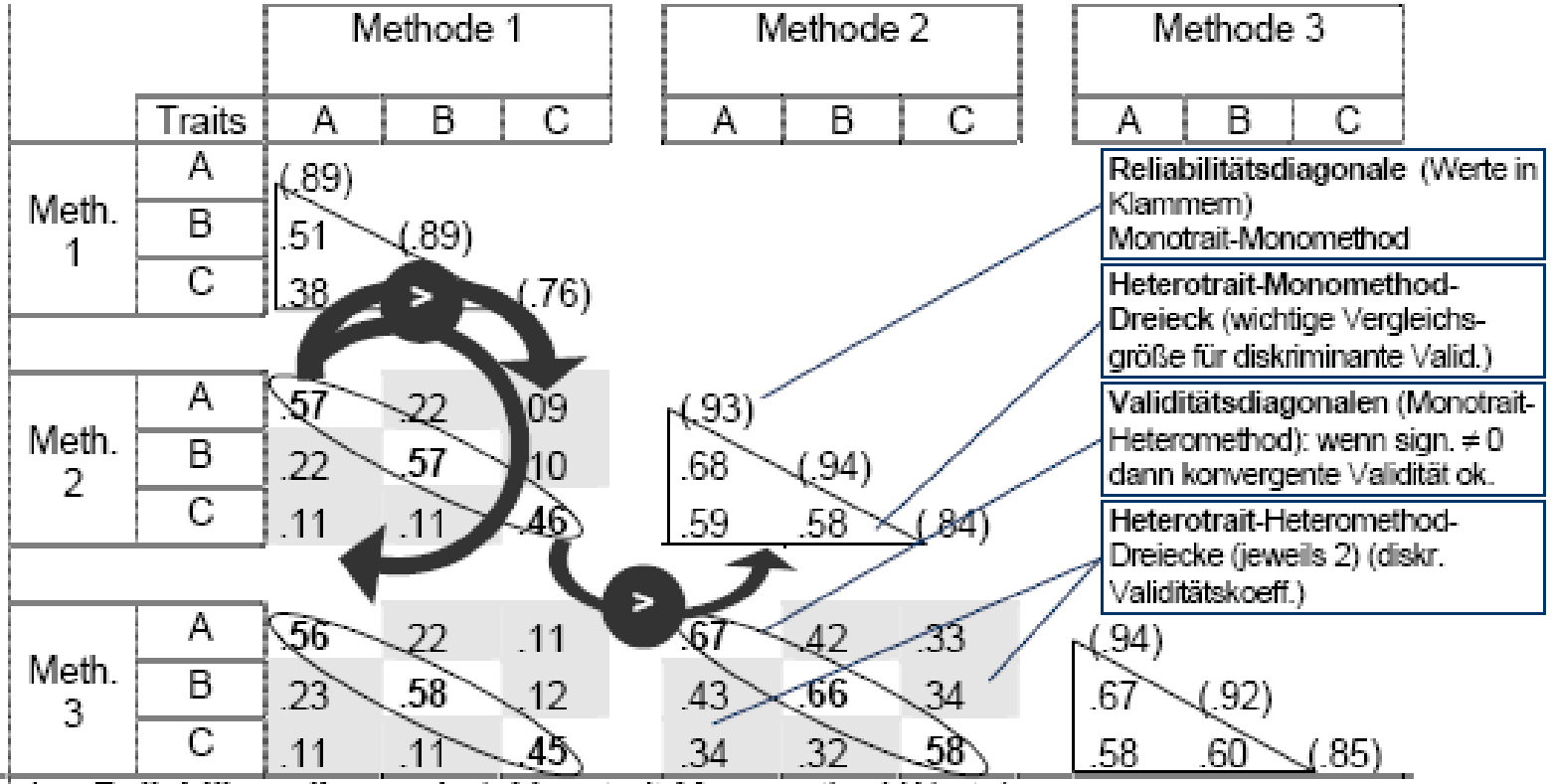
vgl. Ostendorf, Angleitner und Ruch (1986)

Ziel im Kontext Konstruktvalidität	1. wahre Konvergenz und Diskriminanz der Traits beurteilen 2. Aussagen über die proportionalen Anteile der Trait- und Methodenvarianz treffen
Vorgehen	• mehrere Traits (oder auch Konstrukte) mit mehreren (möglichst unterschiedlichen) Methoden erheben • die Ergebnisse interkorrelieren • Höhe und Muster der Korrelationen sind indikativ für konvergente und divergente Validität
Diskutieren Sie die Begriffe konvergente und divergente Val. im Kontext von MTMM!	• konvergent = gleiche oder ähnliche Merkmale kovariieren eng miteinander, unabhängig von der Methode, mit der sie gemessen werden • siehe unten • divergent = wenn verschiedene Konstrukte (mit einer oder mehreren Methoden gemessen) voneinander differenziert werden. Verschiedene Merkmale sollen innerhalb und zwischen Methoden nur niedrig miteinander korrelieren.

MTMM-Matrix beschreiben



MTMM-Matrix beschreiben



4 Teilmatrizen

1. **Reliabilitätsdiagonale** (~Monotrait-Monomethod-Werte)
   - Reliabilitäten werden hier per Konvention eingetragen. z.B. r_{tt}, r_{12}, α
   - muß größer als Monomethod-Korrelationen sein, sonst zuviel Meth.-Varianz
2. **Heterotrait-Monomethod-Dreieck**:
   - dient als Vergleichsgröße für diskriminante Validität

1.+2. sind jeweils ein Monomethod-Block
3.+4. sind jeweils ein Heteromethod-Block

3. **Konvergente-Validitäts-Diagonalen** (~Monotrait-Heteromethod-Diagonalen)
   - Korrelation der gleichen Traits mit verschiedenen Methoden
4. **zwei Heterotrait-Heteromethod-Dreiecke** (jeweils über und unter den konverg.Validitätsdiagonalen)
   - dienen als Vergleichsgröße für diskriminante Validität

Interpretation	konvergente Validität a) wenn konv. Validitätsdiagonale (**Monotrait-Heteromethod**) sign. ≠ 0 ; je höher desto besser; wenn die Werte der Validitätsdiagonalen nicht groß genug sind, kann man sich den Rest quasi schenken. **Gelingt der Nachweis nicht, so messen unterschiedliche Methoden unterschiedliche Konstrukte.** diskriminante Validität b) die **Werte der konvergenten Validitätsdiagonale sollten größer sein als** verschiedene Traits, die mit der gleichen Methode gemessen werden (**Heterotrait-Monomethod**) • verschiedene traits, die mit derselben Methode gemessen wurden, sollten nicht höher kovariieren als Messungen des jeweils gleichen Traits mittels verschiedener Methoden c) die **Werte der Validitätsdiagonale sollten größer sein als die der benachbarten Heterotrait-Heteromethod-Dreiecke** • ist dies nicht der Fall, so diskriminieren die inhaltlich verschiedenenKonstrukte nicht, um die Kovariation ist evt. durch einen generellen Faktor zu erklären, der mehrere Traits der MTMM erfaßt Konstruktvalidität • Konvergente und diskriminante Validität sind Voraussetzungen für eine gute Konstruktvalidität. d) das **gleiche Muster von Trait-Interkorrelationen** soll sich über alle Heterotrait-Diagonalen + wiederholen (nicht unbedingt in den gleichen Korrelationswerten, aber eben in vergleichbaren Mustern der Reihenfolge der Höhe der Korrelationswerte). „Gleich“ wird dabei unterschiedlich interpretiert: Alwin z.B. → gleiche Rangreihe der Koeffizienten

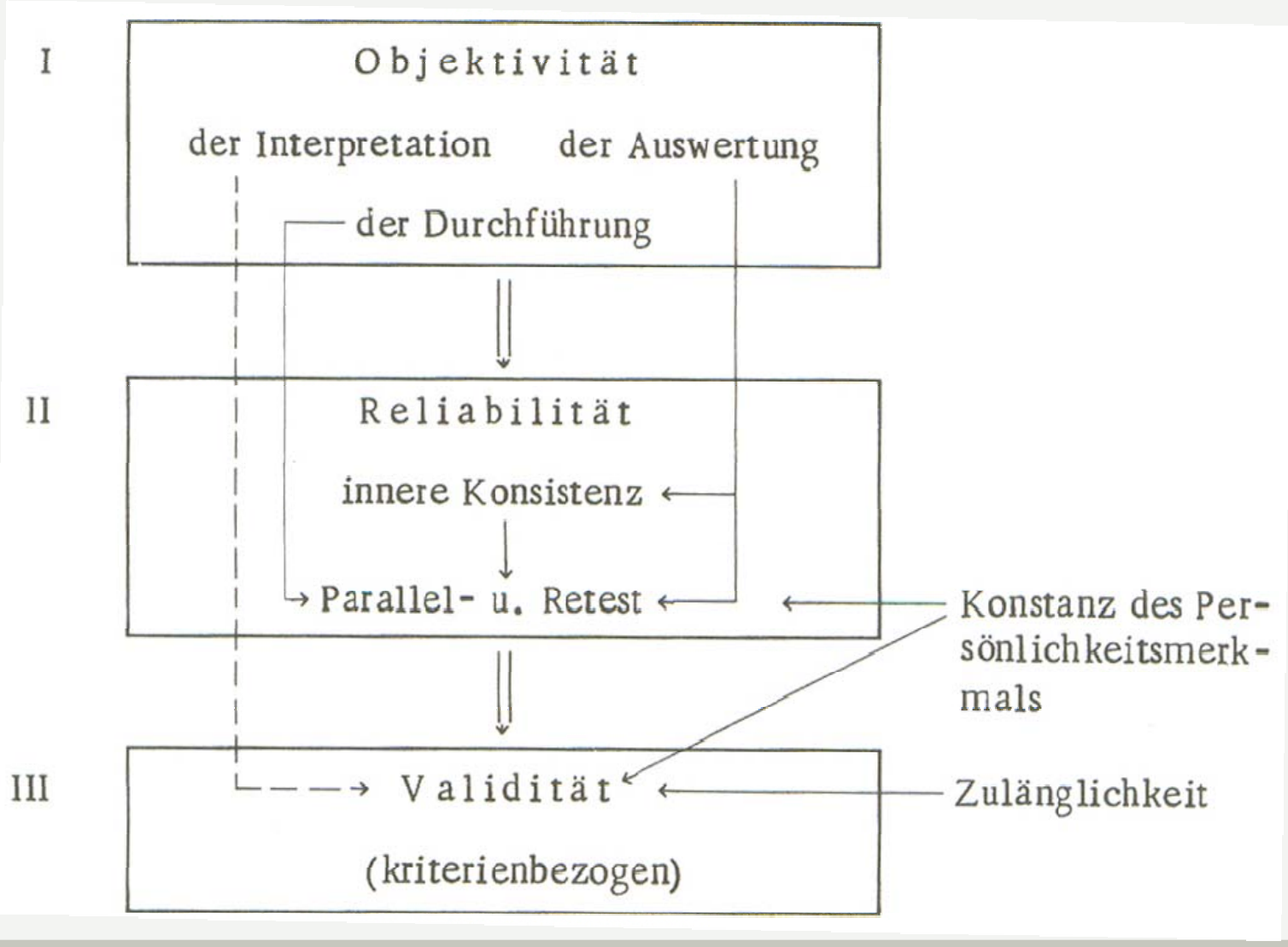
Wechselbeziehungen zwischen Objektivität, Reliabilität und Validität

Objektivität: Bestimmt, wie groß die Reliabilität maximal sein kann.

Reliabilität: Bestimmt wie groß die Validität maximal sein kann (max r_{tc} = Wurzel aus r_{tt}). Durch Erhöhung der Reliabilität kann die Validität erhöht werden (erwarteter Zuwachs erfolgt jedoch asymptotisch, s. Minderungskorrektur). Retest- und Paralleltestreliabilität können zudem nicht größer sein als Objektivität und Konsistenz .

Validität: Ein Test kann jedoch völlig reliabel und objektiv sein und dennoch nicht valide, wenn er nicht das misst, was er messen soll. Ist jedoch die Validität hoch, impliziert dies hohe Objektivität und hohe Reliabilität.

Wechselbeziehungen zwischen Objektivität, Reliabilität und Validität



Wechselbeziehungen

- Die Parallel- oder Retestreliabilität kann nicht höher sein als die Konsistenz oder die Objektivität
- Hohe Validität bedeutet zwangsläufig auch hohe Objektivität, Konsistenz und Reliabilität

Unvereinbarkeiten/Das Reliabilitäts – Validitäts – Dilemma: "Partielle Inkompatibilität der Kardinalkriterien"

Unter diesem Stichwort wird die (teilweise) *Unvereinbarkeit von Reliabilität und Validität* gehandelt:

- Wir wissen ja noch, daß sich bei Erhöhung der Reliabilität (wenngleich um einiges langsamer) auch die Validität erhöht.

Wechselbeziehungen zwischen O., R. und Validität

- Allerdings wird **Reliabilität** vor allem durch **Homogenität** der Items gesichert, **Validität** dagegen durch **heterogene** Items.
- *Unterschiedliche* Aufgabenschwierigkeiten besonders günstig für die *Validität*.
- *Mittlere* Itemschwierigkeit läßt höchste *Reliabilität* erwarten.

Wechselbeziehungen zwischen O., R. und Validität

→ Ausweg aus diesem Dilemma ist die **Kombination homogener Subtests** für ein heterogenes Konstrukt zu einer **Testbatterie**, die insgesamt die Validität sichert, während die Einzelskalen die Reliabilität wahren.

Beispiel: Verwendung von Testbatterien, die homogene Skalen mit guten Reliabilitäten haben und als Ganzes (Testgesamtwert) dann ein komplexeres Konstrukt (z.B. Intelligenz) valide messen.

Normierung (Eichung)

Definition:

Die Normierung eines Tests liefert das Bezugssystem, um die individuellen Testwerte (Rohwerte, die für sich noch keine Aussagekraft haben) im Vergleich zu denen einer größeren (meist repräsentativen) SP von Testteilnehmern einordnen zu können.

Zu diesem Zweck werden die Rohwerte in standardisierte Werte (Prozeß der Standardisierung), für die es wiederum verschiedene Skalen gibt (z-Werte, T-Werte, etc.), transformiert.

Normierung (Eichung)

Standardisierung:

Bezeichnet die Transformation von Rohwerten in Standardwerte, die einen festgelegten Mittelwert (z.B. M=100) und eine festgelegte Streuung (z.B. s=15) haben.

Eine Standardisierung wird mit dem Ziel vorgenommen, Messungen, die auf verschiedenen Maßstäben vorgenommen wurden, vergleichbar zu machen.

Ziele

- Testwerte verschiedener Probanden im gleichen Test oder von denselben Probanden in verschiedenen Tests sollen vergleichbar gemacht werden.
- Darstellen auf einer einheitlichen Skala.
- Transformation der Rohwerte (ohne Änderung des Skalenniveaus).

Arten von Normierungen

Man unterscheidet drei unterschiedliche Arten, um Rohwerte in Standardwerte zu transformieren:

- Äquivalentnormen
- Variabiliäts- und Abweichungsnormen
- Prozentränge

Variabilitäts- oder Abweichungsnormen

Liegen vor, wenn die individuelle Leistung mit der Werteverteilung (meist Normalverteilung, durch Mittelwert und Streuung charakterisiert) einer (repräsentativen) Vergleichspopulation (z.B. gleiche Altersgruppe) in Beziehung gesetzt wird.

Testwerte werden dabei in Werte einer Standardskala transformiert, z.B. in

- z-Werte (M=0, S=1),
- Z-Werte (M=100, S=10) oder
- T-Werte (M=50, S=10).

Beispiel: Wechsler-Intelligenztest, bei welchem die Testergebnisse für jede Altersstufe separat normiert worden sind (M=100, S=15).

Variabilitäts- oder Abweichungsnormen

Vorteile:

- Die Ergebnisse verschiedener Tests können auf derselben Normskala (evtl. nach Transformation) gemessen und verglichen werden.
- Kommt mit den Problemen der Äquivalentnormen zurecht.

Probleme:

- Setzt voraus, daß die Verteilung in der Population über die Zeit hinweg stabil bleibt (was bei Intelligenz empirisch z.B. nicht der Fall ist). Dieses Problem ließe sich durch regelmäßige Nachnormierungen lösen.
- Setzt in der Regel eine Normalverteilung der Populationswerte voraus, da sonst nicht in Standardwerte transformiert werden darf.

Prozentränge

Hier werden die Rohwerte in Prozentwerte transformiert, die darüber informieren, wieviel Prozent aller Mitglieder einer Bezugspopulation (z.B. gleiche Altersstufe) einen höheren oder geringeren Wert erzielt haben.

Beispiel: 20 Punkte in einem IQ-Test könnten z.B. einem Prozentrang von 60 entsprechen, d.h., daß 60% der Vergleichspopulation schlechter und 40% besser abgeschnitten haben.

Vorteile:

- Setzt keine best. Verteilungsform voraus (Bezugswerte werden nur gerangreiht).
- Leichte Verständlichkeit.

Problem:

- Gleiche Prozentrangdifferenzen müssen aufgrund des ordinalen Skalenniveaus nicht gleiche Rohwertdifferenzen bedeuten, oder gleichen Differenzen in Standardnormwerten entsprechen.

Eichstichprobe:

Der Umfang der Eichstichprobe ist *abhängig vom*

- **Allgemeinheit**sgrad des untersuchten Merkmals

  und

- vom **Heterogenität**sgrad der Zielpopulation.

Stichproben

Je **allgemeiner** das Merkmal und je **heterogener** die Population, **desto größer** sollte die Eichstichprobe sein.

Bei **Eignungstests** reichen **einige 100** Probanden, bei allgemeinen **Persönlichkeits- und Leistungstests** können es **einige 1000** sein.

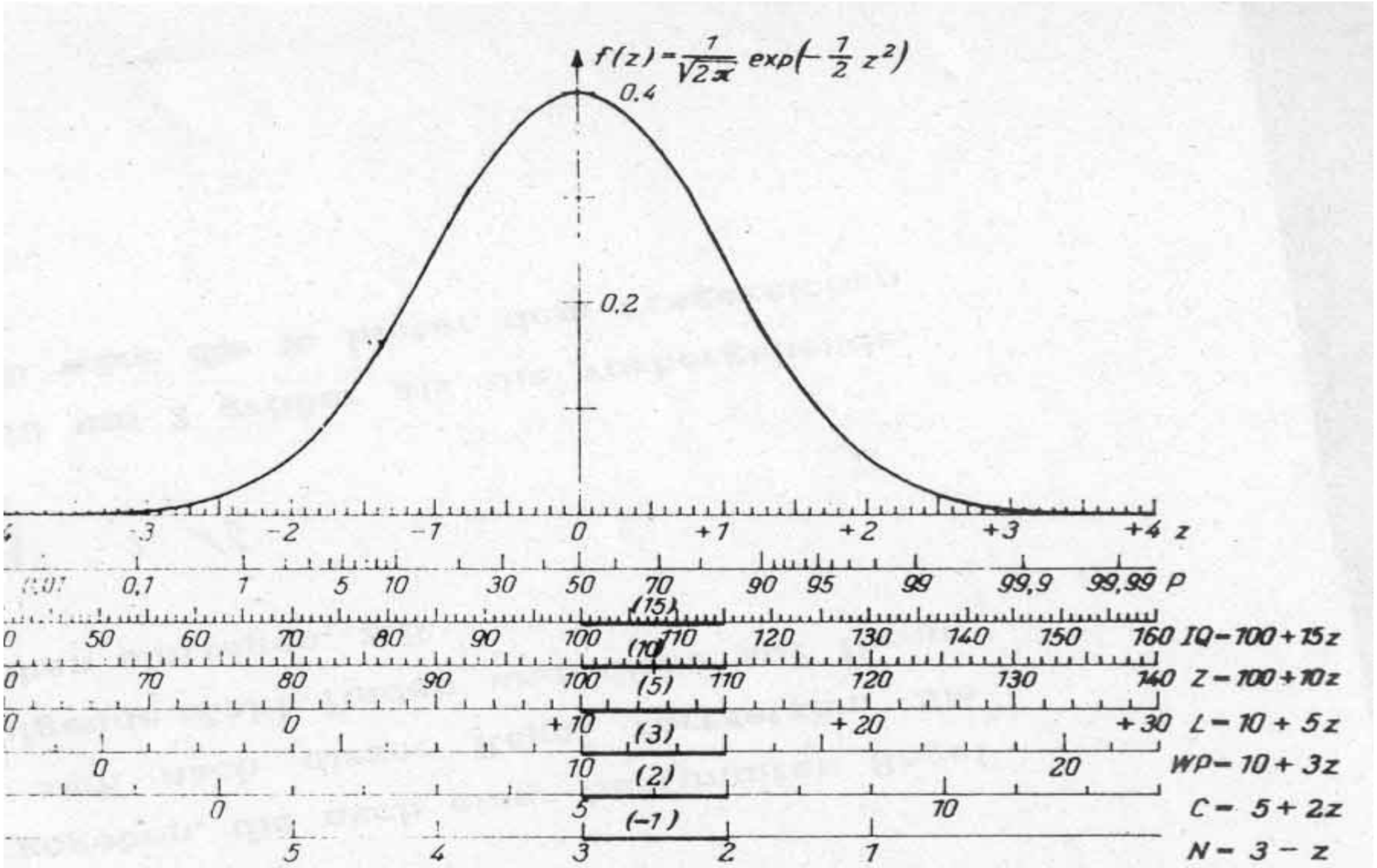


Abb. 15: Normierte GAUSS-Verteilung [$f(z)$], Prozentrangwerte (P) und einige abgeleitete Standardskalen.

Beziehungen verschiedener Transformationen bzw. Skalen bei Vorliegen einer normalen Ausgangsverteilung von Meßwerten

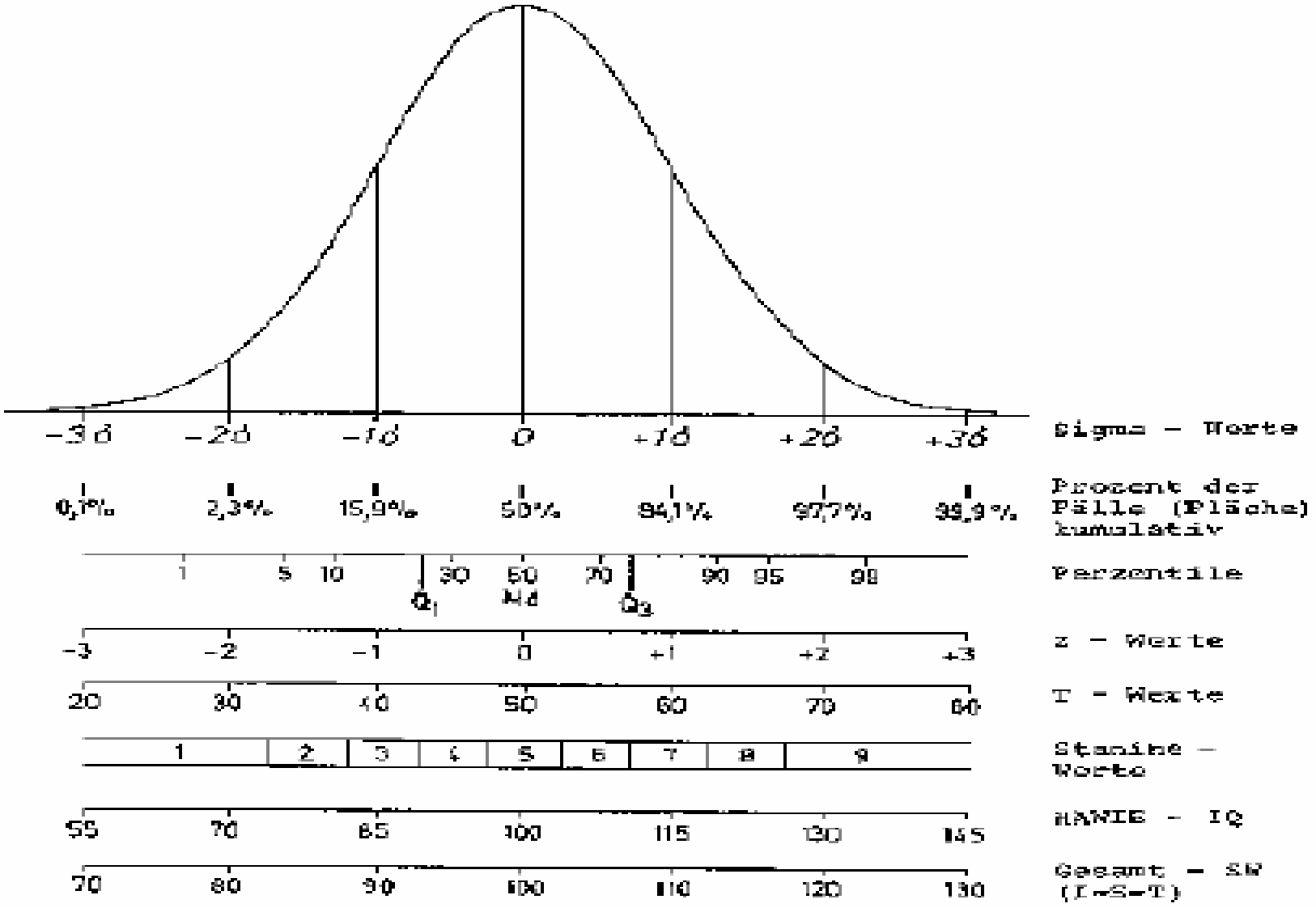


Tabelle A: *Standardwert-Skalen (1)*

z	P	Z	T	L	C	N	IQ
−4,00	0,0032	60	10	−10	−3		40
−3,95	0,0039						
−3,90	0,0048	61	11				
−3,85	0,0059						
−3,80	0,0072	62	12	− 9			43
−3,75	0,0088						
−3,70	0,0108	63	13				
−3,65	0,0131						
−3,60	0,0159	64	14	− 8			46
−3,55	0,0193						
−3,50	0,0233	65	15		−2		
−3,45	0,0280						
−3,40	0,0337	66	16	− 7			49
−3,35	0,0404						
−3,30	0,0483	67	17				
−3,25	0,0577						
−3,20	0,0687	68	18	− 6			52
−3,15	0,0816						
−3,10	0,0968	69	19				
−3,05	0,1144						
−3,00	0,1350	70	20	− 5	−1		55
−2,95	0,159						
−2,90	0,187	71	21				
−2,85	0,219						
−2,80	0,256	72	22	− 4			58
−2,75	0,298						
−2,70	0,347	73	23				
−2,65	0,403						
−2,60	0,466	74	24	− 3			61
−2,55	0,539						
−2,50	0,621	75	25		0		
−2,45	0,714						
−2,40	0,820	76	26	− 2			64
−2,35	0,939						
−2,30	1,072	77	27				
−2,25	1.222						
−2,20	1.390	78	28	− 1			67
−2,15	1,578						
−2,10	1,786	79	29				
−2,05	2,018						
−2,00	2,275	80	30	0	1	5,0	70
−1,95	2,56						
−1,90	2,87	81	31			4,9	
−1,85	3,22						

z	P	Z	T	L	C	N	IQ
+0,00	50,00	100	50	10	5	3,0	100
+0,05	51,99						
+0,10	53,98	101	51			2,9	
+0,15	55,96						
+0,20	57,93	102	52	11		2,8	103
+0,25	59,87						
+0,30	61,79	103	53			2,7	
+0,35	63,68						
+0,40	65,54	104	54	12		2,6	106
+0,45	67,36						
+0,50	69,15	105	55		6	2,5	
+0,55	70,88						
+0,60	72,57	106	56	13		2,4	109
+0,65	74,22						
+0,70	75,80	107	57			2,3	
+0,75	77,34						
+0,80	78,81	108	58	14		2,2	112
+0,85	80,23						
+0,90	81,59	109	59			2,1	
+0,95	82,89						
+1,00	84,13	110	60	15	7	2,0	115
+1,05	85,31						
+1,10	86,43	111	61			1,9	
+1,15	87,49						
+1,20	88,49	112	62	16		1,8	118
+1,25	89,44						
+1,30	90,32	113	63			1,7	
+1,35	91,15						
+1,40	91,92	114	64	17		1,6	121
+1,45	92,65						
+1,50	93,32	115	65		8	1,5	
+1,55	93,94						
+1,60	94,52	116	66	18		1,4	124
+1,65	95,05						
+1,70	95,54	117	67			1,3	
+1,75	95,99						
+1,80	96,41	118	68	19		1,2	127
+1,85	96,78						
+1,90	97,13	119	69			1,1	
+1,95	97,44						
+2,00	97,725	120	70	20	9	1,0	130
+2,05	97,982						
+2,10	98,214	121	71				
+2,15	98,422						

…bedrohen die Validität und sollten deshalb minimiert werden!

Welche **Verfälschungsarten** gibt es?

Absichtliche *Verfälschungen oder Verstellungen (Faking)*: Wenn Versuchspersonen z.B. versuchen

- möglichst hohe Werte zu erzielen (Simulation),
- möglichst niedrige Werte zu erzielen (Dissimulation, „Dummstellen“) oder
- entsprechend der von ihnen wahrgenommenen sozialen Erwünschtheit reagieren.

Welche Verfälschungsarten gibt es?

Unbemerkte/unkontrollierte Verzerrungen:

Wenn Versuchspersonen

- unbewußt bestimmte Antworttendenzen (s. u.) bevorzugen oder sich
- kognitive Effekte aus den Bereichen Gedächtnis (es werden nur ausgewählte, nicht repräsentative Verhaltensaspekte erinnert), Konzentration (nicht-überdachtes Antworten), Informationsverarbeitung, Selbstbeobachtung oder Selbstdarstellung auf die Testergebnisse auswirken.

Ursachen für (absichtliche) Testverfälschungen:

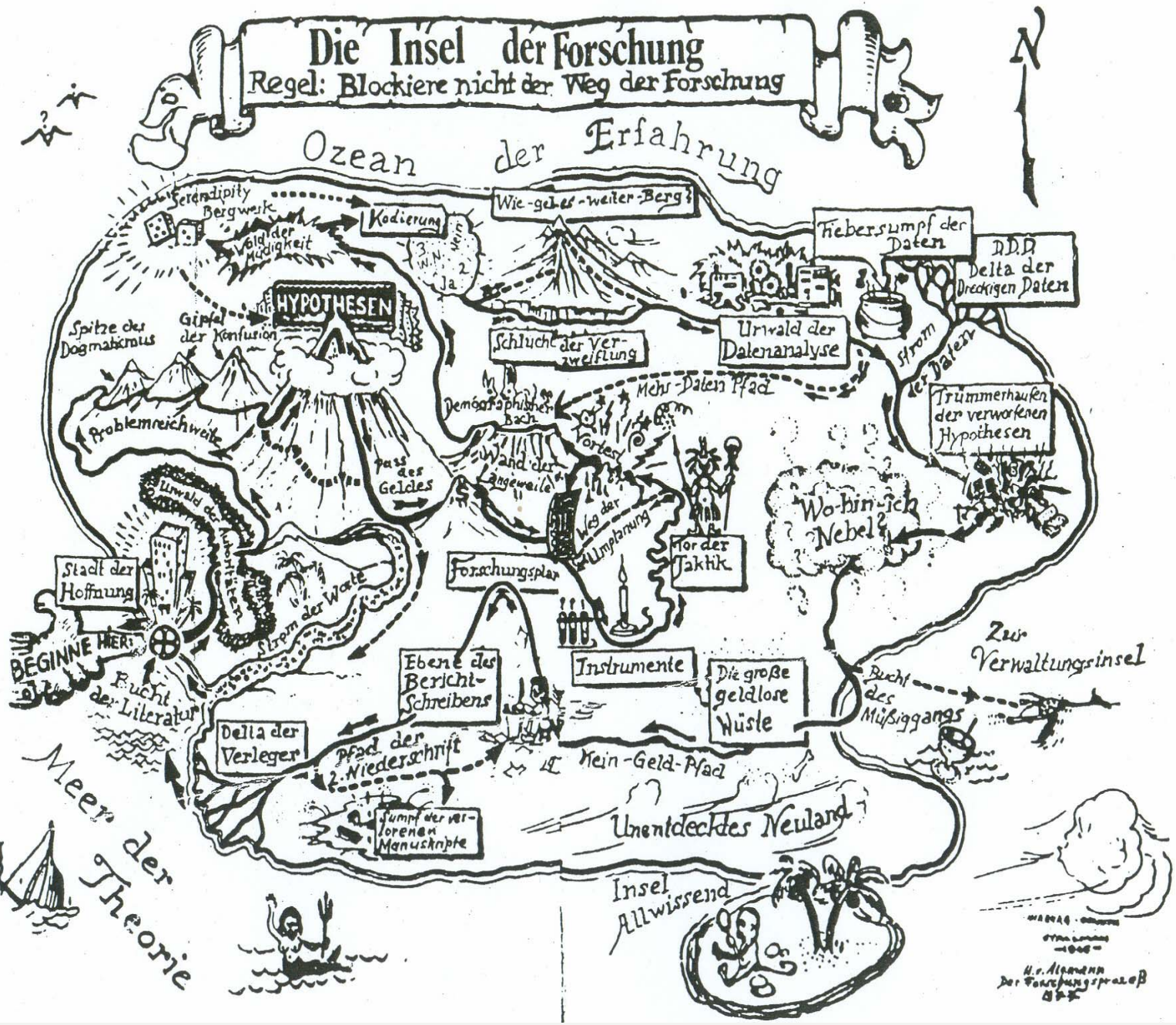
- Da Testergebnisse für die Versuchspersonen oft persönlich sehr wichtig sind (z.B. Einstellungs- oder Selektionstests), haben sie ein Interesse daran, die Ergebnisse so zu „korrigieren", daß das von ihnen angestrebte Ziel besser erreicht werden kann.

- Versuchspersonen können aber auch einfach (z.B. mangels Vertrauen zum Versuchsleiter) nicht geneigt sein, best. Aspekte ihres Verhaltens und Denkens anderen Personen mitzuteilen.

Testverfälschungen

Welche wichtigen (voneinander wohl nicht unabhängigen) **Verfälschungsarten** gibt es?

- Selbstdarstellung
- Soziale Erwünschtheit
- Antworttendenzen
- Urteilsfehler beim Einsatz von Ratingskalen


Die Insel der Forschung
Regel: Blockiere nicht der Weg der Forschung
Ozean der Erfahrung
N
Serendipity Bergwerk
Kodierung
Wie-geht's-weiter-Berg
Ja
Nein
Hebersumpf der Daten
D.D.D. Delta der Dreckigen Daten
Strom der Daten
HYPOTHESEN
Spitze des Dogmatismus
Gipfel der Konfusion
Schlucht der Verzweiflung
Urwald der Datenanalyse
Demographischer Bach
Mehr-Daten Pfad
Trümmerhaufen der verworfenen Hypothesen
Problemreichweite
Pass des Geldes
Wand der Langeweile
Fort
Weg der Umkehrung
Tor der Taktik
Wo-bin-ich Nebel?
Stadt der Hoffnung
Urwald der Theorien
Strom der Worte
Forschungsplan
BEGINNE HIER
Bucht der Literatur
Ebene des Berichtschreibens
Instrumente
Die große geldlose Wüste
Zur Verwaltungsinsel
Bucht des Müßiggangs
Delta der Verleger
Pfad der Niederschrift
Kein-Geld-Pfad
Sumpf der verlorenen Manuskripte
Unentdecktes Neuland
Meer der Theorie
Insel Allwissend