


6. Testtheorie: Klassische Testtheorie KTT

auf der KTT beruhen...	... die meisten Tests
KTT und PTT	<ul style="list-style-type: none"> • PTT ist eigentlich „besser“ - aber deshalb auch mathematisch wesentlich komplexer • nach der KTT entwickelte Tests haben sich empirisch bewährt
Variationen von Messergebnissen	Testergebnisse einer Person fallen beim gleichen Test zu unterschiedlichen Messzeitpunkten unterschiedlich aus
→ unsystematische Fehler	<ul style="list-style-type: none"> • Übungs- und Transfereffekte • äußere und innere Einflüsse (<i>Tagesform, Situation...</i>)
 KTT als Theorie der (Freiheit von) Messfehler(n) → Reliabilität	aus welchen Komponenten bestehen Messwerte ? → woher kommen die inter- u. intraindividuellen Variationen in den Messwerten ? → inwieweit weisen diese Variationen auf Messfehler hin ?
Annahme der KTT	Normalverteilung des Testergebnisses (durch Zufallsziehung und Zufallsfehler)
Existenzaxiom: „Es existiert ein wahrer Wert T“	→ Mittelwert dieser Normalverteilung = wahrer Wert T
„wahrer Wert“	= Verhaltenstendenz einer Person in einer konkreten Situation
Fehler	unkontrollierte und unsystematische Störeinflüsse <ul style="list-style-type: none"> • schwanken auch (<i>wie der wahre Wert</i>) um einen Mittelwert

7. Testtheorie: Klassische Testtheorie KTT

<p>Grundannahme der KTT</p>	$X_i = T_i + E_i$	<p>Messwert X = wahrer Wert T + Fehler E</p>						
<p>=></p>	$E_i = X_i - T_i$	<p>Fehler E = Messwert X - wahrer Wert T</p>						
<p>Folgerungen / Ableitungen</p>								
<p>1.</p>	$M(E)_i = 0$ $M(E)_p = 0$	<ul style="list-style-type: none"> • der Mittelwert M des Fehlers E ist 0 → die Wahrscheinlichkeit, den wahren Wert zu überschätzen, ist gleich der Wahrscheinlichkeit, den wahren Wert zu unterschätzen • bei unendlich vielen Messungen pendelt sich der Fehler bei 0 ein 						
<p>2.</p>	$r(E, T) = 0$ $r(E_A, E_B) = 0$ <p>wenn...</p> $r(E_A, T_B) = 0$	<ul style="list-style-type: none"> • wahrer Wert T und Fehler E sind unkorreliert → sie variieren unabhängig voneinander • Fehler E verschiedener Tests A u. B sind unkorreliert → Fehler eines Tests A variieren unabhängig von Fehlern eines Tests B beide Tests experimentell unabhängig sind → der Punktwert eines Testes B darf nicht vom Punktwert eines Testes A abhängig sein (auch wenn die getesteten Merkmale empirisch zusammenhängen, z.B. „räumliches Denken“ und „Rechenkompetenz“) • Fehler E und wahrer Wert T verschiedener Tests A u. B sind kreuzweise unkorreliert <div style="display: flex; align-items: center; margin-top: 10px;"> <table border="1" style="border-collapse: collapse; text-align: center; margin-right: 10px;"> <tr> <td style="padding: 5px;">E_A</td> <td style="width: 20px;"></td> <td style="padding: 5px;">E_B</td> </tr> <tr> <td style="padding: 5px;">T_A</td> <td style="width: 20px;"></td> <td style="padding: 5px;">T_B</td> </tr> </table> <div style="margin-right: 10px;"> </div> <div style="margin-left: 10px;"> <p>→ Fehler eines Tests A variieren unabhängig von den wahren Werten eines Tests B</p> </div> </div>	E_A		E_B	T_A		T_B
E_A		E_B						
T_A		T_B						

8. Testtheorie: Klassische Testtheorie KTT

Intervallskalenniveau	als Voraussetzung für die Berechnung von Differenzen → in den meisten Studien eher „fraglich“	
„Per fiat“ -Messung	Vertrauen darauf, dass sich auch nicht intervallskalierte Variablen durch Aufsummierung von Antworten einem Intervallskalenniveau annähern	
Reliabilität als...	Anteil wahrer Varianz S^2_T (Varianz des wahren Wertes) an der gesamten Testvarianz: S^2_X (Varianz der beobachteten Werte) = $S^2_T + S^2_E$ (Varianz der Messfehler)	
Schätzung der Testreliabilität	$r_{tt} = \frac{S_T^2}{S_T^2 + S_E^2} = \frac{S_T^2}{S_X^2}$	Verhältnis „systematische Ergebnisse“ und „zufällige Ergebnisse“ → davon abhängig: Reproduzierbarkeit + Vergleichbarkeit
Schätzverfahren	Problem: oft nur eine Messung oder wiederholte Messungen unter nicht identischen Bedingungen → unterschiedliche Schätzverfahren mit spezifischen Vor- und Nachteilen	
Einführung in die Test- und Fragebogenkonstruktion, Pearson Education 2003, Markus Bühner	<ul style="list-style-type: none"> • Zusammenhang zwischen Items, die das gleiche Konstrukt messen sollen, wird ermittelt. • Zwei Tests mit unterschiedlichen Items, die das gleiche Konstrukt messen werden in Zusammenhang zueinander gesetzt. • Ein einzelner Test kann in zwei Hälften aufgeteilt und der Zusammenhang zwischen beiden Hälften bestimmt werden. • Ein Test wird mit einem gewissen Abstand zweimal vorgegeben und der Zusammenhang zwischen erster und zweiter Messung ermittelt. 	
Standardschätzfehler Standardmessfehler	als Ableitungen aus der Reliabilität → Wie verlässlich ist das Testergebnis einer einzelnen Person ?	
Reliabilität ≠ Validität	im „wahren Wert“ T ist alles erfasst, was der Test misst: Zielkonstrukt C_i + systematische Fehler B_i z.B. im Intelligenztest: Intelligenz und Testangst Reliabilität fragt also im Gegensatz zu Validität nicht danach, wie weit der Test (genau das) misst, was er messen soll	
Erweiterte Grundgleichung der KTT	→ $X_i = C_i + B_i + E_i$ $\text{mit } T_i = C_i + B_i$	
Validität	$r_{tt} = \frac{S_C^2}{S_C^2 + S_B^2 + S_E^2}$	$= \frac{\text{Zielkonstruktvarianz}}{\text{Gesamtvarianz}}$ Anteil der Zielkonstruktvarianz an der Gesamtvarianz → wie weit misst der Test genau das, was er messen soll ?

9. Testtheorie: Klassische Testtheorie KTT – Kritische Anmerkungen

<p>Axiome der KTT</p> <ul style="list-style-type: none">• Messwert $X = \text{wahrer Wert } T + \text{Fehler } E$• der Mittelwert M des Fehlers E ist 0• bei unendlich vielen Messungen pendelt sich der Fehler bei 0 ein• wahrer Wert T und Fehler E sind unkorreliert• Fehler E verschiedener Tests A u. B sind unkorreliert• Fehler E und wahrer Wert T verschiedener Tests A u. B sind kreuzweise unkorreliert	<p>mathematisch logisch, aber empirisch nicht überprüfbar</p> <p>mangels „funktioneller Annahmen“ →</p> <p>„Wie kommen die Reaktionen / Antworten / Leistungen überhaupt zustande?“</p> <p>→ könnte also gut sein, dass wahrer Wert und Messfehler doch an irgendeiner Stelle korrelieren</p> <p>→ und dass diese Zusammenhänge die Reliabilität und Validität des Testes (positiv oder negativ) beeinflussen</p>
<p>Messung in Exrembereichen</p>	<p>→ ungenau</p> <p>Problem:</p> <ol style="list-style-type: none">1. Extremfragen dienen jedoch der Erfassung unterschiedlicher Leistungsniveaus.2. „Extrempersonen“ sind eher selten zu Studienzwecken vorhanden → wenig Möglichkeit zur Validierung der Ergebnisse in diesem Bereich
<p>Eindimensionalität</p>	<p>wird ohne Prüfung vorausgesetzt</p> <p>sondern folgt aus der ebenfalls nicht überprüften Annahme:</p> <p>„Keine Korrelation zwischen verschiedenen Fehlerwerten“</p>
<p>stichprobenabhängige Testwerte</p>	<p>Ergebnisse eines Testteilnehmers sind abhängig von der Referenzgruppe</p> <p>→ „im Vergleich zu ... bin ich ... „</p> <p>„Unter den Blinden ist der Einäugige ein König“ 😊</p> <p>(vgl. PTT: stichprobenunabhängige Kennwerte)</p>
<p>trotzdem...</p>	<p>hat sich die KTT in der Praxis bewährt</p> <p>... vielleicht aufgrund der inhaltlich zu begründenden Konstruktion der Item-Skalen ?</p>