

Zuverlässigkeit von Bestehens-/ Nichtbestehensentscheidungen bei Multiple Choice Prüfungen

Lahner, F-M*, Lörwald, A.C*, Guttormsen, S.** Fischer, M***, Huwendiek, S*.

*Abteilung für Assessment und Evaluation (AAE), Institut für medizinische Lehre (IML), ** Institut für medizinische Lehre (IML), ***LMU München, Institut für Didaktik und Ausbildungsforschung in der Medizin.

1. Einleitung

- Zuverlässige Bestehens/ Nichtbestehensentscheidung sind wichtig bei summativen Prüfungen
- Das häufig verwendete Cronbach alpha gibt einen globalen Wert über die Messgenauigkeit der Prüfung, nicht aber über die Messgenauigkeit der Bestehensgrenze¹
- Mit konditionalen Reliabilitäten kann die Messgenauigkeit für jeder erreichte Punktzahl und damit auch an der Bestehensgrenze berechnet werden².
- Konditionale Reliabilitäten können in der Klassischen Test Theorie (CTT) und in der Item Response Theorie (IRT) berechnet werden².

2. Forschungsfrage

Wie verhalten sich konditionale Reliabilitäten in der CTT und der IRT?

Wie ist die Konditionale Reliabilität an der Bestehensgrenze in CTT und IRT?

3. Methoden

Stichprobe

33 Prüfungen (3 schweizer Fakultäten + Eidgenössische Prüfung Humanmedizin); Durchschnittlich 109 Fragen per Prüfung (SD=42) (Kprim und Typ A)

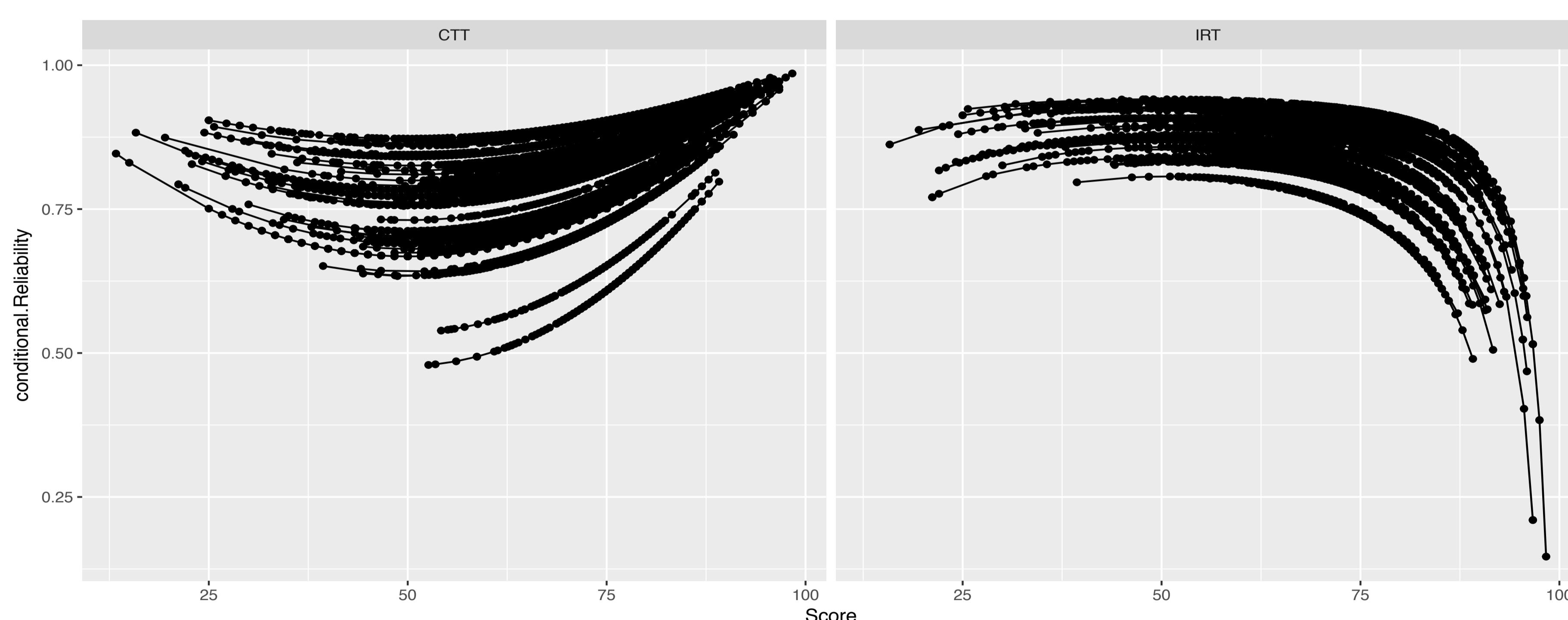
Statistik

Berechnung der konditionalen Reliabilität in der IRT und CTT analog zu Raju et al.² (für IRT Analyse wurde ein 1-PL- IRT Modell verwendet).

ANOVA zum Vergleich der konditionalen Reliabilität an der Bestehensgrenze.

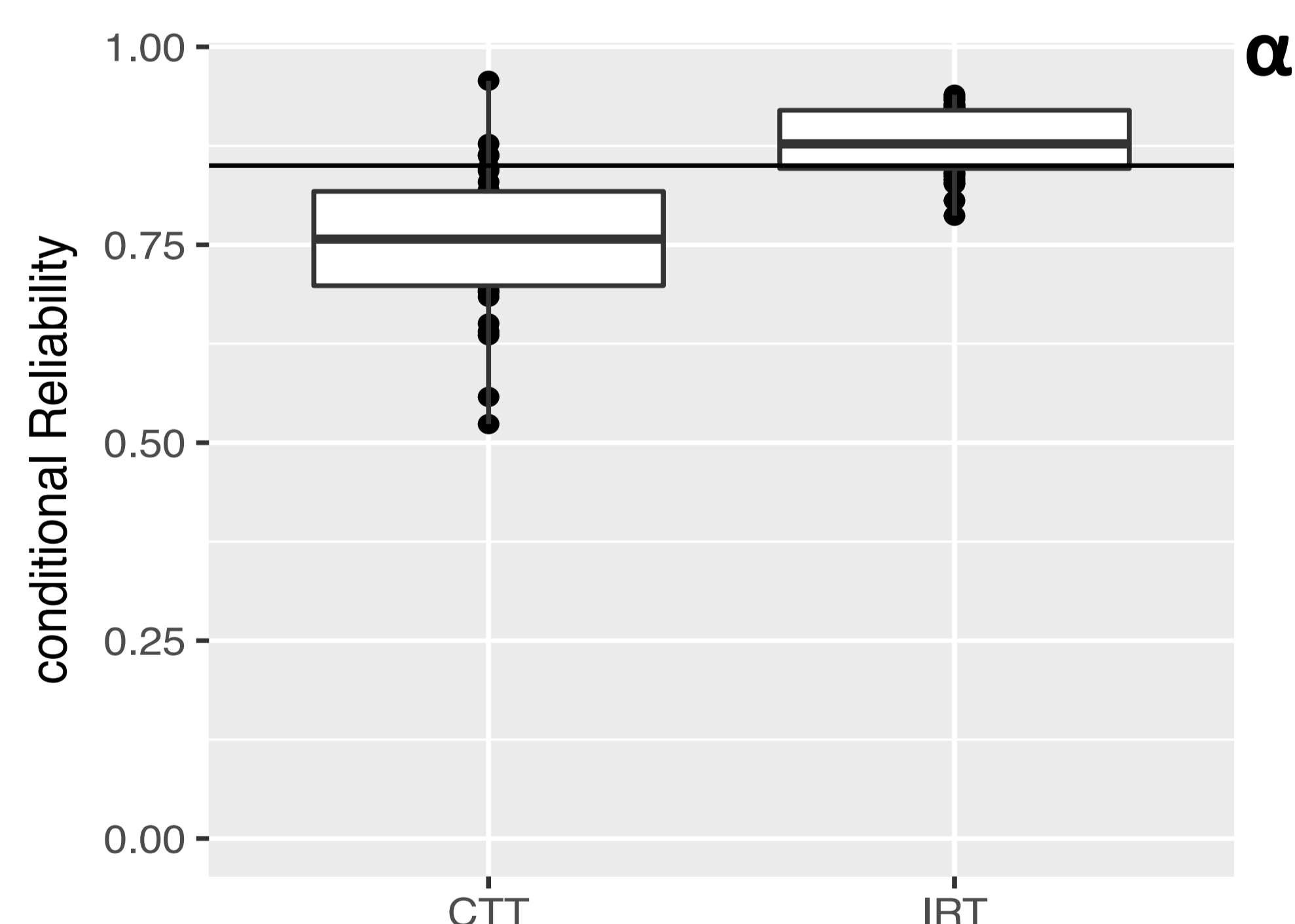
4. Ergebnisse

Verteilung der konditionalen Reliabilität in IRT und CTT



- Messzuverlässigkeit ist nicht linear, bestimmte Bereiche werden genauer gemessen als andere.
- Konditionale Reliabilität in CTT und IRT verlaufen gegenläufig
- Konditionale Reliabilität an der Bestehensgrenze ist in der IRT signifikant höher als in der CTT

Konditionale Reliabilität Bestehensgrenze



5. Schlussfolgerung

- Aufgrund unterschiedlicher Prämissen in den beiden Testtheorien ergeben sich unterschiedliche Verläufe der konditionalen Reliabilität².
- In Abhängigkeit des Zwecks der Prüfung der Prüfung sollte die konditionale Reliabilität in CTT oder IRT berechnet werden.
- Für inhaltsbasierte Bestehensgrenzen empfiehlt sich die Berechnung in der IRT, für normbasierte Bestehensgrenzen in der CTT.

6. Literatur

- AERA, APA, & NCME. (2014). *Standards for educational and psychological testing*: American Educational Research Association.
- Raju, N. S., Price, L. R., Oshima, T., & Nering, M. L. (2007). Standardized conditional SEM: A case for conditional reliability. *Applied Psychological Measurement*, 31(3), 169-180