Wie man unterschiedliche Prüfungen gleich schwer macht: Test-Equating des studentischen kompetenzbasierten Progresstests

A. Möltner, S. Wagener, J. Jünger

GMA-Tagung Leipzig 2015

Studentischer Progresstest

- Fragen von Studierenden konzipiert.
- Zweidimensionaler Blueprint:
  - 8 Fächergruppen
  - 5 Kompetenzbereiche

2013: 140 Items
2014: 114 Items (15 wiederholt)

	Studienjahr						
	1	2	3	4	5	6	Σ
2013:	45	41	74	125	166	18	469
2014:	134	50	55	48	82	12	381

Problem: äquivalente Schwierigkeit

Für eine longitudinale Erfassung der Wissens-entwicklung im Laufe des Studiums durch Progresstest ist eine

äquivalente Schwierigkeit der Tests

in den unterschiedlichen Versionen erforderlich.

Tests lassen sich nicht von vornherein als gleich schwer konzipieren

⇨ Zuordnung von Rohpunktwerten der Tests zu äquivalenten Scores

Lösungsmöglichkeit: IRT

Item-Response-Theorie (logistische Modelle):

- Generierung äquivalenter Scores auf Basis von geschätzten Modellparametern der **Fragen**.
- Attraktive Modelleigenschaften (insbesondere beim Rasch-Modell)
- Verfügbarkeit von Software

Aber:

- In der Praxis sind Modellvoraussetzungen einfacher logistischer Modelle oft nicht erfüllt.
- Komplexe IRT-Modelle erfordern unrealistisch hohe Fallzahlen zur Schätzung.

Lösungsmöglichkeit: OSE

Observed Score Equating

- Generierung äquivalenter Scores auf Basis der empirischen **Summenwerte**
- Bestimmung der Äquivalenzscores durch Verwendung von gemeinsamen Fragen (Ankertest) und Drittvariablen (wie bei IRT)
- Nichtparametrische Schätzung möglich

Ergebnisse IRT

Item-Response-Theorie

Erhebliche Abweichungen für das Rasch-Modell
z. B. für PT 2014: Ausschluss von 96 der 114 Items mit signifikanten Modellabweichungen bei sequentiellem Ausschluss (Itemfit, R-Paket eRm)

Deutliche Abweichung für 3-parametriges logistisches Modell:
z. B. für PT 2014: Ausschluss von 37 der 114 Items bei sequentiellem Ausschluss (item.fit, R-Paket ltm)

Methoden OSE

Observed Score Equating

Design:
Nonequivalent Groups with Anchor Test (NEAT)

Studierendengruppen sind nicht äquivalent (unterschiedliche Fakultäten, verschiedene Zusammensetzung nach Studienjahren)

Verwendung der 15 wiederholten Items als Ankertest und Studienjahr als Kovariate

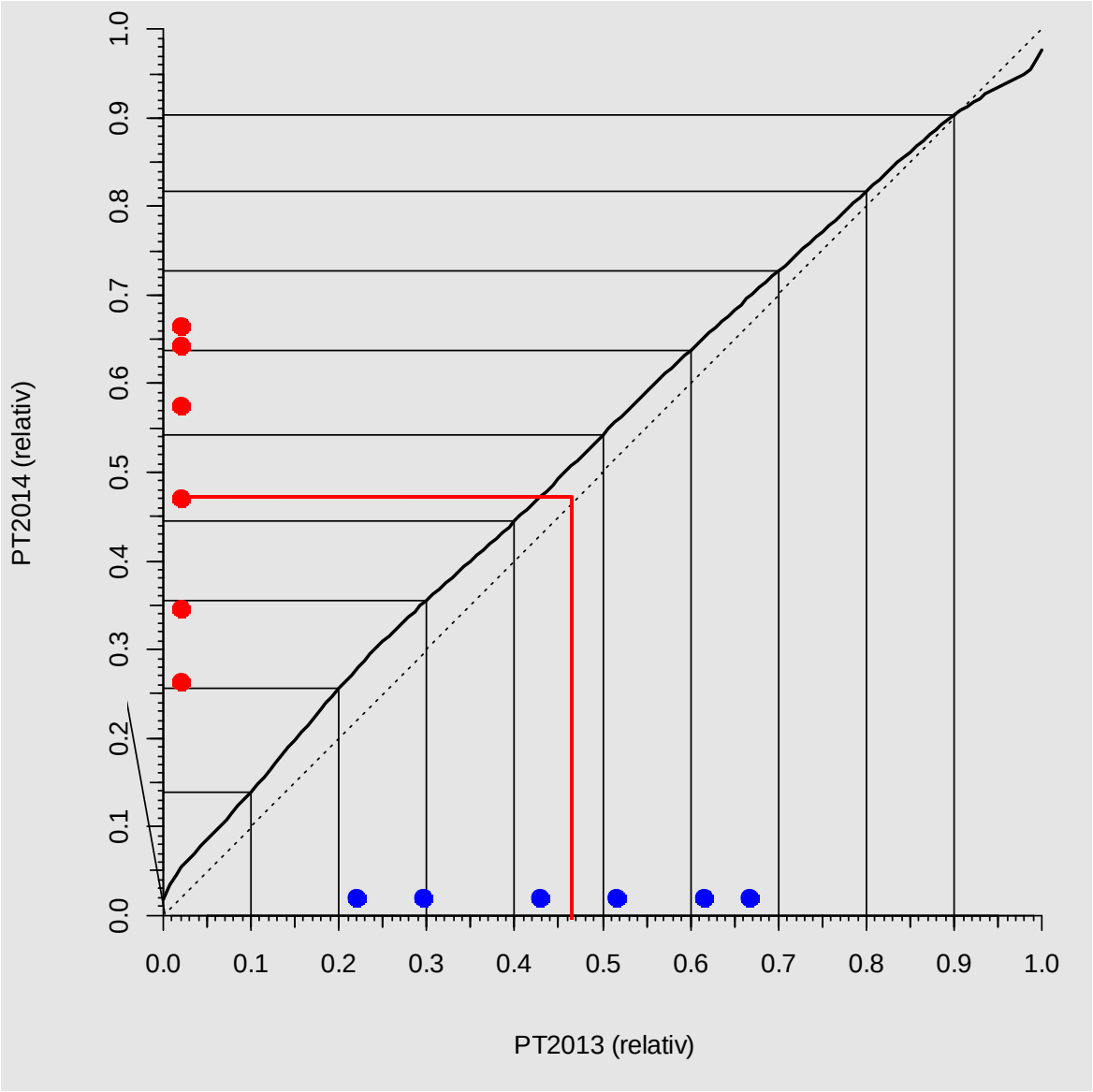
(Programm: R-Paket equate)

Ergebnisse OSE

Korrespondenz relativer Punkt-zahlen 2013/14.

Mittelwert nach Studienjahren

- 2013
- 2014



Zusammenfassung

1) Übliche Modelle der Item-Response-Theorie nur wenig geeignet zum Test-Equating der untersuchten Progresstests

2) Observed Score Equating:

PT 2014 leichter als PT 2013.

Unterschied im Bereich niedriger Scores größer als bei hohen Werten (entspricht knapp ½ Studienjahr)

Lineare Transformation nicht adäquat (Konsequenz für Bestehensgrenzen!)

Literatur

von Davier AA (Ed.). Statistical Models for Test Equating, Scaling, and Linking. Springer Science+Business Media, New York, 2010.

Kolen MJ, Brennan RL. Test Equting, Scaling, and Linking: Methods and Practices. Springer Science+Business Media, New York, 2014.

Wagener S, Möltner A, Tımbıl S, Gornostayeva M, Schultz JH, Brüstle P, Mohr D, Beken AV, Better J, Fries M, Gottschalk M, Günther J, Herrmann L, Kreisel C, Moczko T, Illg C, Jassowicz A, Müller A, Niesert M, Strübing F, Jünger J. Entwicklung eines formativen kompetenzbasierten Progresstests mit MC-Fragen von Studierenden – Ergebnisse einer multifakultären Pilotstudie. GMS - Zeitschrift für Medizinische Ausbildung (akzeptiert).