



Institut zur Qualitätsentwicklung
im Bildungswesen



Testdesign

Sebastian Weirich und Nicklas Hafiz

Institut zur Qualitätsentwicklung im Bildungswesen (IQB)
Humboldt-Universität zu Berlin

Gesis Workshop, Oktober 2024



Wann sind Testdesigns notwendig?



- In Large-Scale Assessments

- Wenn die zu messenden latenten Konstrukte durch eine so große Mengen von Items operationalisiert werden, dass eine Testperson alleine sie (schon aus Zeitgründen) nicht alle bearbeiten kann
 - Welche Testperson bearbeitet welche Items?
 - “item sampling”: Teilstichproben von Personen bearbeiten Teilstichproben von Testitems
- Wenn bestimmte Personengruppe bestimmte Itemgruppen (nicht) bearbeiten sollen
 - Schüler:innen mit sonderpädagogischem Förderbedarf (SPF) bearbeiten Aufgaben/Items, die in ihrer Schwierigkeit für diese Personen angepasst wurden
- Wenn in längsschnittlichen Untersuchungen gewährleistet werden soll, dass Testpersonen nicht wiederholt dieselben Aufgaben bearbeiten (Gedächtniseffekte), aber trotzdem eine Verlinkung zwischen Personen und Zeitpunkten gegeben sein soll



Wann sind Testdesigns notwendig



- Bedeutung von Testdesigns

- Das Testdesign definiert vorab, welche Analysen mit den später gewonnenen Daten durchgeführt werden können
- „schlechte“ Testdesign bewirken, dass zu den ohnehin restriktiven Annahmen des Raschmodells zusätzliche und **nicht testbare** Annahmen hinzukommen, die als gegeben gelten müssen, damit die Parameter sinnvoll interpretiert werden können
 - Bestimmte Analysen liefern dann ggf. keine verlässlichen Parameter mehr
- „Gute“ Testdesigns minimieren die Anzahl dieser zusätzlichen, durch sie implizierten und nicht testbaren Annahmen

- “You can’t fix by analysis what you bungled by design” (Light, Singer, and Willet, 1990)

- „spiralförmige“ Designs

	Items																			
Booklet	I1	I2	I3	I4	I5	I6	I7	I8	I9	I10	I11	I12	I13	I14	I15	I16	I17	I18	I19	I20
B1																				
B2																				
B3																				
B4																				



Large-Scale Assessments



- **Blockdesigns**

- Testitems, Aufgaben, Blöcke und Testhefte sind Designfaktoren
- Gruppieren von Aufgaben/Items zu festen Blöcken mit gleicher Bearbeitungszeit (bspw. 20 Minuten)
 - Designfaktoren Aufgabe/Item und Block sind disjunkt: jede Aufgabe und jedes Item kommt in genau einem Block vor, keine Überlappung
- Gruppieren von Blöcken zu Testheften. z.B. vier Blöcke mit $4 \times 20 = 80$ Minuten Gesamtbearbeitungszeit bilden ein Testheft/Booklet
 - Designfaktoren Block und Testheft sind nicht-disjunkt, sondern gekreuzt: ein Block kann in mehreren Testheften vorkommen
 - Designfaktoren Block und Testheft sind idealerweise balanciert



Large-Scale Assessments



- Wie balanciere ich Testdesigns sinnvoll?
 1. Wissen bzw. vorausplanen, welche Analysen mit den Daten durchgeführt werden sollen
 - Je umfangreicher und vielfältiger die Anzahl/Menge der Analysen, desto komplexer das Design
 2. Wissen bzw. bedenken, welche Faktoren ggf. unbeabsichtigt die Testperformanz beeinflussen, obwohl diese Faktoren *nicht* Gegenstand der Untersuchung sind
 - Bsp.: Moduseffekte, Positionseffekte, Motivationseffekte
 3. Wenn diese Faktoren nicht konstant gehalten werden können (Bsp.: alle Testpersonen bearbeiten den Test am selben Gerät), sollte versucht werden, sie auszubalancieren



Large-Scale Assessments



- Prüfen der Designfaktoren

Tag2_1Vormittag_Nr2_Testdesign.r

Je umfangreicher und vielfältiger die Anzahl/Menge der Analysen, desto komplexer das Design