

PISA-Studie

Ausgewählte Aspekte der Wirtschafts- und Sozialstatistik

Jenny Steindl

LMU München, Institut für Statistik

22.05.2015

- 1 Einführung
- 2 Kompetenzbereiche
- 3 Methodik

- 1 Einführung
- 2 Kompetenzbereiche
- 3 Methodik

- PISA = *Programme for International Student Assessment*
- OECD = *Organisation for Economic Co-Operation and Development*
- KMK = Ständige Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland

Was ist PISA?

- Internationale Schulleistungsstudie der OECD
- Nationaler sowie internationaler Vergleich der Leistungen von Schülern in verschiedenen Kompetenzbereichen
- Zyklisches Programm (alle 3 Jahre)

Teilnehmer:

- 2012: 65 Staaten (2000: 32 Staaten)
- 34 OECD Mitgliedsstaaten und 31 Partnerstaaten
- ca. 500.000 Schülerinnen und Schüler stellvertretend für ca. 28 Mio. 15-jährige
- Deutschland: 5001 Schüler in 230 Schulen

Teilnehmer:

Albanien*	Jordanien*	Republik Serbien*
Argentinien*	Kanada	Rumänien*
Australien	Kasachstan*	Russische Föderation*
Belgien	Katar*	Schweden
Brasilien*	Kolumbien*	Schweiz
Bulgarien*	Korea	Shanghai (China)*
Chile	Kroatien*	Singapur*
Chinesisch Taipeh*	Lettland*	Slowakische Republik
Costa Rica*	Liechtenstein*	Slowenien
Dänemark	Litauen*	Spanien
Deutschland	Luxemburg	Thailand*
Estland	Macau (China)*	Tschechische Republik
Finnland	Malaysia*	Tunesien*
Frankreich	Mexiko	Türkei
Griechenland	Neuseeland	Ungarn
Hongkong (China)*	Niederlande	Uruguay*
Indonesien*	Norwegen	Vereinigte Arabische Emirate*
Irland	Österreich	Vereinigte Staaten
Island	Peru*	Vereinigtes Königreich
Israel	Polen	Vietnam*
Italien	Portugal	Zypern*
Japan	Republik Montenegro*	

Anmerkung: * OECD-Partnerstaaten

Quelle: (Prenzel M., et al. (2013), S. 22)

Durchführung:

- Schriftliche Tests mit Multiple Choice Fragen und offenen Fragen)
- erstmals Hintergrundfragebögen
- Erhebung im Zeitraum 01. März bis 30. Juni 2012
- 2 Stunden / Test

- 1 Einführung
- 2 Kompetenzbereiche
- 3 Methodik

- 1 Einführung
- 2 Kompetenzbereiche
- 3 Methodik

- Lesekompetenz
- Mathematische Grundbildung
- Naturwissenschaftliche Grundbildung

Grundbildungen:

- Mathematische Grundbildung als Schwerpunkt von PISA 2012 (zuletzt PISA 2003)
- Naturwissenschaftliche Grundbildung (zuletzt PISA 2006)
- Lesen als Grundbildung (zuletzt PISA 2009)

Kompetenzbereiche

DIE RADFAHRERIN HEIKE – FRAGE 1

Auf einer Tour ist Heike 4 km in den ersten 10 Minuten gefahren und dann 2 km in den nächsten 5 Minuten. Welche der folgenden Aussagen ist richtig?

- A. Heikes Durchschnittsgeschwindigkeit war in den ersten 10 Minuten höher als in den nächsten 5 Minuten.
- B. Heikes Durchschnittsgeschwindigkeit war in den ersten 10 Minuten und in den nächsten 5 Minuten die gleiche.
- C. Heikes Durchschnittsgeschwindigkeit war in den ersten 10 Minuten niedriger als in den nächsten 5 Minuten.
- D. Es ist nicht möglich, anhand der Angaben etwas über Heikes Durchschnittsgeschwindigkeit zu sagen.

Bewertung

Beschreibung: Vergleich von Durchschnittsgeschwindigkeiten, Streckenangaben und Zeitangaben

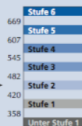
Mathematischer Inhaltsbereich: Veränderung und funktionale Abhängigkeiten

Kontext: Persönlich

Prozess: Anwenden

Aufgabentyp: Einfache Multiple-Choice-Aufgabe

Schwierigkeitsgrad: 440,5



Volle Punktzahl

- B. Heikes Durchschnittsgeschwindigkeit war in den ersten 10 Minuten und den nächsten 5 Minuten die gleiche.

Keine Punkte

- Andere Antworten.
- Keine Antwort.

OECD (2013), S. 135

Kompetenzbereiche

BERGSTEIGEN AM MOUNT FUJI – FRAGE 1

Der Mount Fuji ist für die Öffentlichkeit jedes Jahr nur vom 1. Juli bis 27. August zur Besteigung freigegeben. Ungefähr 200.000 Menschen besteigen den Mount Fuji während dieser Zeit.

Wie viele Menschen besteigen den Mount Fuji durchschnittlich pro Tag?

- A. 340
- B. 710
- C. 3 400
- D. 7 100
- E. 7 400

Bewertung

Beschreibung: Identifizierung eines durchschnittlichen Tageswerts anhand einer Gesamtzahl und eines konkreten Zeitraums (mit Datenangaben)

Mathematischer Inhaltsbereich: Quantitatives Denken

Kontext: Gesellschaftsbezogen

Prozess: Formulieren

Aufgabentyp: Einfache Multiple-Choice-Aufgabe

Schwierigkeitsgrad: 464 *



Volle Punktzahl

- C. 3 400

Keine Punkte

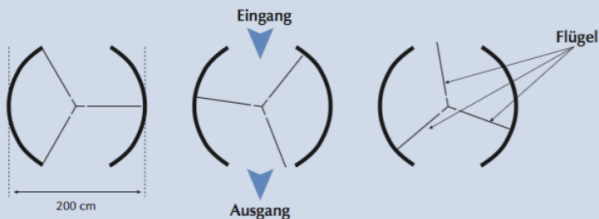
- Andere Antworten.
- Keine Antwort.

OECD (2013), S. 138

Kompetenzbereiche

DREHTÜR

Eine Drehtür hat drei Flügel, die sich innerhalb eines kreisförmigen Raumes drehen. Der Innendurchmesser dieses Raumes beträgt 2 Meter (200 Zentimeter). Die drei Türflügel teilen den Raum in drei gleichgroße Sektoren. Der Plan unten zeigt die Türflügel von oben gesehen in drei verschiedenen Positionen.



DREHTÜR – FRAGE 1

Wie groß ist der Winkel (in Grad), der von zwei Türflügeln gebildet wird?

Größe des Winkels:°

OECD (2013), S. 141

Kompetenzbereiche

Bewertung

Beschreibung: Berechnung des Zentriwinkels eines Kreissektors

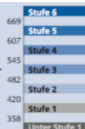
Mathematischer Inhaltsbereich: Raum und Form

Kontext: Wissenschaftsbezogen

Prozess: Anwenden

Aufgabentyp: Aufgabe mit offenem Antwortformat (manuelle Kodierung)

Schwierigkeitsgrad: 512,3 *



Volle Punktzahl

120 [die Angabe des entsprechenden überstumpfen Winkels – 240 – wird ebenfalls akzeptiert].

Keine Punkte

Andere Antworten.

Keine Antwort.

OECD (2013), S. 141

CBA:

- Auswahl aus 3 Kompetenzbereichen Problemlösung, Lesen und Mathematik
- Deutschland entscheidet sich für Problemlösung
- Testungen sind am PC durchzuführen
- Durchführung im Anschluss an schriftliche Tests

- 1 Einführung
- 2 Kompetenzbereiche
- 3 Methodik

- 1 Einführung
- 2 Kompetenzbereiche
- 3 Methodik

Stichprobenerhebung:

- Vollerhebung aus finanziellen und zeitlichen Gründen nicht durchführbar
- Systematische Teilerhebung
- Zwei- oder mehrstufige Auswahlverfahren bei *Large-Scale-Assessments*

Stichprobenplan und Ziehung der Stichprobe:

- Grundgesamtheit: Kohorte der 15-jährigen Schülern ab der 7. Klasse (geboren zwischen dem 01.01.1996 und dem 31.12.1996)
- Zweistufiges Auswahlverfahren in Deutschland eingesetzt

1. Schritt: Erstellung eines Sampling-Frame:

- Bestimmung aller Schulen, die potentiell von 15-jährigen besucht werden
- keine Schule darf mehrfach, fehlerhaft oder unvollständig in den Sampling-Frame aufgenommen werden
- Informationen von 14 statistischen Landesämtern

Eingeholte Informationen auf Basis der Schulstatistik 2010/2011:

- offizielle Schulnummer
- Schulart
- Anzahl der Schüler (geb. 1993-1996)
- Anzahl der Schüler in den 8.-10. Klassen
- Anzahl der Klassen 8.-10.
- Informationen über Schulveränderungen
- Förderschwerpunkte von Förderschulen

2. Schritt: Ziehung der Schülerstichprobe

- Unterteilung des Sampling-Frames anhand mehrerer Stratifizierungskriterien
Vorteile: Erhöhung der Effizienz des Sampling-Frames
- Populationsparameter kann verlässlicher geschätzt werden
- verlässliche Ergebnisse auch mit kleineren Stichproben

- Stratifizierung ermöglicht Überprüfung, ob die Charakteristika der Stichprobe mit jenen des Sampling-Frames übereinstimmen
- Analyse nach bestimmten Kriterien möglich, so dass beispielsweise Ergebnisse einer Hauptschule mit jenen einer Realschule vergleichbar sind

Unterscheidung zwischen expliziter und impliziter Stratifizierung:

Explizite Stratifizierung:

- Zunächst Unterteilung anhand bestimmter Charakteristika in unabhängige Gruppen
- Ziehung separater Stichproben aus den unabhängigen Gruppen

Implizite Stratifizierung:

- Die aus den expliziten Strata gezogenen Stichproben werden in einzelne Merkmalsklassen unterteilt

In Deutschland wird die Grundgesamtheit aller Schulen mittels zweier expliziter und zweier impliziter Stratifizierungsverfahren gezogen:

- ① Unterteilung der Schulen in allgemeinbildende Schulen, Förderschulen und Berufsschulen (explizites Stratifizierungsverfahren)
- ② Aufteilung der Schulen nach Bundesländern (explizites Stratifizierungsverfahren)
- ③ Aufteilung der Berufs- und Förderschulen nach Bundesländern (implizites Stratifizierungsverfahren)
- ④ Aufteilung der Allgemeinbildenden Schulen nach Hauptschulen, Realschulen und Gymnasien (implizites Stratifizierungsverfahren)

→ Aufteilung der Schulen in 18 explizite Strata (16 Bundesländer und Berufs- und Förderschulen)

- Ziehung der Stichproben der jeweiligen Strata durch das internationale PISA Konsortium
- 247 in die Stichprobe gezogene Schulen
- Proportionaler Anteil der Schüler in den Stichproben soll approximativ jenem in der Grundgesamtheit des Staates entsprechen

Methodik

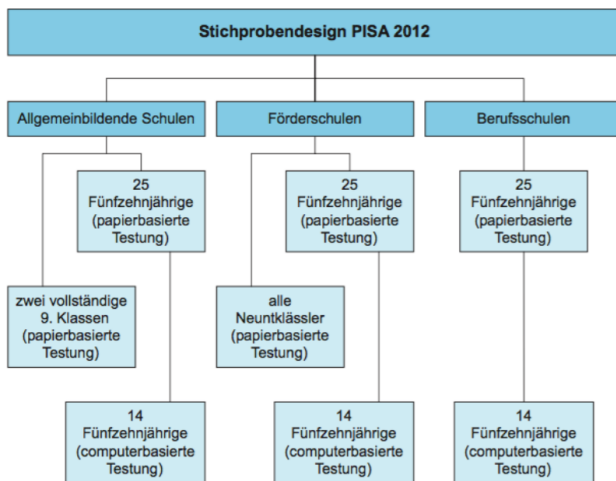
	Haupt- schule	Inte- grierte Gesamt- schule	Schule mit mehreren Bildungs- gängen	Real- schule	Gymna- sium	Förder- schule	Berufs- schule	Gesamt
Baden- Württemberg	2	1	5	13	12	2	4	39
Bayern	10	1	0	13	12	1	6	43
Berlin	0	0	4	0	3	0	0	7
Brandenburg	0	0	2	0	2	1	0	5
Bremen	0	0	1	0	1	0	0	2
Hamburg	0	0	2	0	2	0	1	5
Hessen	2	3	0	5	6	1	1	18
Mecklenburg- Vorpommern	0	0	2	0	1	0	0	3
Niedersachsen	4	2	0	9	8	2	1	26
Nordrhein- Westfalen	10	10	0	15	18	3	2	58
Rheinland-Pfalz	0	2	4	1	4	1	0	12
Saarland	0	0	1	0	1	0	1	3
Sachsen	0	0	3	0	3	1	1	8
Sachsen-Anhalt	0	0	2	0	2	0	0	4
Schleswig- Holstein	0	4	1	0	3	1	0	9
Thüringen	0	0	3	0	1	0	1	5
Gesamt	28	23	30	56	79	13	18	247

Quelle: (Prenzel M., et al. (2013), S. 316)

3. Schritt: Ziehung der Schülerstichprobe:

- Beschaffung von Informationen der Schüler vor Stichprobenziehung: Vor- und Nachname, Geschlecht, Geburtsdatum, Klassenbezeichnung, möglicher Förderbedarf
- Verwendung von Pseudonymen, um Datenschutz zu gewährleisten

- Ziehung einer Stichprobe von 25 Schülern pro Schule aus den 247 ausgewählten Schulen (allgemeinbildende Schule)
- Stichprobe von zwei 9. Klassen (allgemeinbildende Schulden), die vollständig in die Stichprobe mit aufgenommen werden
- Ziehung von 14 Schülern aus den 25 gezogenen, die am Testprogramm CBA teilnehmen
- Vollerhebung bei allen ausgewählten Berufs- und Förderschulen



Quelle: (Prenzel M., et al. (2013), S. 317)

Realisierte Stichproben: Unterscheidung zwischen Brutto- und Nettostichproben

- **Bruttostichprobe:** Alle für die PISA-Studie 2012 ausgewählten Schüler
- **Nettostichprobe:** Alle Schüler, die tatsächlich an der Studie teilgenommen haben

- Bei Nicht-Teilnahme wurden Ersatzschulen bestimmt
- Ausfälle von 17 Schulen

- 230 Schulen nahmen tatsächlich an der PISA-Studie teil
- Mindestvorgabe der OECD Standards erfüllt
- **Teilnehmerquote:** Schulen: 98,3%, Schüler: 93,2%

Gewichtung: Einzelne Ziehungswahrscheinlichkeiten werden durch mehrstufiges Ziehungsverfahren beeinflusst

Unterteilung der Gewichte in:

- Schulbasisgewichte
- Schülerbasisgewichte und Schülergesamtgewicht

Schulbasisgewicht:

- Durch Aufteilung der Sampling-Frames in 18 Strata und die Sortierung der Schulen nach Bundesland haben einige Schulen eine niedrigere Wahrscheinlichkeit in die Stichprobe zu gelangen als andere
- Ungleichheit wird durch umgekehrt proportionale Gewichtung ausgeglichen

Schülerbasisgewichte und Schüलगesamtgewicht

- Ziehungswahrscheinlichkeiten der einzelnen Schüler je nach Größe der Schule unterschiedlich
- Damit Aussage auf Ebene der Grundgesamtheit möglich, wurde jedem Schüler *post hoc* die selbe Ziehungswahrscheinlichkeit zugeordnet (umgekehrt proportional)
- Berechnung des Gesamtgewichts aus einzelnen Schülerbasisgewichten

Notwendigkeit einer Gewichtung: Verzerrung der Schätzungen sollen vermieden werden

→ Korrekturfaktoren

- Kompensierung von Schulausfällen
- Gewichtung auf Staatsebene
- Ausgleich durch Nicht-Teilnahme von Schülern
- *Trimming Factors*

Skalierung von Leistungstests zu Modellierung von Kompetenzen:

Konflikt zwischen praktischer Durchführbarkeit und psychosomatischen Anforderungen

- Erfassung latenter Eigenschaften in den drei Kompetenzbereichen
- Möglichst große Anzahl an Fragen (*Items*)
- begrenzte Testzeit

Bei *Large-Scale-Assessments* wird zur Gestaltung solcher Fragebögen ein sog. *Balanced Incomplete Block Design* verwendet.

Grundüberlegung:

- Nicht jeder Schüler erhält alle entwickelten Fragen
- Erstellung mehrerer Testhefte
- Unterschiedliche Auswahl von Aufgabengruppen in den einzelnen Testheften

→ **Multi-Matrix-Design**

Das PISA-Multi-Matrix-Design und IRT-Skalierung:

Beschreibung der PISA-Items als manifestierte Indikatoren kognitiver Kompetenz

- Gruppierung der einzelnen Items um einen Aufgabenstamm in den Kompetenzbereichen (Testlet oder Unit)
- Testlets oder Units enthalten Textelemente, Grafiken, Tabellen oder Kombinationen hieraus

- Zwischen einer und sieben Items pro Unit
- Insgesamt 87 Units (=207 Items)
- *Link-Unit*: Übernommene Aufgaben aus früheren PISA Erhebungen
- Ausschließliche Verwendung von Link-Units in den Kompetenzbereichen Lesen und Naturwissenschaften
- Über 50% der Aufgaben für den Schwerpunkt Mathematik neu erstellt

Eingesetzte Units und Items

	Units gesamt	Items gesamt	Link-Units	Link-Items
Mathematik	56	110	25	36
Lesen	13	44	13	44
Naturwissenschaften	18	53	18	53
Gesamt	87	207	56	133

Quelle: (Prenzel M., et al. (2013), S. 324)

Aufbau der Testhefte:

- Unterteilung der Units für die drei Kompetenzbereiche in 15 Cluster
- Für den Schwerpunkt Mathematik wurden 9 Cluster erstellt (PM1 - PM5, PM6a+b, PM7a+b)
- PM6a und PM7a wurde in allen Staaten eingesetzt, die im Bereich Mathematik in PISA 2009 mindestens 480 Punkte erreichten; PM6b und PM7b in allen anderen Staaten (leichter)

Kompetenzbereiche Lesen und Naturwissenschaften:

- Erstellung von 3 Clustern (PR1 - PR3 bzw. PS1 - PS3)
- Bearbeitungszeit für jedes Cluster ca. 30 Minuten von insgesamt 120 Minuten

Insgesamt bei der Haupterhebung 15 Cluster in 13 Standardheften, wobei jedes der Testhefte vier Aufgabengruppen enthielt

Aufbau der Testhefte

Testheft													
Dauer	1	2	3	4	5	6	7	8	9	10	11	12	13
30	M5	N3	L3	M6	M7	M1	M2	N2	L2	M3	M4	N1	L1
30	N3	L3	M6	M7	N1	M2	N2	L2	M3	M4	M5	L1	M1
30	M6	M7	N1	L1	M1	L1	M3	M4	M5	N3	L3	M2	N2
30	N2	L2	M3	M4	M5	M6	M7	N1	L1	M1	M2	N3	L3
45	Fragebogen												

Anmerkung: M = Mathematik, N = Naturwissenschaften, L = Lesen.

Quelle: (Prenzel M., et al. (2013), S. 324)

Vergleichbarkeit der Schülerleistungen untereinander:

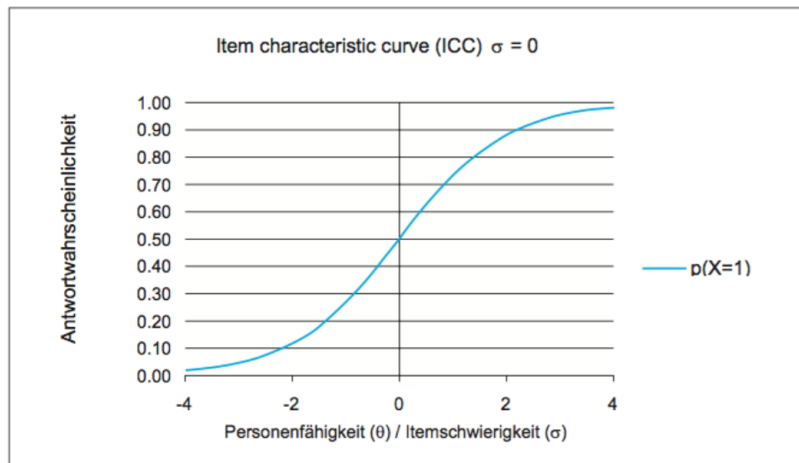
- Jeder Schüler erhält unterschiedliches Testheft mit unterschiedlichen Aufgaben
- Schwierigkeit der Testhefte weitestgehend gleich, Aufgaben dennoch unterschiedlich schwer
- Anzahl der gelösten Aufgaben kann nicht als Vergleichswert herangezogen werden

Latente (kontinuierliche) Personenvariable:

- *Item-Response-Theory* (IRT) zur Auswertung der PISA-Studie
- Latente und individuelle Personeneigenschaften sind anhand des Antwortverhaltens gegeben
- Zusammenhang zwischen Ausprägungen auf der latenten Variable und der beobachtbaren Lösung mittels *Rasch-Modell* und dessen Erweiterung dem *Partial-Credit-Modell*

Das Rasch-Modell:

- Modellierung der Wahrscheinlichkeit, dass ein Schüler eine Aufgabe richtig beantwortet
- Wahrscheinlichkeit wird durch folgenden Parameter modelliert:
- Personenparameter: gibt Fähigkeit der getesteten Person an
- Itemparameter: beschreibt den Schwierigkeitsgrad der Aufgabe
- beschreibt neben den beiden genannten Parametern zusätzlich eine psychologisch plausible, stochastisch probabilisierte Beziehung der beiden Parameter
- Beziehung der Parameter fließt als logistische Funktion in das Modell ein
- Grafische Darstellung mittels *Item-Characteristic Curve*(ICC)



Quelle: (Prenzel M., et al. (2013), S. 327)

Interpretation der *Item-Characteristic Curve*:

- Mit zunehmender positiver Differenz zwischen dem Personenparameter und dem Itemparameter nimmt die Wahrscheinlichkeit der Lösung einer Aufgabe oberhalb der 50% Lösungswahrscheinlichkeit zu
- Lösungswahrscheinlichkeit des betreffenden Items ist größer 50%, wenn Personenparameter $>$ Itemparameter

Formal:

σ Itemparameter

θ Personenparameter

$p(X)$ Lösungswsk.

Person v

Item i

$$p(X_{vi}) = \frac{\exp(X_{vi}(\theta_v - \sigma_i))}{1 + \exp(\theta_v - \sigma_i)}$$

wobei $X \in 0, 1$

- Wenn $X_{vi} = 0$: Es wurde eine falsche Antwort gegeben
- Wenn $X_{vi} = 1$: Es wurde eine richtige Antwort gegeben
- keine teilweise richtigen Antworten zugelassen

→ *Partial-Credit-Modell* als Erweiterung des *Rasch-Modells*

Das Partial-Credit-Modell:

- Mehrstufige Antwortformate zugelassen
(richtig / teilweise richtig / falsch)
- Zerlegung des Itemparameters in einzelne Schwellenwerte

σ Itemparameter, θ Personenparameter, s Schwellenwerte, σ_{ix} realisierter Schwellenparameter, σ_{is} möglicher Schwellenparameter, $p(X)$ Lösungswsk., Person v , Item i , $m + 1$ mögliche Kategorien

$$p(X_{vi} = x) = \frac{\exp((x\theta_v)) - \sigma_{ix}}{\sum_{s=0}^m \exp((s\theta_v) - \sigma_{is})}$$

wobei $x \in 0, 1, \dots, m$

Methoden zur Parameterschätzung

- 3 Verfahren zur Parameterschätzung
- Stützung auf die Likelihood der Daten (wie gut passen bestimmte Werte der Parameter zu den Daten?)
- Ziel: Bestmögliche Erklärung der empirisch beobachteten Daten

Methoden zur Parameterschätzung

- JML (Joint-Maximum-Likelihood-Schätzung)
- CML (Conditional-Maximum-Likelihood-Schätzung)
- MML (Marginal-Maximum-Likelihood-Schätzung)

Methoden zur Parameterschätzung

JML- Methode:

- σ und θ werden gleichzeitig bestimmt
- Aufstellen und Nullsetzen der Likelihood um Maximum zu bestimmen
- Mehrere unbekannte Größen \rightarrow Iterative Lösung
- Schätzung des Itemparameters in der Praxis genauer als Schätzung des Personenparameters
- Problematik: zusätzliche Person in der Stichprobe \rightarrow zusätzlicher zu schätzender Parameter

Methoden zur Parameterschätzung

CML-Methode:

- Parameter werden in zwei Schritten getrennt geschätzt
- Zuerst Schätzung des Itemparameter, auf dessen Basis dann Schätzung des Personenparameters
- konsistente und unverzerrte Item- & Parameterschätzung möglich
- keine a-priori-Annahmen über die Verteilung der latenten Personenvariable
 - Verwendung dieses Designs, wenn sämtliche Personen in der Stichprobe dieselbe Auswahl an Items bearbeiten
 - Schätzung über Scoregruppe (Anzahl gelöster Items)

Methoden zur Parameterschätzung

MML-Methode:

- Parameterschätzung über Normalverteilungsannahme mit Mittelwert und Streuung
- getrennte Schätzung der Parameter um unverzerrte Schätzwerte zu erhalten
- Personenparameter werden über die Verteilungsannahme herauskonditioniert
 - Populationsmodell
 - komplexere Modelle

Vorteil: Stabile Schätzung, da die Anzahl der geschätzten Parameter nicht so extrem ansteigt

Vielen Dank
für Ihre
Aufmerksamkeit!