Psychometrie

- Bias
  - Invloed van groepssamenstelling op testscores
  - Hoe homogener de scores op twee testen, hoe lager de correlatie tussen deze twee testen: restriction of range (naast de aantasting/attenuatie door meetfouten of onbetrouwbaarheid)
- Validiteit
  - Typen
    - Inhoudsvaliditeit: de inhoud van de test moet representatief zijn voor de te meten variabelen
    - Criteriumvaliditeit: voorspellende (predictieve) waarde van een testscore
    - Begripsvaliditeit: correlaties met andere variabelen
  - Een testscore kan niet hoger met een andere variabele correleren dan met zijn eigen ware score. Dus de correlatie met een andere variabele (ρXY), is nooit hoger dan de correlatie tussen de geobserveerde scores en de ware scores (ρXT), wat gelijk is aan de wortel uit de betrouwbaarheidscoefficient (√ρXX')
    - ρXY ≤ ρXT, dus ρXY ≤ √ρXX'
  - Correlatie tussen de ware scores van twee variabelen
    - ρTxTy = ρXY / √(ρxx' * ρyy')
    - Correlatie geobserveerde scores gedeeld door de wortel uit het product van de betrouwbaarheidscoefficienten
    - Als testen 100% betrouwbaar zijn: ρXY = ρTxTy
  - Correlatie tussen de ware scores op de ene test (X) en de geobserveerde scores op de andere test (Y)
    - ρXY / √(ρxx')
  - Als de betrouwbaarheidscoëfficiënt van de scores op de ene test lager zou zijn en de correlatie met de scores op de andere test blijft gelijk, dan zou de correlatie tussen de ware scores op de ene test en de ware scores op de andere test hoger uitvallen, omdat dan voor meer onbetrouwbaarheid wordt gecorrigeerd.
  - Multitrait-multimethod matrix
    - Vaststellen begripsvaliditeit
    - Op de diagonaal staan de betrouwbaarheidscoefficienten
    - Indicatief voor convergerende validiteit: correlatie tussen hetzelfde construct via beide methoden
    - Indicatief voor divergerende validiteit: alle correlaties tussen verschillende constructen (binnen methode A, methode B en tussen methode A & B)
  - Convergerende validiteit
  - Divergerende validiteit
- Klassieke testtheorie
  - Assumpties
    - X = T + E; T = X - E
    - ε(X) = T: T (de ware score) is gelijk aan de verwachte waarde van X (de geobserveerde score)
    - ρET = 0; geen correlatie tussen T en E. Wel possitieve correlatie mogelijk indien één groep (de hoge scoorders) wordt bevooroordeeld tijdens de afname en een ander groep (de late scoorders) wordt benadeeld
    - Ware score onmogelijk te meten omdat er altijd meetfouten zijn en omdat de afnames in de praktijk nooit onafhankelijk zijn
    - Meetfout: niet-systematische of toevallige afwijking van de geobserveerde testscore van een persoon van zijn ware score (= de verwachte geobserveerde score).
    - Iemand met een hogere/lagere geobserveerde X scoort niet per definitie beter/slechter dan een ander. Hangt af van systematische meetfouten (validiteit; meet de test wel het contruct goed) en toevallige meetfouten (betrouwbaarheid; errorvariantie/residu)
  - $\rho^2XT = \sigma^2T/\sigma^2X = \rho XX'$
  - $\rho^2XT = \sigma^2T/\sigma^2X = 1 - \sigma^2E/\sigma^2X$
  - $\sigma^2E/\sigma^2X = \rho^2XE = 1 - \rho XX'$
  - Ware scorevariantie (σ^2T): ρXX' * geobserveerde scorevariantie
  - Error scorevariantie (σ^2E): 1 - ware scorevariantie
  - Betrouwbaarheid van de meting: verhouding error score variantie tov ware score variantie
  - Paralleltests
    - Nut: ware score beter kunnen benaderen en betrouwbaarheid bepalen met behulp van parallelvorm-benadering
    - Paralleltests hebben gelijke geobserveerde score-varianties en gelijke correlaties met andere (geobserveerde) testscores
    - Assumpties
      - Gelijke ware scores: T = T'
      - Gelijke errorscore varianties: $\sigma^2E = \sigma^2E'$
      - Gelijke geobserveerde score varianties: $\sigma^2X' = \sigma^2X$
      - X(gem) = X'(gem)
      - Dezelfde correlaties tussen beide tests en geobserveerde score van andere tests: ρX'Y = ρXY
  - Essentieel tau-equivalent
    - Essentieel tau-equivalent wanneer geldt: T = T' + c
    - Assumpties paralleltest gelden niet
- Componentenanalyse
  - Doel
    - Datareductie (het reduceren van het aantal variabelen)
    - Het bepalen van de dimensionele structuur (naar hoeveel dimensies verwijzen de items)
  - Eigenwaarde: proportie itemvariantie dat verklaard kan worden door een component (eigenwaarde component a / aantal componenten * 100 = x % van de totale itemvariantie wordt door component a verklaard
  - Communalititeit (h2): x % van de variantie van item a kan worden verklaard door de componenten (bijv. 0,390 = 39). Communaliteit te berekenen door correlatie tussen item en componenten te kwadrateren (-,624 in het kwadraat = 0,390)
  - Correlaties tussen items en het component worden ook wel ladingen genoemd
  - Som van alle communaliteiten (gekwadrateerde correlaties tussen items en component) is de eigenwaarde van component 1
  - Convergerende validiteit: welk item heeft de hoogste correlatie met component 1 (of bij welk item wordt de meeste variantie verklaard door component 1)
  - Divergerende validiteit: kan alleen bepaald worden bij 2 of meer componenten
  - Rotatie
    - Het op een andere wijze presenteren van dezelfde data
    - Om de divergerende validiteit te verhogen
    - Het totaal aan verklaarde variantie veranderd niet na rotatie
    - Wel de verklaarde variantie per component veranderd; het verschil tussen de componenten wordt kleiner, op itemniveau wordt het verschil in de proportie verklaarde variantie tussen de componenten groter (het item past na rotatie beter bij het 1 van de x componenten; de divergerende validiteit neemt toe)
  - Validiteit
    - Convergerende validiteit: elk item heeft op een component een hoge lading
    - Divergerende validiteit: elk item heeft op 1 component een hoge lading en op de andere componenten een significant lagere lading
    - Inhoudelijke validiteit: passen de items bij elkaar die onder een component vallen?
- Transformeren
  - Om betekenis aan scores te geven (moeilijk om ruwe scores te interpreteren als je geen vergelijking met andere personen kunt maken)
  - Lineaire transformatie: rechtlijnig verband tussen ruwe en getransformeerde scores (verdeling en correlaties hetzelfde), bijvoorbeeld standaardscores (gem = 0, Sd = 1) of gestandaardiseerde scores (met een bepaald gem. en Sd)
    - Standaardscore: $Z = (X - \mu x)/\sigma x$
    - Gestandaardiseerde scores: $Y = \sigma * Z + \mu$
  - Niet-lineaire transformatie: percentielscores (leidt bij een beperkt aantal mogelijke scores en homogene groepen tot een sterke uitvergroting van kleine verschillen in ruwe scores), genormaliseerde standaardscores (scores worden getransformeerd naar een normaal verdeling, gem. 0, Sd 1; zijn gemakkelijk te interpreteren en met elkaar te vergelijken, voorwaarde is wel dat de onderliggende eigenschap normaal verdeeld is)
    - Genormaliseerde standaardscores (Nx): eerst percentielscores berekenen, dan Z scores
    - T-scores: genormaliseerde scores met gem. 50 en Sd 10: $T = 10 * Nx + 50$
    - Stanines: 1-cijferige genormaliseerde scores van 1 tot 9 (gem. 5, Sd +/- 2)
  - Monotonie: volgorde van de personen in de verdeling verandert niet door transformatie
  - Leeftijdsscores
    - Mediane score op een test op een bepaalde leeftijd
    - Voorwaarde is dat de eigenschap monotoon toe- of afneemt met leeftijd/niveau
    - Niet zo bruikbaar omdat slechts zeer weinig mensen precies de score zullen hebben die verwacht wordt op basis van hun leeftijd/niveau

eoordelaarsovereenstemming

- Proportie overeenstemming (Po)
- Verwachte proportie overeenstemming (Pe; overeenstemming op basis van toeval): som van rijproporties x kolomproporties
- Cohen's Kappa: (Po - Pe) / (1 - Pe); lager dan Po, omdat gecorrigeerd wordt voor toevallige overeenkomsten
- Overeenstemming tussen 3 of meer beoordelaars
  - Simpele methode: alle Kappa's bij elkaar optellen en dan delen door aantal beoordelaars (gemiddelde Kappa)
  - Methode 'Popping': gemiddelde Po en Pe berekenen, dan in formule Kappa deze gemiddelden gebruiken
  - Simultane overeenstemming: hoe vaak zijn alle beoordelaars het met elkaar eens, gedeeld door totaal aantal beoordelingen (Po), dan Pe = som van het product van randproporties per categorie van de beoordelaars, dan formule Kappa gebruiken\
- Beste beoordelaar: Kappa's met andere beoordelaars bij elkaar optellen en delen door aantal overige beoordelaars (middelen)
- Gewogen Kappa (Kw)
  - Met behulp van een gewichtenmatrix
  - Po: som van product van gewicht x proportie overeenstemming
  - Pe: som van product van gewicht x rijproporties x kolomproporties
- Gewichtenmatrix
  - Lineaire gewichtenmatrix: gelijkmatige verdeling van gewichten, gelijke stappen
    - $w_{ij} = 1 - \{ |i - j| / (c - 1) \}$
    - i = index categorie a; j = index categorie b; c = totaal aantal categorieen
  - Kwadratische gewichtenmatrix: groter verschil tussen 2 categorieen wordt zwaarder gestraft dan klein verschil tussen 2 categorieen
    - $w_{ij} = 1 - \{ (i - j)2 / (c - 1)2 \}$
- Intervalniveau
  - 3 variantiebronnen: personen, beoordelaars en interactie personen-beoordelaars (residu)
  - Df
    - Personen (p): n - 1
    - Beoordelaars (b): k - 1 (aantal beoordelaars - 1)
    - Residu: (n-1)(k-1)
    - Totaal: N - 1 (aantal scores - 1)
  - Overeenstemmingscoefficient (wordt lager bij minder beoordelaars; vul bij k in het gewenste aantal beoordelaars of gebruik Spearman-Brown formule met verkorting/verlengingsfactor (vul bij n in), vul bij r de gevonden overeenstemmingscoefficient in)
    - $\frac{\hat{\sigma}^2_p}{\hat{\sigma}^2_p + (\hat{\sigma}^2_b + \hat{\sigma}^2_{res}) / k}$
  - Betrouwbaarheidscoefficient (variantiecomponent van beoordelaars wordt buiten beschouwing gelaten; als deze klein is, dan ook klein verschil, wel iets hoger dan overeenstemmingscoef omdat bron dan wel ietsje afbreuk doet aan overeenstemming; wordt hoger bij meer beoordelaars, zie uitwerking bij overeenstemmingscoef, vul voor r de gevonden betrouwbaarheidscoef in, vul bij k het gewenste aantal beoordelaars)
    - $\frac{\sigma^2_p}{\sigma^2_p + \sigma^2_{res} / k}$
  - Verlengingsfactor bereken: xx' is gewenste coefficient, yy' is gevonden coefficient
    - $N = \frac{\rho_{xx'}(1 - \rho_{yy'})}{\rho_{yy'}(1 - \rho_{xx'})}$
    - $r_{xx'} = \frac{n * r}{1 + (n - 1) * r}$

eschrijvende statistiek

- Variantie: som (Xi - Xgem. in het kwadraat) gedeeld door N-1
- Standaardafwijking: wortel van de variantie
- Covariantie: som(product (Xi-Xgem) x (Yi-Ygem)) gedeeld door N - 1
- Correlatiecoefficient: covariantie gedeeld door product SdX x SdY

emanalyse

- Moeilijkheidsgraad: p-waarde
- Moeilijkheidsindex: volgens ICC de somscore waarbij 50% van de leerlingen het item goed heeft beantwoordt
- Hoe lager de moeilijkheidsgraad (p-waarde), hoe hoger de moeilijkheidsindex
- Discriminerend vermogen: hoe hoger de item-totaalcorrelatie, hoe hoger het discriminerend vermogen
- ICCs: zeggen iets over in welk gebied het item het beste discrimineert: over het hele gebied, of alleen bij de slechtere/betere leerlingen
- Evocatiegraad: mate waarin het item het te meten gedrag uitlokt, bij test-items te bepalen door naar gemiddelde te kijken, hoe hoger het gemiddelde, hoe hoger de evocatiegraad
- Proportie verklaarde variantie: r (correlatie) in het kwadraat. x% van de variantie van de scores op item 2 kan worden verklaard door de scores op item 1

etrouwbaarheid

- Test-hertest, parallelvorm, splithalf, interne consistentie
- Bij paralleliteit (gelijke gemiddelde en varianties): schatting van betrouwbaarheid op basis van correlatie tussen beide helften: 2 x correlatiecoefficient / 1 + correlatiecoefficient
  - $\rho_{xx'} = \frac{2\rho_{yy'}}{1 + \rho_{yy'}}$
- Bij essentieel tau-equivalentie: schatting van betrouwbaarheid op basis van varianties: 2 x (totaalvariantie in kwadraat - (som variantie van beide helften in kwadraat)) / totaalvariantie in kwadraat
  - $\rho_{xx'} = \frac{2[\sigma^2_x - (\sigma^2_{y_1} + \sigma^2_{y2})]}{\sigma^2_x}$
- Nadeel verhogen betrouwbaarheid door verlengen test: test wordt erg (te) lang
- Gestandaardiseerde alpha (alleen bij paralleliteit): gebaseerd op gemiddelde inter-itemcorrelatiecoefficient
  - $\alpha = \frac{k\bar{r}}{1 + (k - 1)\bar{r}}$
- Betrouwbaarheid score(s) (op basis van gemiddelde inter-itemcorrelatie of cronbach's alpha) = proportie ware scorevariantie
- Bij parallele items: geschatte itembetrouwbaarheid = gemiddelde inter-itemcorrelatie = proportie ware itemscorevariantie; proportie error itemscorevariantie = 1 - proportie ware itemscorevariantie
- Bij parallele items: betrouwbaarheid schaal gebaseerd op gemiddelde inter-itemcorrelatie (N = aantal items binnen schaal)
  - $\alpha = \frac{N * \bar{r}}{1 + (N - 1) * \bar{r}}$
- Standaardmeetfout
  - Gebaseerd op betrouwbaarheid van de scores (alpha of gem. inter-itemcorr) en de standaardafwijking van de testscores
  - Schatting van de meetfouten
  - Om de betrouwbaarheidsinterval rondom de ware scores te kunnen bepalen
    - $s_e = s_x(\sqrt{1 - r_{xx'}})$
  - Met 95% zekerheid zal de ware score van van iemand die een geobserveerde score van 100 heeft op de test liggen tussen 100 +/- 1.96 maal de standaardmeetfout

erdeling van testscores

- De scores op de items van deze schaal hangen niet met elkaar samen (de gemiddelde inter-itemcorrelatie is .00): normaal verdeling
- Perfect onderscheid tussen lage en hoge scoorders: U-vormige verdeling
- Een toets met erg gemakkelijke en erg moeilijke vragen, onderscheid/differentieert niet goed tussen personen met een score rond het gemiddelde: leptocurvische verdeling
- Scores op een toets met onbehandelde stof: positief/rechtsscheef (onbehandelde items zullen erg moeilijk zijn, lage p-waarde hebben)
- Scores met een plafond-effect: negatief/linksscheef verdeling
- Toets met erg moeilijke vragen: positief/rechtsscheef verdeling
- Alle items hebben een gecorrigeerde p-waarde van .50 (goed onderscheidend vermogen, afhankelijk van de hoogte van het discriminerend vermogen): platycurvisch, normaal of U-vormige verdeling,
- De ideale verdeling voor selectie-doeleinden: U-vormige verdeling
- Verdeling waaruit systematische verschillen tussen twee groepen blijken: bimodale verdeling
- Van een plafond- of bodemeffect is sprake als de scoreverdeling scheef is, terwijl de verdeling van de te meten eigenschap normaal is. Vaak is een plafond- of bodem-effect het resultaat van een te gemakkelijke respectievelijk te moeilijke toets.