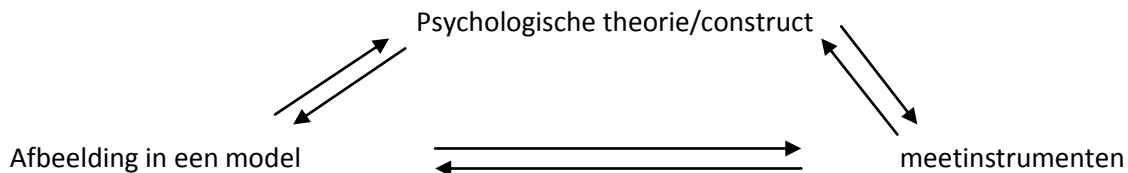


Samenvatting psychodiagnostiek 1 :

1. Wat is psychodiagnostiek? P 110-115 , 119-120 , 134-135

Psychodiagnostiek => als compositie met 3 componenten

- Psychologische theorievorming en constructen over cognitie, intelligentie, emotie, motivatie en gedrag => gedragingen
- Het afbeelden van deze constructen in een model
- De psychologische meetinstrumenten



Verschillende sequenties:

Theorie → model → meetinstrument

Meetinstrument → theorie → model

Model → theorie → meetinstrument

Definities

Beslissen

< Grieks (onderscheiding): onderscheidings-, beslissingsvermogen “een systematisch beslissingsproces” (De Bruyn, 1988)

Vaststellen van individuele verschillen

“een onderzoekstaak om individuele verschillen tussen mensen vast te stellen en om deze kennis toe te passen op het enkele individu” (De Zeeuw, 1983). Het vaststellen van individuele verschillen gebeurt adhv meetinstrumenten

Verzamelen van info over persoon en zijn/haar omgeving

“op wetenschappelijk verantwoorde wijze verzamelen van info omtrent persoon en zijn omgeving met het oog op het nemen van beslissingen” (Van Aarle, 1990)

Hulp bieden

“psychodiagnostisch proces is het proces van hulp bieden aan personen met hun vragen en problemen” (Walsh & Betz, 2001). Diagnostiek mondt uit in behandeling/interventie

Beschrijven van ontwikkelingsniveaus (zwakke én sterke kanten) van cliënt

“developmental assessment op diverse ontw.gebieden zoals cognitie, motoriek, socio-emotionele vaardigheden, ...” (Johnson & Sheeber, 1999)

Theorie

Impliciete theorievorming

Het alledaagse beschrijven en verklaren van ons gedrag.

Impliciete : opvatting van leken over oorzaken en samenhangen tussen gedragingen

Expliciet : theorievorming uit handboeken en vaktijdschriften

3 centrale oriëntaties :

- Individuele verschiloriëntatie **kijken naar kenmerken waarop mensen verschillen van elkaar**
- Ontwikkelingsoriëntatie **sterkere (Piaget) en zwakkere (Erikson)ontwikkelingstheorievorming**
- Contextoriëntatie **efficiënte oorzaken worden aangewezen voor gedragingen => (omgevingsstimuli, sociale situaties => de context)**

Oriëntaties gebaseerd op drie fundamentele menselijke condities

- Mensen kunnen onderling vergeleken worden op **kenmerken** en deze variëren
- De mens is een tijdelijk wezen, dat zich *niet* niet kan **ontwikkelen**
- De mens acteert en reageert onontkoombaar in en op een **omgeving /context**

⇒ we noemen deze vormen van combineren 'hybriden', omdat de gezichtspunten niet tot elkaar te herleiden zijn.

Model

Afbeelding van centrale kenmerken van theorieën, theoretische begrippen (constructen) in een model. Operationaliseren van construct

- meten op een schaal
 - *vb. subtest Cijferreeksen WISC-III -> 'geheugen'*
 - *vaak: breed construct (bv. intelligentie) afgebeeld in dimensies*
- onderbrengen in een categorie die te onderscheiden is van andere
 - *vb. DSM-IV 'depressie'*
- modellen kunnen afgeleid zijn uit de drie oriëntaties
 - Individuele-verschillenoriëntatie
 - *vb. model van Carroll over intelligentie (p. 320)*
 - *vaak gemeten met tests en schalen (=> verbinding met testtheorie is gemakkelijk gelegd)*
 - *als resultaat doorgaans typen of dimensies om mensen te categoriseren/beschrijven*
 - Ontwikkelingsoriëntatie
 - *Als resultaat een afbeelding van de ontwikkeling van gedrag (vb. IQ of neuroticisme van persoon of groep doorheen de tijd)*
 - *Ontwikkelingsstadia kunnen kwalitatief afgebeeld worden of ordinaal met ontwikkelingsstappen*
 - *Staat het verst van de testleer af*

- Contextoriëntatie
 - Causale relatie wordt gelegd tussen onafhankelijke en afhankelijke variabele
 - Niet altijd gebruik van betrouwbare en valide instrumenten
 - Variabelen worden gemeten ,meestal zonder te toetsen, of de afhankelijke variabele unidimensionaal is
 - Vallen niet terug op testleer en moeten dat ook niet
- kwantitatieve modellen ontstaan uit de testleer en de statistiek
- Testleer of psychometrie
 - Een verzameling van modellen om bepaalde eigenschappen van items –en testscores af te beelden en te toetsen (kunnen gebruikt worden voor de verschillende oriëntaties!)
 - Nadruk ligt op unidimensionale intervalschalen en modellen

Enkele belangrijke modellen:

- *Klassieke testtheorie (KTT) (cfr later)*
- *Moderne testtheorie of Item Respons theorie (IRT) (cfr psychodiagnostiek-II)*

Instrument

- Psychologische tests & meetinstrumenten
 - Afbeelding van eigenschappen, kenmerken of gedragingen van personen op dimensies of plaatsen in een categorieën systeem
 - Vbn: interview, enquête, vragenlijst, test, observatieschaal, ...
- Definitie psychologische test volgens Drenth & Sijtsma (1990)
 - Een test is een systematische classificatie –of meetprocedure, waarbij het mogelijk wordt om een uitspraak te doen over één of meer empirisch -theoretisch gefundeerde eigenschappen van de onderzochte of over specifiek niet-testgedrag, door uit te gaan van een objectieve verwerking van reacties van hem/haar, in vergelijking tot die van andere, op een aantal gestandaardiseerde, zorgvuldig gekozen stimuli.
- Soorten tests
 - Persoonlijkheid (1) en prestaties of ‘typical’ en ‘maximum performancetests’ (2)
 - (1) tests voor gedragswijzen, geen goede of slechte antwoorden, maar wat de persoon er zelf van vindt
 - (2) antwoord is indicatie voor intelligentie, vaardigheid of (school)prestatie
 - Papier -en potloodtests vs. speciaal materiaal
 -
 - Complex schema : 3 superordinate categorieën (met subordinate categorieën)
 - PersoonSituatieTests
 - Situaties , zoals intramuraal, arbeid, opleiding; lichaam en ander
 - Vaardigheden en attitudes

- Gedragscategorieën, zoals cognitie, aandacht, perceptie, emoties, behoeften, waarden en motivatie
- PersoonlijkheidsTests
 - Vaardigheden en attitudes
 - Gedragcategorieën, zoals bij PersoonSituatieTests
- SituatieTests
 - Indeling in typen situaties, zoals sociale, taakgerichte, therapeutische, arbeids-, opleidings- en gezinssituaties
- Taak, opdracht of vraag die persoon voorgelegd krijgt
- Typen testconstructie
 - Rationele strategie
 - Gaat uit van een conceptuele analyse van een begrip en formuleert op grond daarvan opdrachten, vragen of items
 - Een strategie om het psychologisch concept te analyseren is facet-analyse
 - Object-facet
 - ✓ Als elementen school, beroeps carrière, persoonlijke ontwikkeling, sociale relaties en vrije tijd
 - Extensie-object
 - ✓ De korte termijn en de lange termijn
 - Empirische strategie
 - Houdt in dat langs empirische weg bepaald wordt of een reeks items discrimineert tussen categorieën personen, die we daadwerkelijk willen onderscheiden
- Testafname
 - Tests en items zijn doorgaans in handleiding precies voorgeschreven => test is gestandaardiseerd om een vergelijking tussen personen mogelijk te maken
 - Bij 'Testing the Limits' en 'Leerpotentieel-tests' vergelijken we de testprestaties onder standaardcondities waarbij de proefpersonen geholpen worden bij het maken van tests
 - De scoring van de taken, vragen of items is voorgeschreven de interpretatie van scores geschiedt doorgaans aan de hand van normtabellen bij 'norm referenced' tests
 - Een nieuwe ontwikkeling is het afnemen van test met behulp van de computer
 - Test hoeft niet helemaal eer afgenomen te worden om betrouwbare schatting te krijgen
 - Items moeten wel de latente waarde nauwkeurig schatten en moeten geordend zijn naar moeilijkheidsgraad en discriminerend vermogen

Doel p 27-29 en 139

Diagnostiek is geen zelfstandige subdiscipline van de psychologie, maar activiteit ervan is van belang in elke subdiscipline. Er wordt gebruik gemaakt van theoretisch kennisbestand, methodologie en data-analysemethoden van alle subdisciplines. Wordt gebruikt omwille van 4 doelen (+ 2 extra doelen)

- Beschrijving

- Betrouwbare en geldige beschrijving van gedrag van cliënt, zijn sociale omgeving en diens ontwikkelingen
- Er worden gedrags-, emotionele- en cognitieve problemen onderzocht
- Onderzoek naar condities, verbanden en verloop
- Diagnose
 - Activiteit van experts om een persoon in een bepaald categorieënsysteem onder te brengen.
 - Inventariseren en ordenen van (niet) functionele clusters van gedragingen, cognities en emoties
 - Voorbeeld categorieënsysteem binnen psychologie/ psychiatrie: Diagnostic and Statistical Manual of mental Disorders(DSM-IV; APA, 1994). Berust voor een deel op wetenschappelijke kennis, maar vooral op intercollegiale overeenstemming over de indeling van allerlei stoornissen.
 - Nuttig bij differentiaaldiagnoses waarbij onderscheidt wordt gemaakt tussen twee categorieën met gedeeltelijke gelijke symptomen.
 - Nuttig bij het vaststellen van co-morbiditeit (het voorkomen van 2 of meer stoornissen bij een persoon).
 - APA: Amerikaanse psychiatrische beroepsvereniging
- Voorspelling (predictie)
 - Voorspelling kans van optreden van (dis)functionele gedragingen, cognities en emoties in toekomst (met/ zonder behandeling)
 - Kansuitspraak doen over gerelateerde gedragingen(vb. wat is de kans op zelfmoord?)
 - Nut van kansuitspraak
 - Praktijk
 - Wetenschappelijk onderzoek: aantonen van theoretisch veronderstelde samenhang tussen gedragingen via empirisch onderzoek
- Verklaring
 - Diagnostisch onderzoek naar de oorzaken van een gedrag(sprobleem)
 - Geeft antwoord op het ‘waarom’ van het voorkomen van de (dis)functionele gedragingen
 - Deze vier doelen niet altijd strikt te scheiden
 - Ze zijn wel te onderscheiden en dekken vrijwel alle activiteiten van het diagnostisch onderzoek bij cliënten
 - Niet in alle subdisciplines van de psychologie wordt elk van de vier doelen even belangrijk gevonden
 - Hypothesen Toetsend Model (htm)
 - Structureert het diagnostisch zoek-, oordeels- en toetsingsproces
 - Voorlopige omschrijving van psychologische diagnostiek
 - “een onderzoeks –en oordeelsproces dat als doel heeft te komen tot een beschrijving, diagnose, voorspelling of verklaring van klachten van een cliënt of een antwoord te geven op zijn vragen”
- Indicatie
 - Keuze tussen wel of niet behandelen
 - Keuze uit verschillende mogelijke behandelingen

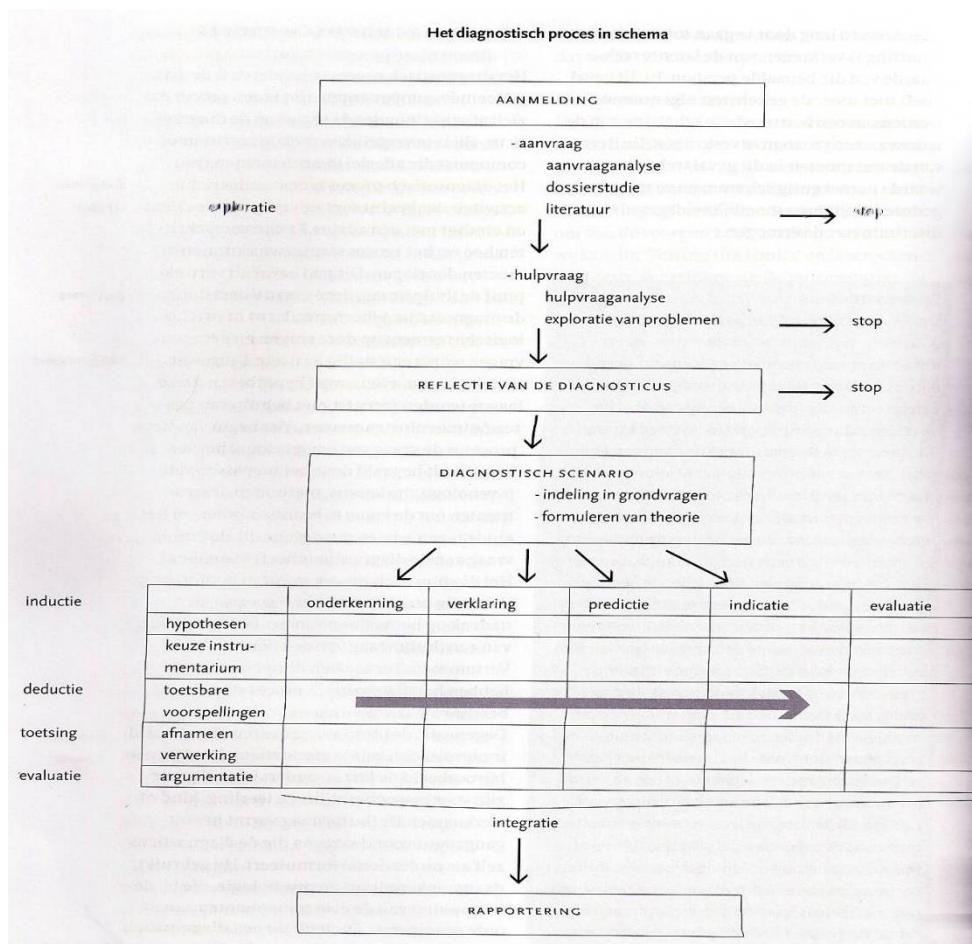
- Hoe effectief zijn ze?
- Wat is de kans dat cliënt ze kan uitvoeren of wil ondergaan?
- Evaluatie
 - Het bepalen van het succes of falen van de behandeling
 - Effect van interventie nagaan (wat hielp wel en niet? Vooruitgang?)

Het diagnostisch scenario in 5 stappen

- Omzetten van theorie/construct in concrete, toetsbare hypothesen
- Kiezen van tests, instrumenten, procedures om de hypothesen te toetsen
- Kiezen van criteria om de hypothesen (antwoorden) te accepteren of niet
- Het afnemen van tests, instrumenten en/of toepassen van procedures
- Concluderen of de hypothesen niet of wel verworpen worden

Het diagnostisch proces p 137 - 139

Het **diagnostisch proces** (steunt op 3 componenten : theorie, model, meetinstrument) is een doelgerichte activiteit dat begint met de vraag van de cliënt en eindigt met een advies. We kunnen het proces stapsgewijs doorlopen. Dit pad bevat als vertrekpunt de vraag, deze vraag wordt door de diagnosticus (= regisseur) geherformuleerd in psychologische termen; op deze vragen en deelvragen wordt een stellig antwoord gegeven in de vorm van evenzoveel hypothesen (deze worden getoetst met behulp van (meet)instrumenten). Hieruit volgt het **diagnostisch proces** waarin de vragen van de cliënt worden beschreven en hypothesen worden geformuleerd. = 5 stappen.



2. Geschiedenis van psychodiagnostiek

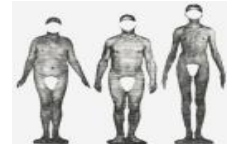
Inleiding p. 29 – 32

- Al zeer lang sprake van tests en beoordelen
 - bv. wie is de beste, snelste, slimste, handigste,...?
 - Selectiesystemen in China, meer dan 1000 jaar
 - Ontstaan vanuit praktisch probleem
 - Binet-Simon test (1905)
 - Eerste wetenschappelijke test
- Ontstaan eind 19^{de} eeuw (officieel 1870-1880)
- Voorlopers van de testbeweging leidde tot procedures voor bepalen van individuele verschillen in intelligentie en persoonlijkheid => mensen selecteren
- Het testen is inmiddels al uitgegroeid tot de professionele activiteit bij uitstek van psychologen en pedagogen in de 21^{ste} eeuw.
- De geschiedenis van de diagnostiek wordt min of meer gelijkgesteld aan de opkomst en de ontwikkeling van het testen voor selectie en plaatsing.
- De geschiedenis van psychodiagnostiek is ook de geschiedenis van testen, testtheorie, psychologische theorievorming en het diagnostisch proces
- Geschiedenis diagnostiek vooral over opkomst, verspreiding en gebruik van psychologische tests en meetinstrumenten; deze is na een Europese aanzet in de VS verder ontwikkeld
- Geschiedenis bestuderen is nuttig omdat daardoor kan worden voorkomen dat theorieën en methoden die niet succesvol waren, opnieuw worden uitgevonden en toegepast

Historische voorlopers p 32 – 35

- Opbouw van het ambtenarenapparaat in China
 - Selectieprocedure ambtenaren voor de keizers (de mandarijnen)
 - Schriftelijke examens
 - Getrapt examensysteem
 - In 1906 werd het systeem afgeschaft
- Selectie in het Oude Testament
 - Selectie van de Gideonsbende
 - 'Ratrace' (met veel afvallers =< vormde mede de klasse van de 'literati')
- Beoordelen en selecteren in het antieke Griekenland
 - Gedragsverschillen → Typologieën
- **Hippocrates:** temperamenttypes op grond van lichaamssap
 - Sanguinisch type (bloed)
 - Cholerisch type (gal => opvliegend)
 - Flegmatisch type (slijm => lusteloos)
 - Melancholisch type (zwarte gal => depressief)
- **Plato:** persoonstypes op grond van de plaats van het gemoed in het lichaam
 - In het hoofd (dichters, filosofen,...)
 - In de borst (soldaten)
 - In het onderlijf (gewone man en vrouw)
- Hedendaagse 'ratraces'
 - Overheid
 - Politie

- Bedrijfsleven
- Diagnostiek avant la lettre
 - 13^{de} eeuw: schoolprestaties via mondelinge examens op universiteit
 - 16^{de} eeuw: selectie van 'hoog begaafden' (Huarte : Spaanse arts)
 - 18^{de} & 19^{de} eeuw: persoonlijkheid op basis van uiterlijk
 - Typologie op grond van constitutie
 - Gall: intelligentie op basis van schedelomtrek (schedelknobbeltheorie)
- Typologie op grond van constitutie
 - Kretschmer
 - Picnisch (houden van het leven) (1)
 - Atletisch (breedheid, luidruchtig) (2)
 - Leptosoom (lengte, meer intelligent, niet uitbundig)(3)
 - Sheldon
 - Endomorf (1)
 - Mesomorf (2)
 - Ectomorf (3)



Ontwikkeling in psychiatrie en experimentele psychologie p 35 – 39

Bijdrage psychiatrie

- Inzichten van de psychiatrie (medische denkmodellen) in de aard van krankzinnigheid en zwakzinnigheid (Frankrijk)
- Taxonomisatie - Classificatie
- Begin 19de eeuw: onderscheid zwakzinnigheid en psychiatrische stoornissen (krankzinnigheid)
 - Pinel: interesse
 - Esquirol: diagnose
 - Sequin: behandeling
- Eind 19de eeuw: ontwikkeling testmethoden
 - intelligentie, vermoeidheid, schizofrenie & persoonlijkheid
- Begin 20ste eeuw: classificatiesysteem voor psychiatrische stoornissen
 - Kraepelin → Taxonomisatie
- Pinel (1793) => interesse
 - Directeur van een hospitaal
 - Dankzij hem een andere aanpak in instellingen
 - Patiënten onderverdelen
 -
 - Bicêtre : mannen (4000)
 - Nieuwe ideeën in het leven geroepen om patiënten te benaderen
 - Salpêtrière
 - Populatie van La Salpêtrière in 1701 (instelling Parijs): 4646 vrouwen
 - 1894 kinderen jonger dan 15 jaar
 - 329 meisjes jonger dan 16 jaar (kreupel, kolerig, enz)
 - 594 oudere blinden en verlamden
 - 262 ouderen hoger dan 70 j
 - 380 veroordeelden, zedelozen en prostituees
 - 465 schooiers en vagebonden
 - 330 kindsvrouwen van hoge leeftijd
 - 300 gewelddadige gekken
 - 92 epileptici van verschillende leeftijden

- Esquirol (1838): “des maladies mentales”
 - Kwantitatieve en kwalitatieve differentiatie
 - Irreversibel
 - Zwakzinnigheid
 - permanent, irreversibel
 - ongeneeslijk
 - Psychiatrische stoornis
 - kan ontstaan op latere leeftijd
 - kan verbeteren
 - Esquirol schetsen
 - Idiote
 - Dementie
 - Manie
 - melancholie
- Seguin (1837): student Esquirol
 - Reversibel
 - Trainingsmethode behandeling
- Kraepelin (1883)
 - “Compendium der psychiatrie” eerste classificatie in de psychiatrie: o.a. manisch-depressie & dementia-precox, Alzheimer... => symptomen gaan ordenen
 - Bestaat uit verschillende symptomen
 - Oorzaken proberen te geven van elke stoornis
 - Fysiologische oorzaak
 - Meting noodzakelijk
- Taxonomisatie psychiatrie
- De essentie
 - mentale retardatie ≠ psychiatrische beelden
 - ... is “een ziekte” (fysiologische ...)
 - ... kan behandeld worden
 - Behandeling door training
 - Meting is gewenst voor differentiatie tussen de beelden



Bijdrage experimentele psychologie

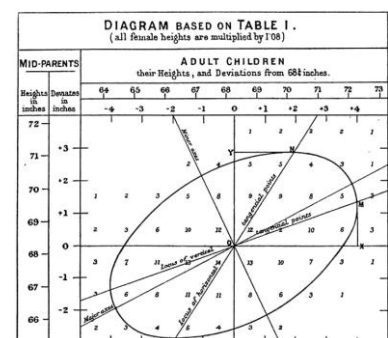
- Meetnauwkeurigheid (Duitsland)
 - Vaststellen en eten van algemeen geldende wetmatigheden in de waarneming
 - Mensen in condities onderbrengen
 - Vanuit ander uitgangspunt
- Individuele verschillen tussen mensen: waarneming
 - Wundt: denksnelheid
 - Ebbinghaus: geheugenspanne/capaciteit
 - Ging bij verschillende mensen na : betekenloze lettergrepen aanleren
- Meetinstrumenten
 - Bv. Tachistoscoop (1903) (Toestel waarmee visuele stimuli kunnen worden aangeboden voor een zeer korte tijdsspanne (msec))
- Manipulatie van onderzoekscondities
 - Nauwkeurige schattingen van fouten (storende factoren)
 - Standaardisatie-eis voor tests
- Toetsen van hypothesen
 - Diagnostisch proces: Hypothesen toetsend model

Enkele belangrijke historische personen p. 39-47

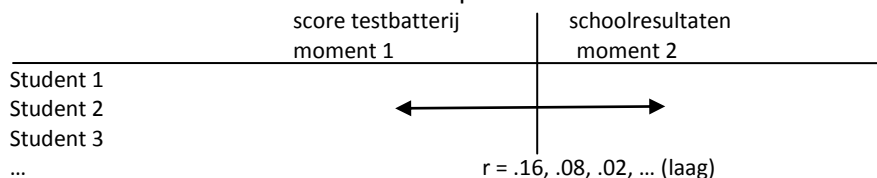
- Inleiding historische personen
 - Vooral klemtoon op individuele verschillen in persoonlijkheid en intelligentie (dus ontwikkelingsoriëntatie en contextoriëntatie krijgen minder aandacht)
 - Belang van maatschappelijke veranderingen voor ontwikkelingen in psychodiagnostiek
- Galton (1822 – 1911) GB
 - Brede interesse (vb. statistiek, erfelijkheid, efficiëntie van gebed)
 - Filosofische overtuiging: Engels empirisme
 - Kennen is alleen mogelijk via zintuigen; dus wie de beste zintuiglijkheid bezit functioneert ook het best mentaal
 - 'Meten is weten'
 - Erfelijkheid van individuele verschillen
 - Geïnspireerd door evolutieleer van Darwin: belang van erfelijkheid van lichamelijke eigenschappen
 - Galton wil aantonen dat psychische eigenschappen (vb. genialiteit) ook erfelijk zijn en dezelfde wetten volgen als lichamelijke eigenschappen
 - Cfr. Eugenetica ("goed geboren")
 - Metingen zouden als basis kunnen dienen om de "besten" te selecteren
 - Psychofysische metingen voor het mentale vermogen
 - Belangrijk individueel verschil: het mentale vermogen
 - Metingen van het mentale vermogen gebeurde via psychofysische metingen van elementaire functies (vb. lichaamslengte, gewicht, lengte van hoofd, spierkracht, reactiesnelheid, capaciteit longen, hoogst hoorbare toon, ...)
 - Waarom? Elementaire functies zijn eenvoudig en exact te meten, zouden weinig beïnvloed worden door milieu en opvoeding, zijn representatief voor het mentale functioneren (cfr. Engels empirisme)
 - -> ontwikkeling objectieve tests
 - -> gebruik van gestandaardiseerde procedures
 - Psychofysische metingen voor het mentale vermogen
 - Protocol van Galton
 - Wiskundige technieken
 - Maakt gebruik van 'co-relations' en regressie-analyses om bv. verband uit te drukken tussen kenmerken van ouders en kenmerken van hun kinderen
 - Samengevat:
 - individuele verschillen (cfr. genetica)
 - -> systematisering onderzoekstechnieken (cfr. experimentele psychologie)
 - -> wiskundige technieken
 - bv. scores uitdrukken in termen van afwijkingen tov. het gemiddelde (1e stap naar normatief denken nl. prestatie beoordelen door vergelijking met anderen)

F. Galton

Individual			Father and mother		Date of birth		Eye color		Sex		Height		Color		Occupation	
Self	Father	Mother	Inches and inches		Year and day		Light blue		M		M		Normal		Peculiar	
Day	Month	Year	Inches	Tenths	Inches	Tenths	Inches	Tenths	Inches	Tenths	Inches	Tenths	Inches	Tenths	Inches	Tenths
24	2	44	7	1	6	1/2	68	7	7	1	70	7	1	70	7	1
Height			Length of arm		Length of middle finger		Hearing		Hearing		Reaction time		Reaction time		Reaction time	
Inches			Inches		Inches		Inches		Inches		Inches		Inches		Inches	



- Net als Galton geïnteresseerd in individuele verschillen
- Ontwikkeling van tests voor meting individuele verschillen
 - Link tussen Engelse traditie (meten) en Amerikaanse behoefte aan tests (konden gebruikt om immigranten te selecteren in de VS; VS is eind 19e – begin 20e eeuw een immigratieland)
 - 'Mental tests' weerspiegelen werkwijze van Galton (psychofysische metingen): handgrip (dynamometer), discrimineren gewichten, oordeel over verstreken tijd, druk nodig om pijn te voelen
 - Overtuiging: mentale en lichamelijke energie kunnen niet van elkaar onderscheiden worden
 - Ontwikkeling van psychofysiologische laboratoria en testbatterij. Jaarlijkse testafname bij schoolkinderen en studenten met oog op bepalen intelligentie en voorspellen schoolprestaties bv. meten van spierkracht, snelheid bewegingen, pijn gevoeligheid...
- Onderzoek naar validiteit
 - (deel 5 'geldigheid scores') => meten we wat we willen meten?
 - Wissler (student van Cattell) onderzoekt
 - of scores op testbatterij schoolresultaten van studenten kunnen voorspellen

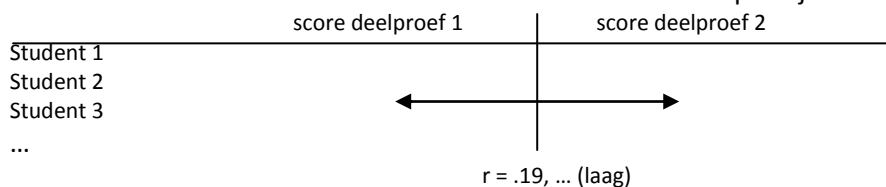


= criteriumvaliditeit

verband testscores – externe variabele

correctheid van voorspellingen over te verwachten gedrag in andere situaties dan de testsituatie op basis van de testresultaten

- in hoeverre de verschillende proefjes onderling samenhangen



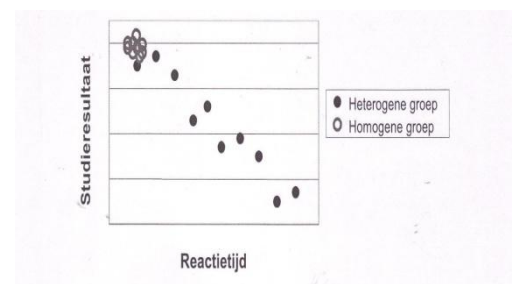
= constructvaliditeit

interne structuur van de testbatterij

zijn scores op deelopproeven indicatoren voor hetzelfde construct (vb. 'mentaal vermogen')?

-> omwille van lage r: psychologen stoppen met gebruik van psychofysiologische metingen voor intelligentie

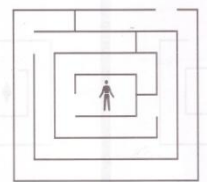
- Mogelijke verklaring voor lage correlaties:
 - Homogene steekproef ('restriction of range')
 - Variatie in testcores en in studieresultaten was beperkt
 - Recentere correlaties zijn groter dan deze gevonden door Wissler



- 3 grote invloedslijnen:
 - Psychiatrie:
 - Classificatie, diagnose en niveau van geestelijke afwijkingen
 - Experimentele psychologie:
 - Standaardisatie onderzoeksprocedures

- Genetica:
 - Individuele verschillen
 - tot nu toe gericht op 'elementaire functies'
 - hierin komt verandering vanaf Binet
- Binet (1857 – 1911) F
 - Stichter van tijdschrift L'Année Psychologique
 - Veelzijdig en zeer actieve persoon (zie verschillende onderwerpen in zijn boeken en artikels)
- Hogere mentale processen
 - Bekend artikel "La psychologie individuelle" (1895) van Binet en Henri:
 - afwijzing van sensorische tests voor onderzoek naar intelligentie
 - individuele verschillen in intelligentie bepalen adhv. hogere mentale processen (bv. geheugen, begrip, problemen oplossen, verbeeldingskracht)
 - voorgestelde tests: geheugentests, verbeelding met behulp van inktvlekken, aandachtsonderzoek, verschil tussen een paar begrippen opschrijven,...
- Binet-Simon test
 - Verzoek van Ministerie van Onderwijszaken (schoolplicht was recentelijk ingevoerd in Frankrijk + groot aantal mislukkingen): test ontwikkelen voor selectieve maatregelen voor kinderen die wel/niet het gewone onderwijs konden volgen
 - 30 taken om intelligentie na te gaan (zie box 1.7 p. 43):
 - Vb. 7) Het aanwijzen van hoofd, neus, oor, enz. op de vraag: "Waar is je...?".
 - ~ WPPSI-r subtest Informatie
 - "Wijs eens aan waar je neus is"; "hoe noem je dit (duim)"
 - 11) Herhaalt drie voorgelezen paren getallen
 - ~ WISC-III subtest Cijferreeksen
 - "Ik ga enkele cijfers opnoemen; luister goed want als ik klaar ben moet je ze nazeggen"
 - 14) Herhaalt een zin van 15 woorden
 - ~ WPPSI-r subtest Zinnen
 - "Ik ga iets zeggen en jij moet het nazeggen. Luister goed en zeg hetzelfde als wat ik zeg. Mama werkt hard"
 - 15) Definieert eenvoudige woorden en begrippen door middel van hun functie
 - ~ WISC-III subtest Woordkennis
 - "Wat is een huis?"
 - 20) Vertelt hoe twee bekende voorwerpen overeenstemmen, bv. vlinder en vlieg
 - ~ WISC-III subtest Overeenkomsten
 - "Wat is de overeenkomst tussen melk en water?"
 - Eerste versie verschijnt in 1905
 - Kenmerken:
 - niet bedoeld om alle aspecten van intelligentie te meten waar wel als maat voor algemene mentale ontw. van kinderen
 - korte en praktische test
 - aantrekkelijk zodat kinderen er hun aandacht bij konden houden
 - intelligentie wordt beschouwd als oordeelsvermogen en niet als 'lagere' vaardigheden

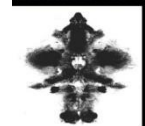
- nog geen gestandaardiseerde instructie (doel = classificatie en niet het exact meten van begaafdheden)
 - Er volgen een aantal herwerkingen van de eerste versie
- 'Mentaal niveau' of 'Mentale leeftijd' (1908)
 - Binet-Simon test wordt bij 300 kinderen afgenomen en per leeftijdsgroep wordt nagegaan hoeveel kinderen een item goed beantwoorden
 - Items worden geordend volgens leeftijd
 - Bepaling van mentaal niveau of mentale leeftijd:
 - bepaling van het basale mentale niveau (mentale niveau waarvoor hoogstens één proef niet was gelukt)
 - voor elke 5 gelukte proeven boven basale niveau wordt één extra jaar toegekend
- Latere ontwikkelingen van de Binet-Simon test
 - Stern (1911) $IQ = \frac{\text{mentale leeftijd}}{\text{chronologische leeftijd}}$
 - Vertalingen
 - Terman (1916): Stanford-Binet (Amerikaanse versie)
 - Standaardinstructies
 - Normen
 - nieuwe scoringswijze $IQ = \frac{\text{mentale leeftijd}}{\text{chronologische leeftijd}} \times 100$
- Goddard (1866 – 1957) VS
 - Verspreiding van tests en testmisbruik
 - Belangrijke rol in verspreiding van tests binnen VS
 - Indeling van prestaties adhv. van beledigende termen bv. 'moron'
 - Onderzoek naar prevalentie van mentale achterstand
 - Verspreiding van tests en testmisbruik
 - Organiseerde testing van immigranten op Ellis Island in VS
 - Problemen:
 - oorspronkelijk Franse test -> vertaald in Engels -> vertaald in Russisch, Italiaans,....
 - immigranten waren zwaar vermoeid bij afname van test
 - scores werden geïnterpreteerd met Franse normen
 - Conclusie Goddard: 83% Joodse, 80% Hongaarse, 79% Italiaanse, 87% Russische immigranten waren 'zwakzinnig' !
 - Verspreiding van tests en testmisbruik
 - Ontwikkeling nieuwe non-verbale tests voor immigranten die nu nog terug te vinden zijn in intelligentietests
- Verdere ontwikkelingen
 - Groepstests
 - Door WO I ontstond noodzaak om grote groepen personen voor uiteenlopende functies te selecteren (bv. Yerkes)
 - Amerika:
 - Army alpha test (vb. rekenopgaven, synoniemen, informatie)
 - Army beta test (non-verbale opdrachten)
 - Probleem: vele rekruten scoorden 0 (niet omwille van mentale achterstand!)
 - Brigham:
 - Testresultaten worden gebruikt (misbruikt) om verschillende bevolkingsgroepen te vergelijken
 - "Rasvermenging zal onherroepelijk leiden tot aftakeling van Amerikaanse intelligentie"



- Later herroept Brigham zijn uitspraken
- Tests voor specifieke vaardigheden
 - Na WOII heeft men nood aan selecteren van personen voor gespecialiseerde taken
- Trend vanaf WO II tot heden
 - Toename in diversificatie van tests en testgebruik
 - Sterkere en kritische bezinning over methodische grondslagen van testgebruik (met meer toepassing van statistische analyses)

Ontwikkeling van persoonlijkheidstests p. 49-50

- Persoonlijkheidsvragenlijsten
 - Woodworth
 - legerrekruten WO I in VS selecteren (vatbaarheid 'psychoneurose')
 - Woodworth Personal Data Sheet (1917)
 - Schaal op basis van 116 symptoombeschrijvingen
 - Vb. Komen gedachten bij jou op waardoor je niet kan slapen?
 - Beschouwde men jou vroeger als een stout kind?
 - Heb je last van een gevoel dat de dingen niet echt zijn?
 - ☐ ja
 ☐ neen
 - Som van antwoorden geeft mate van aanpassing weer
 - Basis voor latere vragenlijsten
 - MMPI (1943)
 - Minnesota Multiphasic Personality Inventory
 - Koppeling van onaangepastheid aan psychiatrische categorieën
 - Verschillende klinische schalen: bv.
 - Depressie ("ik heb perioden gehad dat ik niets deed omdat ik niet op gang kon komen")
 - Schizofrenie ("één of meerdere keren in mijn leven had ik het gevoel dat iemand me dingen liet doen door me te hypnotiseren").
 - Verschillende validiteitschalen: bv.
 - Leugenschaal ("Ik zou een spel liever winnen dan verliezen")
- ☐ akkoord
 ☐ niet akkoord
- Projectieve technieken
 - Rorschach (1921) publiceert over ervaringen met inktvlekken en legt verband tussen het duiden van inktvlekken en persoonlijkheid
 - 10 symmetrische vlekken (keur-zwart/wit)
 - "Wat stelt dit voor?"
 - Thematic Apperception Test (TAT)
 - Voor het eerst beschreven door Morgan & Murray (1935)
 - 'verhalentest'
 - Tekeningen van personen in onduidelijke situaties
 - "wat doen deze mensen, wat is er vooraf gebeurd, wat zal er gebeuren?"
 - Aanvullen van zinnen
 - Ik hou van _____
 - De gelukkigste tijd _____
 - Mijn vader _____
 - Mijn grootste angst _____
 - Ik kan niet _____



- Stiekem (in het geheim) _____
- Toen ik nog kind was _____
- Ik wil _____
- **voor- en tegenstanders van projectieve technieken**
 - brede band techniek, diepere lagen persoonlijkheid betrouwbaarheid, validiteit
- **Waarom is studie van geschiedenis van psychodiagnostiek zinvol?**
 - Inzien dat diagnostiek ook misbruikt kan worden (vb.)
 - Hier kritisch over blijven: tests ontwikkelen die voldoen aan wetenschappelijke eisen is nog geen garantie dat deze instrumenten ethisch verantwoord gebruikt worden!
 - Huidige testinhouden en testvragen begrijpen vanuit hun historische oorsprong
 - Inzien dat diagnostiek aansloot en ook nu nog aansluit bij maatschappelijke vragen
- **Voorbeeldvragen**
 - De empiristische opvatting dat kennis vooral door de zintuigen verworven wordt treffen we aan bij
 - a. **Galton**
 - b. Kraepelin
 - c. Binet
 - d. Wundt
 - De beste 'verkoper' van tests was
 - a. Galton
 - b. Wundt
 - c. **Goddard**
 - d. Mc Keen Catell

3. Meten

Waarom?

- Vb1. Jobselectie
- Vb2. Hoe komt het dat Katrien zo moeilijk kan volgen in de klas?
- Vb3. Hoe kunnen we de aandachtsproblemen van Jan best aanpakken?
- Scores interpreteren

Operationaliseren

- METEN = toekennen van cijfers of categorielabels
 - We willen bijvoorbeeld meten:
 - hoe groot iemand is.
 - wat iemands lichaamstemperatuur is.
 - hoe intelligent iemand is.
 - in welke mate iemand dyslexie heeft.
 - Begrippen uit de 4 meetbaar maken
 - Categorielabel of cijfers
 - Verschil tussen de vb'n : 1x2 makkelijk, 1x4 moeilijk
 - => begrippen uit de 4

Empirische constructen (1x2)	Abstracte constructen (1x4)
Directe meting: Observatie -> meetwaarde	Indirecte meting: Observatie -> score -> meetwaarde

Vb. lengte, gewicht, temperatuur	Vb. aangepastheid, intelligentie, neuroticisme
Begrip = zijn operationalisering	Begrip $\neq$ zijn operationalisering
Direct duidelijkheid hoe je het moet meten	Nadenken hoe je het moet meten

- **OPERATIONALISEREN** van een construct
 - = Bepalen wat men precies zal registreren en welke variabelen hiervoor gekozen worden.
 - Vb. een mogelijke operationalisering van het construct intelligentie is de WISC
 - Oefening: Zijn de volgende constructen abstracte of empirische constructen?
 - Sociaal economische status (**a**)
 - Body mass index (**e**)
 - Gemoedstoestand (**a**)
 - Oriëntatievermogen (**a**)
 - Leessnelheid
 - **E = snelheid**
 - **A=leessnelheid nog niet echt duidelijk hoe te meten**

Schalen: antwoordschalen en meetschalen

- **INSCHALEN** = toekennen van cijfers aan objecten of aan waarnemingen betreffende objecten volgens een regel
 - Omcirkel uw antwoord:
 - Mogen er immigranten in uw land leven? **Ja/nee**
 - Mogen er immigranten in uw straat leven? **Ja/nee**
 - Mag een immigrant uw kind huwen? **Ja/nee**
- **ANTWOORDSCHAAL** = de manier waarop je de antwoorden op beweringen verzamelt

Zet een kruisje bij uw antwoord:	Helemaal oneens	Eerder oneens	Noch eens, noch oneens	Eerder eens	Helemaal eens
Mogen er immigranten in uw land leven?					
Mogen er immigranten in uw straat leven?					
Mag een immigrant uw kind huwen?					

- **(MEET)SCHAAL** = de verzameling van mogelijke categorieën of scores van een variabele.
 - De scores voor de verschillende objecten bekomen we via een welbepaalde procedure, een regel.

(Meet)schaal	Antwoordschaal
Resultaat van een procedure	Gebruikt om antwoorden te verzamelen

Elk item in de schaal heeft een waarde	Een item heeft geen waarde
Verwijst naar een set van items (procedure toepassen)	Gebruikt voor een item

- Antwoordschalen
 - Dichotoom
 - Je hebt maar 2 mogelijkheden om te antwoorden => tegenovergesteld aan elkaar
 - kiezen tussen gegeven antwoordmogelijkheden

Vraag 2: Kruis aan of u de genoemde tabletten gebruikt of heeft gebruikt gedurende de laatste twee maanden

A Paracetamol

☐ Ja

☐ Nee

B Aspirine

☐ Ja

☐ Nee

- Semantische responsvorm
 - Kiezen tussen gegeven antwoordmogelijkheden

Ik stel het gezelschap van vreemden op prijs	
Antwoordmogelijkheid	Score
0-Volledig akkoord	5
0-Min of meer akkoord	4
0-Neutraal	3
0-Min of meer niet akkoord	2
0-Totaal niet akkoord	1

- Multiple choice
 - 1 antwoord is juist (verschil met gedwongen keuze vragen)

Vraag 1: Men gooit twee dobbelstenen. **Wat is** de kans dat de ene dobbelsteen een EVEN aantal ogen heeft **en de andere een ONEVEN** aantal ogen?

A. ½ B.1/3 C.1/4 D.1/6 E. 1/9

- Gedwongen keuze vragen

Vraag 1: Ik zou het liefst...

- Het ontstaan van files onderzoeken
- Werken met kinderen
- Mentor zijn van nieuwe werknemers in een bedrijf
- ⇒ Alle antwoorden zijn juist (verschil met multiple choice)
- ⇒

- Grafische responschaal

- Streepje zetten op een lijn waar jij je bevindt als antwoord op de vraag

Vb. Ik geef er de voorkeur aan een leider te zijn.

Altijd ————— Nooit

- Checklist

- Aanduiden wat er allemaal op jou van toepassing is

De laatste twee weken voelde ik mij vaak:

- Gelukkig
- Zenuwachtig
- Opgewonden
- ...

- Scoringsindex = $\frac{\text{aantal correcte}}{\text{benodigde tijd}}$

- Verschillende correcte antwoorden delen door gebruikte/benodigde tijd

Zoek و	و	ه	ه	و	و	و	و	و	و	و	و	و
Zoek ن	ن	ل	ل	ك	ف	ل	ك	ر	ن	ن	ل	ز
Zoek ...												

Het doel van meetschalen:

- 1) Hypothese toetsen

Vb. Is intelligentie één of meerdimensioneel?

- 1) Exploratief

Vb. zijn er onderliggende dimensies terug te vinden in een persoonlijkheidsvragenlijst?

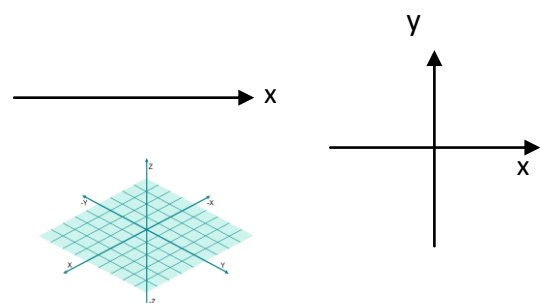
- 1) Bepaald aspect scoren of meten

Vb. totaalscore op de houding ten opzichte van migranten voor een bepaalde persoon

Dimensionaliteit van meetschalen

- Dimensionaliteit:
 - Uni-dimensioneel: 1 dimensie
 - Twee-dimensioneel: 2 dimensies
 - Meer-dimensioneel: >2 dimensies

=> Een schaal kan uit meerdere dimensies bestaan



Empirische constructen	Abstracte constructen
Het aantal dimensies is makkelijk af te leiden	Het aantal dimensies is moeilijker af te leiden
Vb. hoogte, gewicht, BMI	Vb. zelfzekerheid, altruïsme, aandacht, intelligentie

Vbn : IQ: VIQ en PIQ

- Verbaal : woordenschat
- Performaal : ruimtelijk inzicht , probleemoplossend denken
- sociale oriëntatie
 - 2 dimensies
- causale attributietheorie
 - Oorzaken die je toeschrijft aan iets => vb
 - slecht examen => slecht gestudeerd (intern)
 - slecht examen => prof moet mij niet (extern)
- Likert schaal: unidimensioneel
 - Constructie v/e unidimensionele Likert schaal:
 - STAPPEN:
 - 1) Omschrijving van unidimensionele construct
 - 2) Veel items opstellen die aansluiten bij deze omschrijving
 - 3) Items voorleggen aan proefgroep (deel van populatie)
 - 4) Itemscores en schaalscores analyseren

Meetniveau

METEN = toekennen van cijfers of categorielabels

⇒ METEN is vaststellen welke waarde een object op een variabele heeft

Het niveau van de meting verschilt van variabele tot variabele.

Vb. geslacht <-> IQ

Verschillende meetniveaus:

- 1) Nominaal:
 - Classificatie in categorieën
 - De categorieën zijn enkel kwalitatief te onderscheiden, niet kwantitatief
 - Vb.:
 - Geslacht
 - Partijkeuze
 - Diagnostische categorieën uit de DSM-IV
 - De sport die iemand beoefent

NIET	WEL
Rekenkundige bewerkingen	Frequenties tellen

2) Ordinaal:

- Classificatie in categorieën
- De categorieën kunnen gerangordend worden
- Vb.:
 - Scholingsgraad: LO, SO, HO

- SES
- Graden: voldoening, onderscheiding, grote onderscheiding, grootste onderscheiding
- Likert schaal

NIET	WEL
Optellen en vermenigvuldigen => Je mag geen gemiddeldes berekenen	Frequenties, $x > y$, mediaan, modus, minimum, maximum

3) Interval:

- Afstanden tussen twee waarden hebben betekenis
- Wel meeteenheid, maar kan veranderd worden
- Geen absoluut nulpunt
- Vb.:
 - Temperatuur
 - (IQ)
 - (Angstschaal)

NIET	WEL
$x = 2y$ $x = \frac{1}{2} y$	Wiskundige bewerkingen zoals gemiddelde, standaarddeviatie, correlatie,...

4) Ratio:

- Afstanden tussen twee waarden hebben betekenis
- Wel meeteenheid, maar kan veranderd worden
- Ook absoluut nulpunt (oorsprong is vast)
- Vb.:
 - lengte
 - Massa

NIET	WEL
$x = 2$ $x = y + 3$	Wiskundige bewerkingen zoals gemiddelde, standaarddeviatie, correlatie, ... $x = 2y$ $x = \frac{1}{2} y$

5) Absoluut:

- Er is geen andere schaal mogelijk
 - Meeteenheid vast
 - Absoluut nulpunt
 - Vb.:
 - Aantal kinderen in een gezin
- Hoeveel paar schoenen iemand

NIET	WEL
	ALLES

- Oefening: Wat is het meetniveau van de volgende variabelen?:
 - hartslagfrequentie (r)
 - rangpositie van voetbalteams (o)
 - leessnelheid (r)
 - voorkeur (n)
 - tijdstip van de dag (i)
 - de afstand tussen je woon- en werkplek®
 - aantal fouten (a)

Scoring en transformatiewaarden

- RUWE SCORE = basisscore waarin een testresultaat in eerste instantie wordt uitgedrukt => later basisscore transformeren

Voorbeelden:

- Aantal correcte antwoorden
- Aantal items waarvoor de antwoorden in dezelfde richting wijzen
- Benodigde tijd om opdrachten uit te voeren

NADEEL: op zich weinig betekenis

- TRANSFORMATIEMEETWAARDE = een aangepaste ruwe score die interpreteerbaar is

Voorbeelden:

- Stefaan gaf bij 16 van de twintig items een antwoord dat wijst op depressie (vgl met max)
- Jan haalde 80% op fysica (vgl met max)
- Joris scoorde meer dan het gemiddelde van de klas op de rekentoets (vgl met anderen, willekeurige set van anderen))
- Katrien had veel minder tijd nodig voor de aandachtstest dan de gemiddelde tijd die kinderen van haar leeftijd nodig hebben (vgl met anderen, representatieve groep van van anderen)
- Je kan er een betekenis aan koppelen

Absolute transformatie-meetwaarden (nog niet vgl met anderen)	Vergelijkende transformatiemeetwaarden	
	Toevallige transformatie-meetwaarden	normen
Onafhankelijk van prestaties van anderen	Afhankelijk van prestatie van een toevallige groep anderen	Afhankelijk van prestatie van representatieve groep van vergelijkbare anderen
Vb. % juist Vergelijking met leeftijd	Vb. Rangnummers Percentielen standaardscores	Vb. Leeftijdsnormen Percentiëlnormen standaardnormen

=> representatieve steekproef : als het alle lagen van de populatie vertegenwoordigt.

- **Percentage juist** (= vergelijking met perfecte prestatie)

$$100 \times \frac{\text{aantal correcte antwoorden}}{\text{totaal aantal items}}$$

Vb. 14 correcte antwoorden bij 20 vragen = 70%

- Nadeel: Hoe moeilijk was de test?
- => kunnen we pas te weten komen door te vergelijken met anderen

- **Vergelijking met leeftijd**

Vb. Intelligentie Quotiënt (IQ)

- **Binet:** proeven in categorieën van 3 tot 12 jaar:

Vb. Een proef behoort tot de categorie [5 jaar] als:

- 2/3 (of 3/4) van de vijfjarigen de proef met succes afleggen
- Minder dan 2/3 (of 3/4) van de vierjarigen de proef met succes afleggen (voorwaarde)
- Pas als het 75% is dat de proef kan oplossen is het een proef voor die leeftijd

Vb. met 3/4	7-jarigen	8-jarigen	9-jarigen	10-jarigen	categorie
Proef A	52%	63%	70%	80%	10 jaar = leeftijdsschaal 10
Proef B	63%	77%	81%	95%	8 jaar = leeftijdsschaal 8

⇒ Dan is het een proef voor 5-jarige

100x Mentale leeftijd

Chronologische leeftijd

- **Mentale leeftijd** = hoogste leeftijdsschaal waarvan men alle proeven (uitgez. 1) met succes kan uitvoeren + 1/5 jaar voor elke proef uit een latere leeftijdscategorie die men toch met succes kan afwerken

Vb. Stel: elke leeftijdsschaal heeft 5 proeven

- En: de proefpersoon is 4 jaar en zes maanden oud

	Proef 1	Proef 2	Proef 3	Proef 4	Proef 5
Leefijdsschaal 3	X	X	X	X	X
Leefijdsschaal 4	X	X	X	X	X
Leefijdsschaal 5	X	X	X	X	X
Leefijdsschaal 6	X	X	X	X	X
Leefijdsschaal 7	X	X	X	X	X
Leefijdsschaal 8	X	X	X	X	X

- Groen is juist , rood is fout
- Bezwaren :
 - Samenhang tussen toenemende leeftijd en toenemende cognitieve prestatie verdwijnt na 15-16 jaar
 - Moeilijk om mentale leeftijden te bepalen bij een proef
- Opm: eigenlijk geen *absolute* transformatiewaarde want er is vergelijking met prestaties van een groep kinderen.

- **Vergelijking met toevallige referentiegroep**

- Mieke: 14/20

Leerling	Punten op Rekentoets (op 20)	Rangnummer
Els	17	2
Jan	13	6

Nathalie	16	3.5
Eline	12	7
Thijs	18	1
Veronique	9	9
Stijn	14	5
Sara	16	3.5
Jens	10	8

- **Rangnummers**

- Ordenen van laag naar hoog of andere som => 2x hetzelfde : gem nemen van de 2 rangnummers
- Nadelen:
 - Groepsgrootte is van belang
 - Interpretatie hangt af van de kwaliteit of aard van de groep

- **Fractielen**

- De verdeling van de scores wordt in gelijke fracties verdeeld volgens een bepaalde regel:
- Decielen: verdeling verdelen in 10 gelijke fracties
- Percentielen: verdeling verdelen in 100 gelijke fracties
- Vigintielen: verdeling verdelen in 20 gelijke fracties
- Kwartielen: verdeling verdelen in 4 gelijke fracties

percentiel	$P = \frac{F_j - 0,5f_j}{N} \times 100$
deciel	$D = \frac{F_j - 0,5f_j}{N} \times 10$
vigintiel	$V = \frac{F_j - 0,5f_j}{N} \times 20$
kwartiel	$K = \frac{F_j - 0,5f_j}{N} \times 4$

- F_j = cumulatieve frequentie van score j
- f_j = absolute frequentie van score j
- N = totaal aantal scores

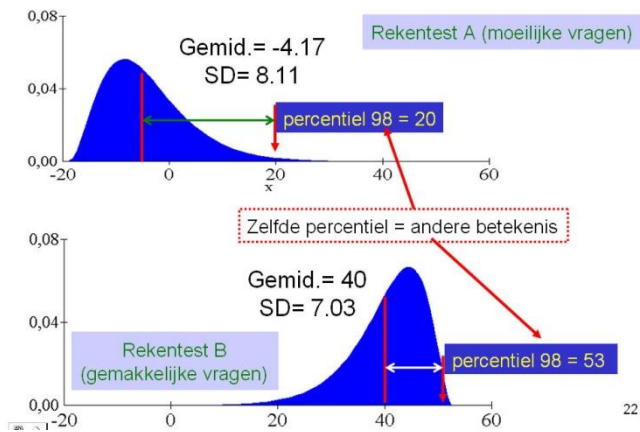
Voordelen:

- Grootte van de referentiegroep zit vervat in een fractielscore => niet meer afhankelijk van groepsgrootte
- Gemakkelijk toepasbaar en inzichtelijk (vb. Percentiel 80 = 80 % van de mensen behaalde een lagere score en 20% een hogere score)

Nadelen:

- Kwaliteit van de referentiegroep
- Afstanden tussen de ruwe scores blijven niet bewaard
- Fractielscores zijn sterk afhankelijk van de verdeling

- **Afstanden tussen ruwe scores blijven niet bewaard**
- **Fractielscores zijn sterk afhankelijk van de verdeling**



- Scores niet echt even goed!
- Score beter bij test A : scores liggen verder van zijn score
- SD bij A veel groter dan bij B
- **Standardscores** (= Het aantal standaardafwijkingen verschil tussen de meetwaarde en het gemiddelde)
 - De standaardafwijking erbij betrekken
 - Standaarddeviatie:

$$s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}}$$

- Variantie : zelfde formule , maar zonder de vierkantswortel

ppn	Test A	Score - gem.	(score - gem.) ²	Test B	Score - gem.	(score - gem.) ²
1	1	-4	16	3	1	1
2	2	-3	9	3	1	1
4	3	-2	4	3	1	1
5	5	0	0	4	0	0
6	7	2	4	5	1	1
7	8	3	9	5	1	1
8	9	4	16	5	1	1
	Gem=5		Som = 58	Gem=4		Som = 6
			Var = 58/7 = 8,29			Var = 6/7 = 0,86
			Stdev = 2,88			Stdev = 0,93

lineair	genormaliseerd
Hoeveel standaardafwijkingen ligt de ruwe score boven of onder het gemiddelde in de WAARGENOMEN verdeling?	Waargenomen verdeling wordt eerst omgezet naar een normaalverdeling. Hoeveel standaardafwijkingen ligt de score boven of onder het gemiddelde in de NORMAALverdeling?

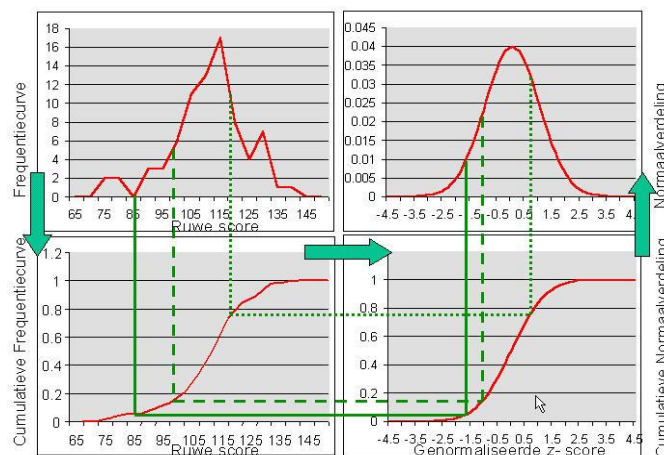
- Z=1 : 1x de standaarddeviatie afwijken van het $\bar{X}$
- Z=2 : 2x de standaarddeviatie afwijken onder het $\bar{X}$

- **z-scores**

- Omzetting van Z-scores
 - Om de decimalen te gaan wegwerken (.10)
 - Meer kans op positieve waarden (makkelijker) (+50)
- Deviatie
 - Afwijking ten opzichte van het gemiddelde
- lineair

$$Z_{x_i} = \frac{X_i - \bar{X}}{s_x}$$

- X_i = de ruwe score waarvan men de standaardscore wil bepalen
- $\bar{X}$ = de gemiddelde ruwe score
- s_x = de standaarddeviatie van de ruwe scores
- Gemiddelde van de z-scores = 0
- Standaarddeviatie van de z-scores = 1
- genormaliseerd



- z-tabel
- Vb. percentielscore = 6% = 0,06 (6 procent van de observaties liggen onder de score van deze persoon)
- => z-score = -1,55 (deze persoon presteert ongeveer anderhalve standaarddeviatie onder het gemiddelde)

- **t-scores**

- lineair

$$T_{x_i} = 10 \left(\frac{X_i - \bar{X}}{s_x} \right) + 50$$

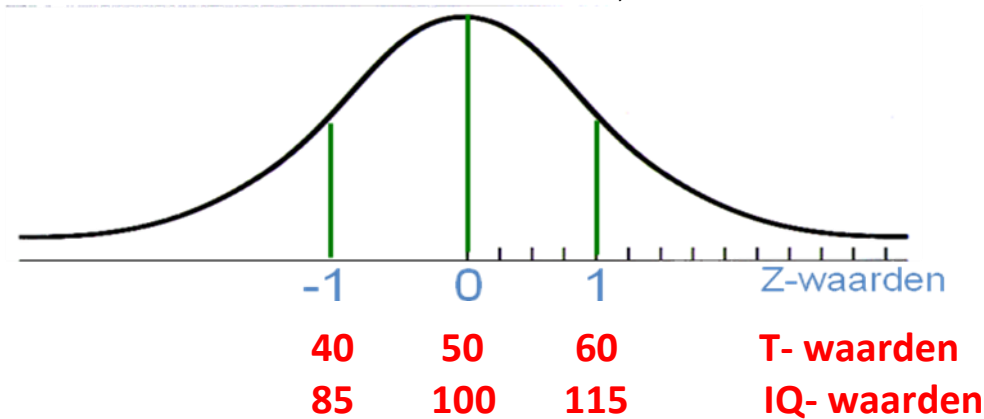
wegwerken van decimalen →
 wegwerken van negatieve waarden →

- X_i = de ruwe score waarvan men de standaardscore wil bepalen
- $\bar{X}$ = de gemiddelde ruwe score
- s_x = de standaarddeviatie van de ruwe scores

- Gemiddelde van de t-scores = 50
- Standaarddeviatie van de t-scores = 10
- Vb. CBCL

Voordeel:

- negatieve waarden en decimalen weggewerkt
- Genormaliseerd
- Verband tussen t-scores, z-scores en deviatie-IQ: $Z_{X_i} = \frac{T_{X_i} - 50}{10}$



- **deviatie-IQ**

- X_i = de ruwe score waarvan men de standaardscore wil bepalen
- $\bar{X}$ = de gemiddelde ruwe score
- S_x = de standaarddeviatie van de ruwe scores
- Gemiddelde van deviatie-IQ = 100
- Standaarddeviatie van deviatie-IQ = 15

$$\text{DeviatieIQ} = 15 \left(\frac{X_i - \bar{X}}{S_x} \right) + 100$$

- **Stanines**

- Standard nines => stanines
- De stanines worden als volgt vastgesteld:
- Breedte van een stanine = stdev/2
- Gemiddelde van de verdeling ligt in het midden van de 5^e stanine
- Vergelijking met een standaardnormale verdeling

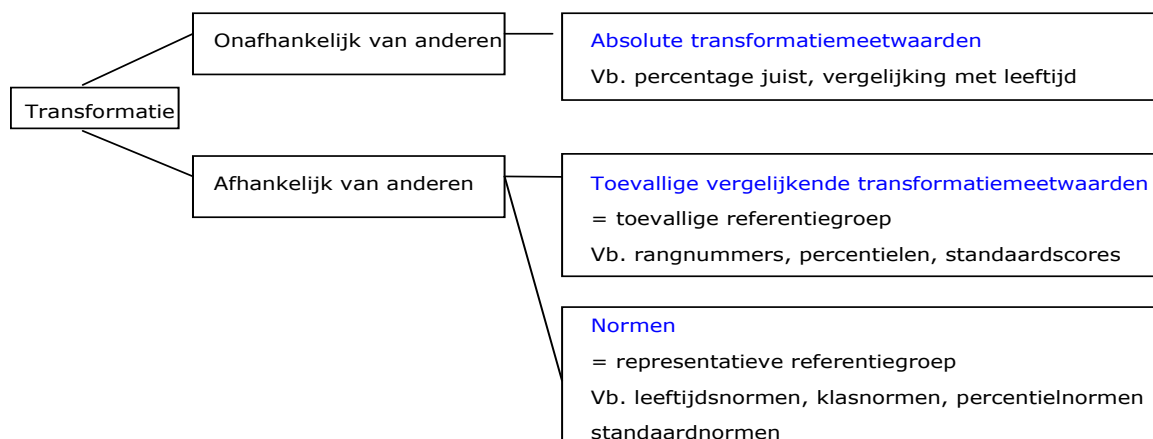
	4%	7%	12%	17%	20%	17%	12%	7%	4%
Stanine	1	2	3	4	5	6	7	8	9

- **C-scores**

- Zelfde als stanines, maar meer klassen (11)
- Vergelijking met standaardnormale verdeling:

	1%	3%	7%	12%	17%	20%	17%	12%	7%	3%	1%
C score	0	1	2	3	4	5	6	7	8	9	10

- **Transformatiemeetwaarden**



- Het verschil ligt in de groep waarmee we gaan vergelijken
- **Representatieve vergelijking (normen)**
 - Voorbeeld: test ontwikkelen voor zelfbeeld van kinderen ts. 9-12 jaar
 - Welke kenmerken zijn belangrijk voor het construct waarin we geïnteresseerd zijn? Geslacht, leerjaar in basisonderwijs,...
 - Bij de testconstructie nemen we de test af bij een representatieve steekproef van kinderen ts. 9-12 jaar (populatie is immers onmogelijk!); in de steekproef is de verdeling van de kenmerken (bv. geslacht, leerjaar) dezelfde als in de populatie
 - → Representatieve steekproef = steekproef is een goede afspiegeling van de populatie
 - Norm
 - transformatiemaatwaarden die werden bekomen voor een representatieve referentiegroep bv. vlamingen, leeftijdsgroep
 - Normeren
 - vastleggen van de normen van een test door test af te nemen bij een representatieve steekproef
 - de handeling, het vastleggen
 - Ruwe scores vergelijken met norm, onafhankelijk van prestatie in toevallige groep

- **Leeftijdsnormen** (score van persoon vergelijken met score van leeftijdsgenoten)

- **Klasnormen** (niet vergelijken met 1 specifieke leeftijd, testresultaten gaan vergelijken per leerjaar, tabellen en grafieken voor klasnormen, bv. 1^e leerjaar = 5 en 6 jaar)

- **Percentiëlnormen** (zelfde als percentielen, afgeleid op basis van een normgroep, niet toevallig bepaald, gemakkelijk af te lezen => niet zelf uitrekenen)

- **Standaardnormen** (ruwe score en standaarddeviatie vergelijken met bepaalde scores van normgroepen)

Raw Score	Equiv.		Raw Score	Equiv.		Raw Score	Equiv.		Raw Score	Equiv.	
	Age	Grade		Age	Grade		Age	Grade		Age	Grade
1	6-1	-	26	7-11	2.6	51	9-1	3.8	76	10-8	5.5
2	6-2	-	27	7-11	2.7	52	9-2	3.9	77	10-10	5.6
3	6-4	-	28	8-0	2.7	53	9-2	3.9	78	10-11	5.7
4	6-5	-	29	8-0	2.8	54	9-3	4.0	79	11-0	5.8
5	6-6	-	30	8-1	2.8	55	9-4	4.0	80	11-2	5.9
6	6-7	-	31	8-1	2.9	56	9-4	4.1	81	11-3	6.0
7	6-8	-	32	8-2	2.9	57	9-5	4.1	82	11-5	6.1
8	6-9	-	33	8-3	3.0	58	9-6	4.2	83	11-6	6.2
9	6-10	-	34	8-3	3.0	59	9-6	4.2	84	11-7	6.3
10	6-11	-	35	8-4	3.1	60	9-7	4.3	85	11-9	6.5
11	7-0	-	36	8-4	3.1	61	9-8	4.4	86	11-11	6.6
12	7-1	-	37	8-5	3.2	62	9-9	4.4	87	12-0	6.8
13	7-2	-	38	8-5	3.2	63	9-9	4.5	88	12-2	6.9
14	7-3	-	39	8-6	3.3	64	9-10	4.6	89	12-4	7.1
15	7-3	-	40	8-6	3.3	65	9-11	4.6	90	12-6	7.2
16	7-4	-	41	8-7	3.3	66	10-0	4.7	91	12-7	7.3
17	7-5	-	42	8-7	3.4	67	10-0	4.8	92	12-9	7.4
18	7-6	-	43	8-8	3.4	68	10-1	4.8	93	12-10	7.5
19	7-7	-	44	8-8	3.5	69	10-2	4.9	94	13-0	7.7
20	7-7	2.2	45	8-9	3.5	70	10-3	5.0	95	13-2	7.9
21	7-8	2.2	46	8-10	3.6	71	10-3	5.1	96	13-4	8.0
22	7-8	2.3	47	8-10	3.6	72	10-4	5.1	97	13-6	8.2
23	7-9	2.4	48	8-11	3.7	73	10-5	5.2	98	13-8	8.3
24	7-9	2.4	49	9-0	3.7	74	10-6	5.3	99	13-11	8.5
25	7-10	2.5	50	9-0	3.8	75	10-7	5.4	100	14-2	8.7

Percentiel punten	Chronologische leeftijd in jaren													
	8	8½	9	9½	10	10½	11	11½	12	12½	13	13½	14	
95	38	39	41	43	45	48	50	51	51	52	52	53	53	
90	34	36	38	41	43	45	47	49	49	50	50	51	52	
75	24	29	32	34	37	39	41	43	45	46	47	48	48	
50	18	21	24	28	30	33	35	37	39	41	43	44	44	
25	-	14	16	18	20	23	26	29	32	34	35	37	38	
10	-	-	-	13	13	15	16	18	22	25	27	28	28	
5	-	-	-	-	13	14	15	16	17	19	21	23		

Schaal per sexe	Mannen (N = 958)		Vrouwen (N = 1390)		Alle groepen (N = 2415)	
	M	SD	M	SD	M	SD
Neuroticisme	29.6	7.8	32.2	8.2	31.1	8.2
Extraversie	39.8	6.5	40.3	6.6	40.1	6.6
Openheid	35.4	6.6	36.3	6.3	35.9	6.4
Altruïsme	42.5	5.1	45.1	5.0	44.1	5.2
Conscientieusheid	45.3	5.7	45.3	5.6	45.3	5.6

=> conscientieusheid : nauwgezet werken

- **Ipsatieve scores**

- Normscores
 - Scores vergelijken met prestaties van een groep
- Ipsatieve scores
 - scores van individu vergelijken met eigen prestaties op andere facetten
 - Gebruik: gedwongen keuze vragen (gedwongen om 1 keuze te maken boven alle andere)
 - Bv. interesseltest
- Nadelen
 - geen persoonlijke waardering alle items. Men is gedwongen de ene activiteit boven de andere te kiezen terwijl men beide even prettig of onprettig kan vinden
 - afstand tussen de voorkeuren komt niet tot uiting

vragen	schaal	ipsatief antwoord	normatief (likert) schaal
ik ben empathisch	S	x	1 2 3 4 5
ik win meestal	D		1 2 3 4 5
ik heb veel vrienden	S	x	1 2 3 4 5
ik heb veel ideeën	D		1 2 3 4 5
totale score : D = 0/10			D= 5/10
S = 10/10			S= 7/10

Likert schaal : op een schaal aangeven voor hoever iets bij jou van toepassing is.

~ > sociaal wenselijk gedrag invullen: dit zal wel beter zijn voor die job waarschijnlijk (bv.)

Een klein verschil in voorkeur wordt in de ipsatieve scores sterk overtrokken

Schaalscores zijn niet onafhankelijk: kiezen voor S = minder punten voor D

- **Norm-, criterium- en communicatie-georiënteerde instrumenten**

- Interpretatie van testresultaten
 - Normgeoriënteerde instrumenten
 - Normgroep of referentiegroep (ruwe scores vgl met een bepaalde normgroep om de scores te interpreteren)
 - Vergelijkende transformatiemaatwaarden: bv. T-score, IQ-score, C-score, ...
 - Gebruik binnen diagnostisch proces
 - screening grote groepen
 - bepalen van de aanwezigheid en ernst van een afwijking of achterstand
 - Voorbeelden: intelligentietest, CBCL (Child Behavior Check List : maakt gebruik van t-scores) / YSR
 - Criteriumgeoriënteerde instrumenten
 - = 'Criterion referenced' tests of (criterium)toetsen
 - Vergelijking tussen het gedrag van de onderzochte persoon met gewenst gedrag; vastgelegd doel; criterium
 - Gebruik binnen diagnostisch proces
 - plannen van interventie
 - Bijv. prestatietoets, examen (vgl met een vastgelegd doel => bepaald criterium : 10/20)
 - Communicatiegeoriënteerde instrumenten
 - Communication-referenced tests
 - Doel: het ontdekken van het individueel beleven en functioneren van de onderzochte persoon

- Geen vergelijking met groep of expliciet criterium
- Algemeen welbevinden van een persoon in kaart brengen
- Vanuit de psychologische achtergrond gebruiken om te interpreteren
- Bijv. onderzoek binnen sociaal-emotionele domein; anamnese

Aard van het gemeten gedrag: testindeling

- **Instrumenten en tests die betrouwbare, precieze en valide informatie verstrekken over**
 - Welke soorten bestaan er om te gaan meten?
 - eigenschappen, kenmerken of gedragingen van personen
 - toename in verscheidenheid
 - dus noodzakelijk om testen in te delen => verschillende soorten van classificatiesystemen
 - ontwikkeling van personen
 - kenmerken van omgeving
 - Verschillende soorten tests → categorisatie van tests
- **Soorten psychologische instrumenten**
 - **Test**
 - Verzameling verbale, performale en/of motorische opdrachten of een toestel om één of meer variabelen te meten
 - Verzameling van allerlei opdrachten
 - **Vragenlijst**
 - Aantal vragen en/of beweringen dat één of meer variabelen meet
 - Papier en potlood test ; bepaalde beweringen, vragen, stellingen
 - **(Klinische) Beoordelingsschaal**
 - Schriftelijke gedragsbeoordeling mbv psychologische kennis
 - **Observatieschaal**
 - Schriftelijke gedragsbeoordeling mbv observatie. Het betreft concrete omschrijvingen van gedrag.
 - Op basis van kennis een beoordeling geven
 - **Evaluatieschaal**
 - Aantal vragen en/of beweringen om variabelen te meten met doel personeel, trainingen, onderwijs te evalueren
 - **Batterij**
 - Combinatie van verschillende tests of vragenlijsten, die ook elk afzonderlijk kunnen worden gebruikt
 - Verschillende testen of in combinatie met vragenlijsten => om algemeen functioneren in kaart te brengen
 - **(Meet)schaal**
 - Aantal beweringen dat slechts één variabele meet
- Indeling obv testgedrag
 - **Prestatieniveau tests of 'maximum' performance**
 - Bepaald niveau gaan meten van iets en in kaart brengen, zicht krijgen op de maximale prestatie van de persoon
 - Een goed en een fout antwoord
 - Geeft de mate van iets (niveau)
 - Bijv. intelligentie, vaardigheden, geheugen, redeneren, inzicht of (school)prestaties
 - **Tests voor gedragswijze ('typical')**
 - Persoon beter leren kennen
 - Geen goede of foute antwoorden
 - Een kwaliteitsmeting – wat persoon er van vindt

- *Bijv. neuroticisme, extraversie, persoonlijkheid, temperament, obsessies, depressie, angst...*
- Indeling obv testmateriaal
 - **Papier –en potloodtesten**
 - Antwoord aankruisen of antwoord geven op vraag
 - *Bijv. persoonlijkheidsvragenlijst*
 - **Tests met speciaal materiaal**
 - Opdrachten via speciaal materiaal zoals puzzels, blokken, spel materiaal, computergestuurd testen
 - Materiaal gebruiken buiten de vragelijsten
 - *Bijv. intelligentietests, projectieve tests, persoonlijkheidstests en neuropsychologische tests op PC*
- Indeling obv categorieën (COTAN) => boek met allerlei testindelingen
 - **PersoonSituatieTests**
 - Situaties
 - Vaardigheden en attitudes worden nagegaan
 - Gedragscategorieën: oa. cognitie, aandacht, perceptie, emotie...
 - Wat die persoon vindt van een bepaalde situatie, hoe hij zich gedraagt
 - **PersoonlijkheidsTests**
 - Vaardigheden en attitudes
 - Gedragscategorieën
 - **SituatieTests**
 - Situaties
 - Documentatie van Tests en Testresearch in Nederland (Evers et al., 2000)
 - Beschrijving van alle instrumenten die voor handen zijn in Nederland
 - van de Commissie voor Test Aangelegenheden Nederland (COTAN)
 - ontwikkeling door personengroep
 - Bevat testbeschrijvingen en beoordelingen (cfr PSD-II)
 - Oa. doelpopulatie, meetpretentie, verkrijgbaarheid, wijze en duur afname, gebruik, uitwerking, beoordeling, ...
 - Beschrijvingen voor wie het is bedoeld, meetpretentie (wat bedoelt de test te meten), verkrijgbaarheid, wijze van afnemen (individueel ↔ groep), duur afname + beoordeling (kwaliteitsoordeel)
- Indeling obv type item
 - **Item: taak, opdracht of vraag**
 - **Antwoordvorm**
 - Gesloten of open (interview, projectieve testen,...)
 - Meerkeuze of multiple choice
 - Antwoord zelf produceren
 - **Aard van antwoord**
 - Discreet (dichotoom, polytoom) meerdere antw. mogelijk
 - *Bijv. Ben je tevreden met je werk? 0 Ja 0 Neen*
 - *Bijv. Ik voel me tevreden: bijna nooit soms vaak bijna altijd*

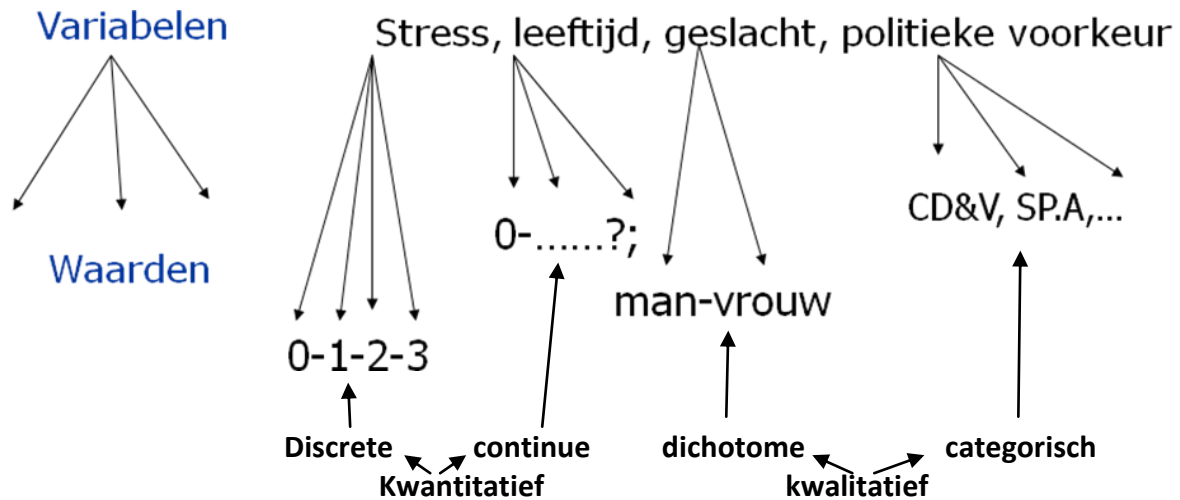
De NIO, Nederlandse Intelligentietest voor Onderwijsniveau (Van Dijk & Tellegen, 2004) is door de COTAN als volgt beoordeeld:

COTAN-beoordeling NIO 2004

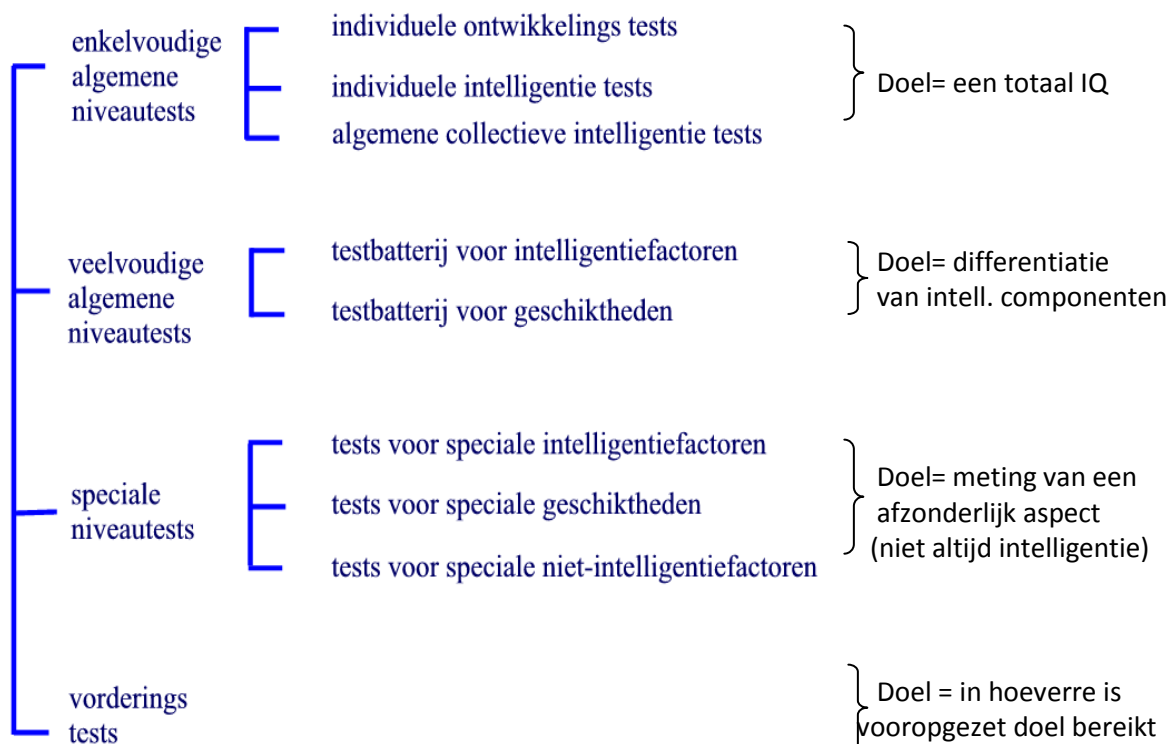
I	Uitgangspunten bij de testconstructie	goed
IIa.	Kwaliteit van het testmateriaal	goed
IIb.	Kwaliteit van de handleiding	goed
III	Normen	goed
IV.	Betrouwbaarheid	goed
Va.	Begripsvaliditeit	goed
Vb.	Criteriumvaliditeit	goed

- Continu
 - Bijv. responstijd

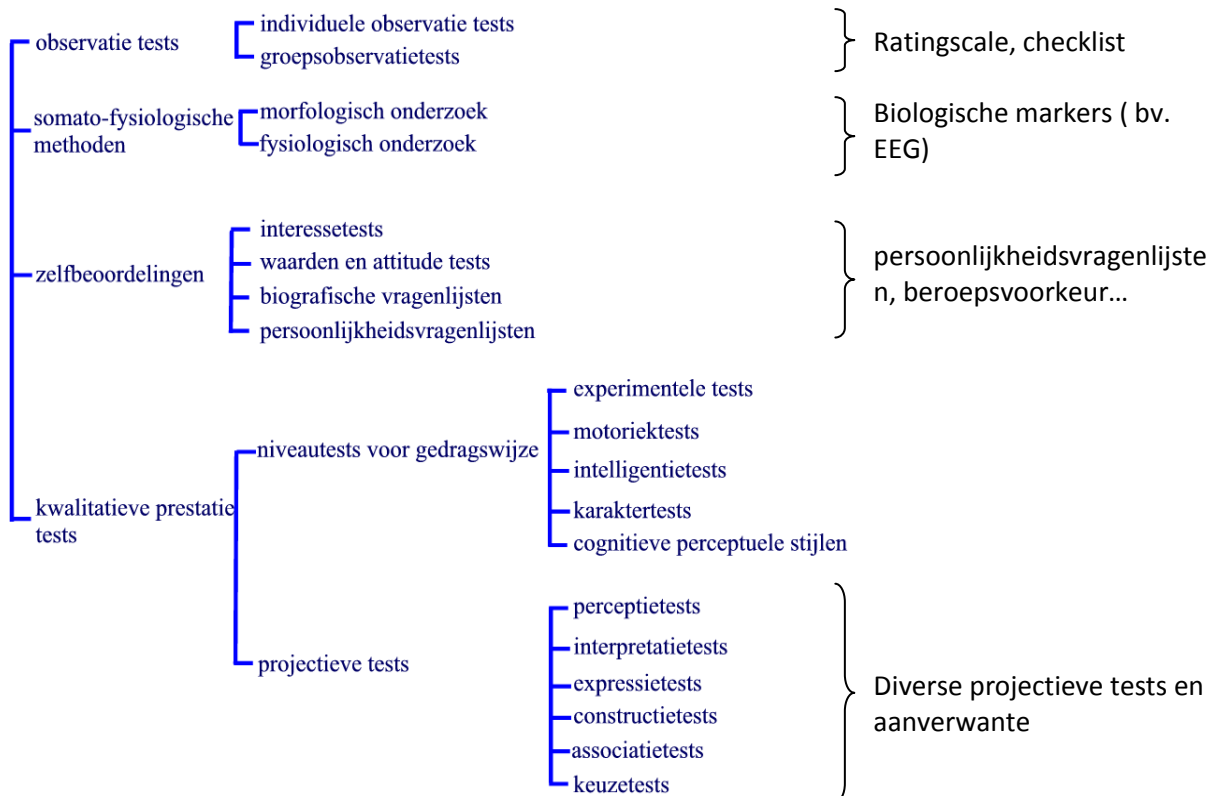
Begrippen: soorten variabelen



- Individueel vs in groep
 - **Individuele tests**
 - Voordelen: Observatie, zeer veel informatie
 - Nadelen: Hoge kost, rol van onderzoeker is complex
 - **Groepstests**
 - Voordelen: Efficiënter, economischer, minder tijds intensief, eenvoudige vergelijking, beperkte rol van onderzoeker
 - Iedereen dezelfde instructies, dezelfde tijd => gestandaardiseerd => mogelijkheid om te vergelijken
- **Tests voor prestatieniveau**



- **Tests voor gedragswijze**



- **Enkelvoudige algemene niveautests**

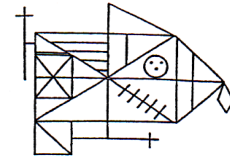
- Individuele ontwikkelingstests
 - Kinderen en jongeren bv. WPPSI-R, WISC-III
 - Individuele intelligentietests
 - Volwassenen bv. WAIS, GIT, KAIT
 - tests die bestaan uit heterogene reeks testopgaven ofwel uit een samenstelling van reeksen van homogene opgaven (subtests)
 - quotering van de prestaties in punten = ruwe scores; omzetting van de ruwe scores in standaardscores en IQ-score
- Collectieve algemene intelligentietests
 - Groepsgewijs en schriftelijk
 - Vb: NDT ("algemene collectieve schoolintelligentietest")

- **Veelvoudige algemene niveautests**

- Testbatterij: bestaat uit een aantal relatief op zichzelf staande homogene subtests
 - testbatterij voor intelligentiefactoren
 - testbatterij voor geschiktheden
- De score op elk van deze subtests wordt afzonderlijk berekend en kan zowel op zichzelf als in vergelijking met scores op andere subtests worden geïnterpreteerd.
- Itt enkelvoudige niveautests waarvan subtests peilen naar specifieke aspecten van intelligentie (bijv. WISC-III: VIQ wordt berekend op basis van verschillende subtests)
- Testbatterij voor intelligentiefactoren
 - Deeltest meet een homogene dimensie van intelligentie
 - Vb. Drenth's testserie voor hoger niveau: TNVA, VAT'69, NAT'70
- Testbatterij voor geschiktheden
 - pretendeert vermogens te meten die iemand in staat stellen zich voor een maatschappelijke taak of schooleisen te bekwamen

- Vb: General Aptitude Test Battery (GATB), Differential Aptitude Testbattery (DAT)
- Vb. General Aptitude Test Battery: Motorische Subtest
- Aptitude Test: 3-dim. Ruimtelijk Denken
 - Voor elke blok met een letter. Met hoeveel andere blokken komt dit blok in contact
- Aptitude Test: Abstract Redeneren
 - Wat is het volgende plaatje?
- Aptitude Test: Mechanisch Redeneren
 - Een bal wordt geworpen, de andere valt:
 - Welke bal raakt eerst de grond?
- Aptitude Test: 2-dim Redeneren
 - Welke figuur rechts komt overeen met de linkse?
- Aptitude Tests: Sorteren, Klasseren en Ordenen
- Aptitude Tests: Verbale Aanleg Test
 -staat tot boot zoalsstaat tot auto
 - 1. Duwen, 2. Pedaal, 3. Varen, 4. Rollen
A. ophalen, B. trekken, C. heffen, D. rijden
- Aptitude Tests: Numerieke Aanleg Test
 - Wat is het volgende cijfer?
 - 1 3 6 8 11 13 (+2, +3)
 - 21 18 15 12 9 (-3)
 - 1 2 4 7 11 16 (+1,+2,+3,+4...)
- Aptitude Tests: Rekenkundige Aanleg Test
 - Wat is het gewicht van een kist van 10 dekens indien elk deken 3.2 kg weegt?
 - Wat zal een arbeider verdienen indien er 40 uren wordt gewerkt tegen 11 Euro/u en 6 uren tegen 14 Euro/u?
- **Speciale niveautests**
 - **Test voor speciale intelligentiefactoren**
 - Gericht op speciale intelligentiefactoren, bv. ruimtelijk inzicht, woordkennis, creatief denken, ...
 - Vb. 'Alternatief Gebruik': *Geef zoveel mogelijke toepassingsmogelijkheden van een krant, buiten het 'normale gebruik'*
 - **Tests voor speciale geschiktheden**
 - Gericht op specifieke vaardigheden zoals leergeschiktheden (geheugentests, tempotests, concentratietests), technische aanleg, werkgeschiktheid ... bv. Bourdon-Wiersma-test
 - **Tests voor speciale niet-intelligentiefactoren oa. neuropsychologische tests**
 - aandachts-, concentratie- en geheugentests bv. Figuur van Rey
 - sensorische tests
 - algemene motorische tests
 - Vb: test voor speciale intelligentiefactoren
 - **Raven's Progressive Matrices (1938)**
 - ➔ Oplossing voor de sterke cultuurverankering van intelligentietests
 - Non-verbale "**Culture Fair**" test voor algemene intelligentie
 - Geen (of minimale) invloed van scholing of cultuur

- Kan eventueel toch nog ongelijk zijn
 - *Culturen die weinig gebruik maken van visuele informatie*
 - Visueel ruimtelijke capaciteiten
- Vb. Speciale niveautest: Complexe Figuur van Rey (CFR)



• Vorderingentests

- Doel: nagaan in hoeverre de onderzochte het doel van een opleiding heeft bereikt
- specifiek op schoolvorderingen gerichte kennistest
vb. examen, LVS
- algemeen op vaardigheden (buiten de directe cognitieve sfeer en kennis) georiënteerde vaardigheidstest
vb. goed kunnen koken
- Vb. Hamilton depression rating scale

For each item, write the correct number on the line next to the item. (Only one response per item)

- 1. DEPRESSED MOOD** (Sadness, hopeless, helpless, worthless)

 0= Absent
 1= These feeling states indicated only on questioning
 2= These feeling states spontaneously reported verbally
 3= Communicates feeling states non-verbally—i.e., through facial expression, posture, voice, and tendency to weep
 4= Patient reports VIRTUALLY ONLY these feeling states in his spontaneous verbal and non-verbal communication
- 2. FEELINGS OF GUILT**

 0= Absent
 1= Self reproach, feels he has let people down
 2= Ideas of guilt or rumination over past errors or sinful deeds
 3= Present illness is a punishment. Delusions of guilt
 4= Hears accusatory or denunciatory voices and/or experiences threatening visual hallucinations
- 3. SUICIDE**

 0= Absent
 1= Feels life is not worth living
 2= Wishes he were dead or any thoughts of possible death to self
 3= Suicidal ideas or gesture
 4= Attempts at suicide (any serious attempt rates 4)

- Vb. Somato-fysiologische methoden
 - Galvanic Skin Respons:
meet elektrische geleiding (of weerstand)
 - Geleiding verhoogd bij
 - stress,
 - angst,
 - opwinding of
 - shock
- Vb. Somato-fysiologische methoden
 - **Actigraph:** meting van motorische activiteit over lange periodes

Standaardisatie

- Standaardisatie in testafname
- Standaardisatie in scoring & verwerking
→ Handleiding van de test
- Gestandaardiseerde afname
 - Tests en items afnemen volgens precies voorgeschreven condities
 - Doel: objectieve vergelijking tussen personen
- Zelfde condities voor alle personen
 - Plaats: lokaal, ruimte apart,..

- Niet veel lawaai
- Samen: voldoende ventilatie
- Testafname
 - Testsituatie: het lokaal: licht, warmte, geluid, ruimte, aanwezigheid van afleiders, ...
 - Testmateriaal: testboekjes, puzzels, blokjes, toetsenbord, monitor, ...
 - Testinstructies: evenveel tijd, niet meer of minder hulp bieden, zelfde afbreekregels, ...
 - Houding en gedrag van de testleider
- Verwerking
 - Scoring: voorgeschreven richtlijnen, objectief
 - Normering

Gestandaardiseerde afname

Uitzonderingen

- 'testing the limits', leerpotentieel-tests
 - prestaties onder standaardconditie vgl met prestatie onder conditie met hulp
- Ongestructureerde observatie
- Computergestuurd testen
 - Papier en potloodtest: volledige testafname
 - Computergestuurd testen: mogelijkheid stoppen tot betrouwbare schatting dmv items volgens oplopende moeilijkheidsgraad en discriminerende waarde

4. Betrouwbaarheid

Betrouwbaarheid als alledaags begrip p. 187 – 189

- **Gebruik in uitspraken over personen of zaken:**
 - Mijn auto is helemaal niet betrouwbaar want er is regelmatig een probleem.
 - Mijn vriend is betrouwbaar want hij is er steeds wanneer ik hem nodig heb.
 - De Belgische treinen zijn niet betrouwbaar want ze rijden nooit op tijd.
 - -> Betrouwbaarheid verwijst naar de mate van consistentie van iets, herhaalbaarheid in identieke situaties, naar een zeker mate van vertrouwen
- **3 betekenissen van betrouwbaarheid mbt. personen:**
 - **Kenmerk dat gewaardeerd wordt**
 - te vertrouwen, eerlijk, je kan er beroep op doen,...
 - **Persoonlijkheidskenmerk waarop mensen verschillen**
 - sommige mensen vinden we betrouwbaar, anderen niet
 - is gevolg van hoe we mensen zien over verschillende situaties heen (thuis, op werk, in vriendenkring) en over de tijd heen (vroeger, nu)
 - verband met Big Five persoonlijkheidskenmerken: betrouwbaarheid ~ consciëntieusheid
 - onbetrouwbaarheid : wel veel fouten bij het vaststellen, invloeden die ervoor zorgen dat je metingen onbetrouwbaar zijn
 - **Verantwoord vaststellen van kenmerken van personen**
 - Betrouwbaarheid als kenmerk van het instrument waarmee je kenmerken van personen vaststelt
 - Onbetrouwbaar: fouten maken bij het vaststellen/de meting van kenmerken
 - -> betrouwbaarheid van instrumenten
 - **Betrouwbaar instrument:**
 - Metingen zijn consistent of herhaalbaar
 - Metingen bevatten weinig onsystematische/toevallige meetfouten

Betrouwbaarheid van instrumenten

Klassieke testtheorie (KTT)

Herhaling: Model als component van psychodiagnostiek

Psychodiagnostiek is een compositie bestaande uit drie componenten (cfr les 2):

- **Theorie**
 - psychologische theorievorming over cognitie, intelligentie, emotie, motivatie en gedrag
 - theorieën zijn onder te verdelen in 3 oriëntaties: individuele verschillen, ontwikkeling, context
- **Model:** afbeelding van de theorie en constructen
 - formele, kwantitatieve modellen
 - kwalitatieve modellen
- **Instrument:**
 - middelen die we gebruiken om iets te weten te komen over cognitie, emotie, motivatie, persoonlijkheid,...
 - psychologische tests en meetinstrumenten

Model

- **Afbeelding** van centrale kenmerken van theorieën, theoretische begrippen (constructen) in een model
- **Operationaliseren van een construct** (= meetbaar maken)
 - meten op een schaal
 - *vb. subtest Cijferreeksen WISC-III -> 'geheugen'*
 - onderbrengen in een categorie
 - *vb. DSM-IV 'depressie'*
- **Modellen kunnen afgeleid zijn uit de drie oriëntaties**
 - Individuele-verschillenoriëntatie
 - Ontwikkelingsoriëntatie
 - Contextoriëntatie
- Daarnaast: **kwantitatieve modellen** ontstaan uit de testleer en de statistiek
- **Testleer of psychometrie**
 - een verzameling van modellen om bepaalde eigenschappen van items –en testscores af te beelden en te toetsen (kunnen gebruikt worden voor de verschillende oriëntaties!)
- Enkele belangrijke modellen:
 - *Klassieke testtheorie (KTT)* (dit jaar)
 - *Moderne testtheorie of Item Respons theorie (IRT)* (volgend jaar)

Klassieke Testtheorie (KTT) (p.120-122)

- Vertrekpunt KT
 - Uitleggen aan de hand van een fictief voorbeeld : verschillende items gaan hercoderen => omgekeerd zetten
 - Vb. vragenlijst trek-angst (ZBV, van der Ploeg, 2000)
 - hercodering items 1, 3, 6, 7, 10, 13, 14, 15, 16, 19; somscore
 - -> hoe hoger de somscore, hoe meer trek-angst
 - Vb. vragenlijst trek-angst bij dezelfde persoon afnemen
 - Dag 1: somscore 45
 - Dag 2: somscore 46
 - Dag 3: somscore 42
 - ...

soms licht verschil tss de scores : construct blijft hetzelfde, maar somscore gaat variëren
 - Bij herhaald testen van hetzelfde kenmerk – terwijl het kenmerk ongewijzigd blijft – is de kans groot dat je niet dezelfde somscore bekomt → KTT

- Hoe komt dit?
 - Eerste keer zeer alert
 - 2^{de} keer bv: minder geconcentreerd, minder geslapen
 - => factoren die kunnen zorgen voor onsystematische meetfouten (meetfouten die buiten de test liggen)
- Meetfouten
 - Hoe komt dit?
 - persoon (vb. psychisch, fysisch), omgeving (vb. licht, temperatuur, tijdstip), procedure (vb. instructie, verwerking)
 - = toevallige omstandigheden die leiden tot onsystematische meetfouten
 - Kunnen allemaal de testuitslag gaan beïnvloeden
 - Uitgangspunt KTT: variatie in somscore is het gevolg van onsystematische (toevallige, random) meetfouten
 - Het is wenselijk dat testcores zo weinig mogelijk beïnvloed worden door toevallige omstandigheden
 - onsystematische meetfouten: de ene keer ligt de score hoger, de andere keer lager; niet altijd zelfde, systematische invloed
 - meetfouten kunnen ook systematisch zijn:
 - operationalisering van het beoogde construct is niet perfect, je meet ook een beetje iets anders dan je wil meten (bv. motoriek bij intelligentietest); heeft een systematische, regelmatige invloed bij elke testafname
 - dus:

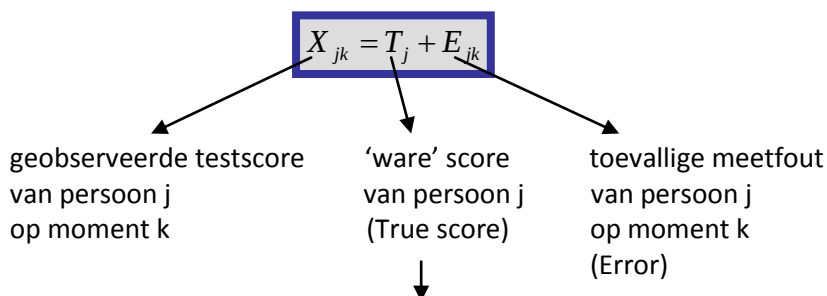
Testscore = beoogde construct
 + systematische meetfouten
 + onsystematische meetfouten

KTT : schatting van grootte van random meetfout bij test

KTT: geen aandacht aan systematische meetfouten

KTT biedt ons de mogelijkheid om de grootte van de onbetrouwbaarheid te schatten/berekenen

- Model en assumpties
 - Model A: Herhaalde metingen bij dezelfde persoon



Op elke moment hetzelfde dus de k moet er niet bij, ware score maakt niet uit op welk moment
 => construct blijft hetzelfde

- Tj: 'Ware' score van persoon j
 - Onafhankelijk van het moment waarop je de test afneemt van persoon j (daarom geen k index)
 - gemiddelde testscore over en groot aantal (q) denkbeeldige onafhankelijke replicaties (herhaalde metingen) bij persoon j
 - omwille van niet-systematische meetfouten zal de geobserveerde testscore (X) de ene keer wat groter zijn en de andere keer wat kleiner => elkaar opheffen
 - Gevolg: de positieve en negatieve afwijkingen in de geobserveerde scores tov de 'ware' scores zullen elkaar opheffen

- Constant gedeelte
- Onafhankelijk van het moment waarop je test afneemt
- vb. (fictief) $T_j = 6$

$X_{j1} = 7$ ($E_{j1} = +1$)
 $X_{j2} = 5$ ($E_{j2} = -1$)
 $X_{j3} = 9$ ($E_{j3} = +3$)
 $X_{j4} = 4$ ($E_{j4} = -2$)
 $X_{j5} = 5$ ($E_{j5} = -1$)

$$T_j = \frac{1}{q} \sum_{k=1}^q X_{jk} = \overline{X}_j$$

$$X_j = \frac{X_{j1} + X_{j2} + X_{j3} + X_{j4} + X_{j5}}{5} = \frac{7+5+9+4+5}{5} = \frac{30}{5} = 6$$

Assumptie: KTT gaat ervan uit dat afnames onderling onafhankelijk zijn (bv. geen geheugeneffecten, geen leereffecten) Realistisch?

○ **Ejk: Meetfout van persoon j op moment k**

- De verwachte waarde van alle toevallige meetfouten is 0. De toevallige fouten kunnen net zo goed leiden tot een geobserveerde testscore die de 'ware' score overschat als onderschat.
- Vb. (fictief): $\frac{1-1+3-2-1}{5} = \frac{0}{5} = 0$

$$E_v(E_{jk}) = 0$$

Som van elke meetfout / aantal replicaties

○ **S(Ej): Standaardmeetfout van persoon j**

- = standaarddeviatie van de foutencomponent E_j
- Afwijkingen van meetfout ten opzichte van gemiddelde meetfout = Standaardmeetfout => gemiddelde van elke meetfout is 0 => dus we kunnen het gemiddelde schrappen
- We weten dat $E_{jk} = 0$ en $E_{jk} = X_{jk} - T_j$ => dus $S(E_j) =$
 - = afwijking van de testscores van persoon j op momenten k tov de 'ware score' van persoon j
 - = indicatie voor hoe 'goed' we gemeten hebben bij persoon j
 - een kleine standaardmeetfout betekent dat de metingen dicht bij T_j liggen
 - een grote standaardmeetfout betekent dat er veel fluctuatie op de metingen zit tov. T_j
- De standaarddeviatie van de meetfout van persoon j (standaardmeetfout van persoon j) =
- De standaarddeviatie van alle metingen (geobserveerde scores) bij persoon j
- We hebben het liefst dat deze waarde zo klein mogelijk is
- **Assumptie:**
 - In KTT gaat men ervan uit dat de standaardmeetfout voor alle personen dezelfde is op een bepaalde test.
 - Realistisch?(bv. minder vaardigheid leidt tot meer gokken)
- In KTT gaat men ervan uit dat de standaardmeetfout S(Ej) voor alle personen dezelfde is op een bepaalde test.
- Waarom heeft KTT deze assumptie nodig?
- In realiteit is het niet mogelijk om van één bepaalde persoon heel veel keren de test af te nemen om de standaardmeetfout te kunnen berekenen.
- Oplossing: we nemen van een zeer grote groep personen de test één maal af; op basis van die gegevens kunnen we dan iets te weten komen over de standaardmeetfout van de test.

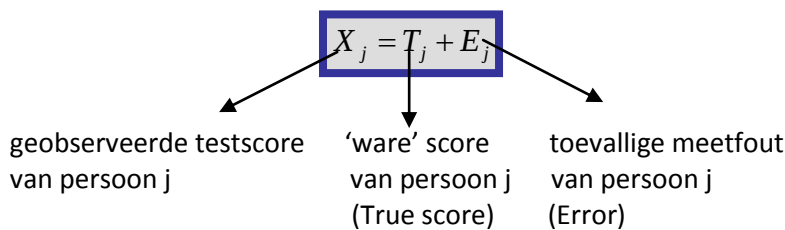
$$s_x = \sqrt{\frac{1}{N} \sum_{i=1}^N (X_i - \overline{X})^2}$$

$$S(E_j) = \sqrt{\frac{1}{q} \sum_{k=1}^q (E_{jk} - \overline{E}_{jk})^2}$$

• **Model B: Metingen bij verschillende personen in populatie (p. 121)**

○ **op één moment**

- -> k valt weg in model A



Elke persoon heeft deze 3 elementen

Alle meetfouten optellen en delen door verschillende personen, dan krijg je 0 => ze heffen elkaar terug op

$$E_v(E_j) = 0$$

- De gemiddelde meetfout in een populatie van n personen is gelijk aan 0
- In een groep personen correleert de meetfout nergens mee tenzij ze zelf deel uitmaakt van wat gemeten wordt
- Correlatie
 - Samenhang tussen twee interval-variabelen (sterkte + richting verband)
 - Betekenis kennen : verband aangeven tussen 2 variabelen op intervalniveau (zonder oorzaak/verband), de mate van samenhang tussen...
 - - : betekenis van negatieve correlatie : $x \blacktriangleup$ en $y \blacktriangledown$
 - + : betekenis van positieve correlatie : $x \blacktriangleup$ en $y \blacktriangleup$
- In een groep personen correleert de meetfout nergens mee tenzij ze zelf deel uitmaakt van wat gemeten wordt
 - Geen verband tussen meetfout en ware score. De meetfout is immers niet systematisch. Personen met een hoge T hebben geen systematische grotere of kleinere E dan personen met een lage T.
 - Er is een positief verband tussen E en X. E maakt immers deel uit van X. Dus hoe hoger E, hoe hoger X; hoe lager E, hoe lager X.
- In een groep personen correleert de meetfout nergens mee tenzij ze zelf deel uitmaakt van wat gemeten wordt
 - Geen verband tussen meetfout en willekeurige variabele waarvan E geen deel uitmaakt.
- In een groep personen is de gemiddelde geobserveerde score gelijk aan de gemiddelde 'ware' score.
 - Immers $E = 0$ (zie slide 27)
- Variantie van geobserveerde scores
 - Variantie van geobserveerde scores = variantie van ware scores + variantie van meetfout => formule zeer belangrijk om te komen tot de betrouwbaarheid.
- Meetfout is niet systematisch

$$S^2(X) = S^2(T) + S^2(E)$$

Waarom KTT nodig?

- Betrouwbaarheid testscore gemeten in populatie van personen

$$= \frac{S^2(T)}{S^2(X)}$$
- Je krijgt een verhouding, de ware score variantie gedeeld door de variantie in de geobserveerde scores => betrouwbaarheid : altijd tussen 0 en 1
- Bv. .85 betekent dat 85% van de variantie in de testcores (X) afkomstig is uit de variantie in de 'ware' scores van personen
- Betrouwbaarheid : zekerheid, herhaalbaarheid, consistentie
- Betrouwbaarheid

$$= \frac{S^2(T)}{S^2(X)}$$

$$= \frac{S^2(X) - S^2(E)}{S^2(X)} = \frac{S^2(X)}{S^2(X)} - \frac{S^2(E)}{S^2(X)}$$

$$= 1 - \frac{S^2(E)}{S^2(X)}$$

- $S^2(E) = S^2(X) \cdot (1 - \text{betrouwbaarheid}) \Rightarrow$ Standaardmeetfout van testscore

Enkele bezwaren bij KTT (p. 131-13)

- De standaardmeetfout is voor iedereen gelijk (zie slide 23)
Vb. Mensen met meer vaardigheid zullen minder 'gokken' op prestatietest met meerkeuzevragen (dus $S(E)$ is kleiner)
- Men kijkt enkel naar de somscore op een test
Twee personen met zelfde somscore kunnen op een verschillende manier tot die somscore komen. Dit is enkel gerechtvaardigd als ieder item in de test uitwisselbaar is met de andere items. Maar dit is meestal niet het geval.
Vb. sommige items zijn moeilijker dan andere in een vaardigheidstest; sommige items zijn een betere meting van het construct dat je wil meten dan andere items
Cfr. Betrouwbaarheid voor nog andere bezwaren

Betrouwbaarheid : meetfouten en KTT

KTT houdt zich bezig voornamelijk met onsystematische meetfouten

Testscore = beoogde construct

+ systematische meetfouten

+ onsystematische meetfouten

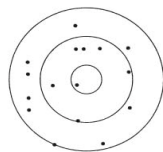
Validiteit

Meet de test wat hij hoort te meten?

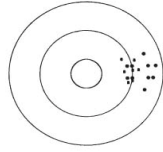
betrouwbaarheid

Bekomen van dezelfde testscore als we dezelfde test herhaaldelijk afnemen bij dezelfde persoon
Herhaalbaarheid en consistentie

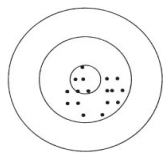
- Betrouwbaarheid en validiteit gaan beide over het beperken van meetfouten, maar gaan elk over een andere soort meetfout



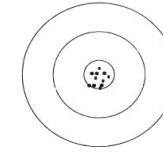
Niet valide,
Niet betrouwbaar



Niet valide,
Betrouwbaar



Matig valide,
Niet betrouwbaar



Valide,
Betrouwbaar

- Middelste cirkel : het construct dat je wil meten
- Puntjes : herhaalde meetresultaten
- Hoe dichter de puntjes bij elkaar liggen, hoe betrouwbaarder
- Hoe dichter de puntjes bij of in het middelste cirkeltje liggen, des te meer valide is de test
- Hoe kleiner de spreiding van onsystematische fouten, hoe 'herhaalbaarder' (= betrouwbaarder) je meting is
- Validiteit wordt ook deels beïnvloed door betrouwbaarheid $\Rightarrow$ betrouwbaarheid basis voor validiteit

$$S^2(X) = S^2(T) + S^2(E)$$

hoe kleiner de spreiding van onsystematische fouten, hoe 'herhaalbaarder' (= onbetrouwbaarder) je meting is

- variantie van geobserveerde testscore = var. van ware score + var. onsystem. Meetfout
- Betrouwbaarheid = $\frac{S^2(T)}{S^2(X)}$
 - Betrouwbaarheid = hoeveel van de variantie is de geobserveerde testscore X is afkomstig uit de variantie van de ware scores T? (getal ts. 0-1)
 - Vb. Als $S^2(T)$ bijna gelijk is aan $S^2(X)$ (dus $S^2(E)$ is zeer klein) dan is de betrouwbaarheid van meting hoog (nl. $S^2(T)/S^2(X) \sim 1$)
 - Als $S^2(T)$ veel kleiner is dan $S^2(X)$ (dus $S^2(E)$ is groot) dan is de betrouwbaarheid van de meting kleiner (nl. $S^2(T)/S^2(X) \ll 1$)
- Standaardmeetfout $S(E) = S(X)\sqrt{1 - \text{betrouwbaarheid}}$
 - Een hoge $S(E)$ betekent: een grote standaarddeviatie van de onsystematische meetfouten bij herhaalde metingen (maw. herhaalbaarheid is laag)
 - $S(E)$ speelt een rol bij het bepalen van betrouwbaarheidsintervallen (zie volgende les) vb. IQ
 - Men wil van een test de betrouwbaarheid of standaardfout kennen.
 - probleem: men kan deze niet berekenen want $S^2(T)$ is onbekend
 - gevolg: de betrouwbaarheid van een test moet worden geschat
 - hoe wordt de betrouwbaarheid geschat? via het berekenen van correlaties
 - Betrouwbaarheid kan je nooit berekenen
 - Enkel schatten, omdat je de ware score niet kan berekenen
 - "berekenen" adhv correlaties

Hoe wordt de betrouwbaarheid geschat? via het berekenen van correlaties tussen testcores op twee 'identieke' tests

	Vb. zelfde test afnemen op twee momenten	
	Moment 1	Moment 2
Pers 1	$X_{11} = T_1 + e_{11}$	$X_{12} = T_1 + e_{12}$
Pers 2	$X_{21} = T_2 + e_{21}$	$X_{22} = T_2 + e_{22}$
Pers 3	...	...
....		

$$r(X_1, X_2) \Rightarrow B$$

- B
 - hoge correlatie $\Rightarrow$ hoge betrouwbaarheid
 - lage corr. $\Rightarrow$ meer meetfouten
- Wanneer er veel onsystematische meetfouten zijn dan zal dit nu eens tot een hogere testscore leiden, en dan eens tot een lagere testscore. Gevolg: $r(X_1, X_2)$ zal klein zijn als er veel meetfouten zijn. Dus r zegt iets over betrouwbaarheid.

1. dezelfde test 2 x afnemen: test hertest methode
 2. twee parallelle tests afnemen: parallelvorm methode
 3. één test afnemen: halveringsmethode
 4. één test afnemen: interne consistentiemethode
- } 2 afnames
 } 1 afname
- 1 tot 4 leveren 'betrouwbaarheidscoëfficiënten'
 - 4 manieren om betrouwbaarheid te gaan meten $\Rightarrow$ 4 betrouwbaarheidscoëfficiënten
 - Generaliseerbaarheidcoëfficiënt (p. 191) = coëfficiënt die rekening houdt met de verschillende soorten betrouwbaarheidscoëfficiënten

Methoden voor het schatten van betrouwbaarheid

Test-hertestmethode

Dezelfde test twee maal afnemen van dezelfde personen met een tussentijd

	Moment 1	Moment 2	
Pers 1	X11	X12	
Pers 2	X21	X22	
Pers 3	X31	X31	
...	...	...	

$r(X \text{ afname1}, X \text{ afname 2})$

= schatting van betrouwbaarheid
= $S^2(T)/S^2(X)$
= $r_{xx}(\text{stabiliteit})$
= stabiliteitscoëfficiënt
= $r_{XX}(\text{stabiliteit}) = r(X \text{ 1}^e \text{ afname}, X \text{ 2}^e \text{ afname})$

- **Nadelen:**

- Niet geschikt om betrouwbaarheid te schatten van tests die constructen meten die niet stabiel zijn over tijd (bv. gemoedstoestand)
- r_{xx} wordt dan immers ook bepaald door verandering op construct (T)
- Niet geschikt om betrouwbaarheid te schatten van tests waar er een leer- of geheugeneffect kan optreden
- Leereffect: mensen leren uit oefeningen uit 1e afname; treed vooral op bij inzichtsopgaven (het systeem doorhebben)
- Geheugeneffect: opgaven die duidelijk in het oog springen worden onthouden en bij volgende afname sneller herkend en opgelost
- Geen probleem voor r_{xx} indien leer- en geheugeneffect voor iedereen op dezelfde wijze zou gebeuren, maar in realiteit verschillen mensen hierin
- Leer- en geheugeneffecten spelen vooral een rol wanneer tijdsinterval tussen twee testafnames klein is.
- Een te groot tijdsinterval houdt dan weer het risico in dat sommige mensen ondertussen veranderd zijn op het construct dat je wil meten waardoor r_{xx} geen goede maat is voor betrouwbaarheid.
- Vb. risico op leer- en geheugeneffecten
 - *Progressive Matrices Raven*
 - *Geheugenproef 15 woorden*

Paralleelvormmethode

- Twee inwisselbare maar niet identieke tests afnemen van dezelfde personen (ipv. twee replicaties van dezelfde test)
- ⇒ problemen ivm leer- en geheugeneffecten verminderen
- niet de identieke test, maar een paralleltest
 - wil hetzelfde construct meten
 - equivalent testen
 - dus minder leer- en geheugeneffecten
- Wanneer zijn 2 test parallel? $T_j = T_j'$
- De ware score van een persoon op test J is dezelfde als de ware score op test J'
- ⇒ dit kan je echter nooit nagaan
- ⇒ criteria waaraan de testscores X_j en X_j' moeten voldoen
 1. $X_j = X_j'$

$$2. S^2(X_j) = S^2(X_j')$$

$$3. r(X_j, Y) = r(X_j', Y)$$

- Voorwaarden :
 - Ware construct dat je meet moet dezelfde zijn
 - Kan je niet berekenen
 - Gemiddelde gaan berekenen

	Test J		Test J'		Test Y
Pers 1	X1		X1'		Y1
Pers 2	X2		X2'		Y2
Pers 3	X3		X3'		Y3
...	...	←	...	→	...
		$r(X, Y)$			
			$r(X', Y)$		

$\bar{X}_j = \bar{X}_j'$	}	gem. moet gelijk zijn
$S^2(X_j) = S^2(X_j')$		var. Moet gelijk zijn

$$r(X_j, X_j') = \text{schatting van betrouwbaarheid}$$

$$= S^2(T)/S^2(X)$$

$$= r_{xx'} \quad (\rho_{XX'} = \rho_{XT} = \sigma^2 T / \sigma^2 X)$$

- **Nadelen:**
 - - Paralleltesten maken is niet eenvoudig:
 - Items moeten gelijkwaardig zijn maar mogen niet identiek zijn (dus je moet vragen hebben die er anders uitzien maar hetzelfde meten)
 - Vaak nog wel mogelijk voor cognitieve of vaardigheidstests, maar vaak moeilijk voor persoonlijkheidsvragenlijsten
 - In praktijk zullen twee paralleltests niet perfect parallel zijn. Gevolg: de betrouwbaarheid wordt onderschat
 - Opm. paralleltests worden soms gebruikt bij hertesting van een persoon (bv. evaluatie na hulpverlening)

Halveringsmethode

- Eén test afnemen bij alle personen en test splitsen in parallelle helften. Werkwijze:
 - Volledige test afnemen
 - Score op helft 1 en score op helft 2 berekenen voor iedere persoon (in 2 delen splitsen)
 - Correlatie tussen score op helft 1 en score op helft 2 (tss de 2 testhelften)
 - Correctie op correlatie om betrouwbaarheid van hele test te schatten

	Test J helft 1	Test J helft 2
Pers1	X1 helft1	X1 helft 2
Pers 2	X2 helft 1	X2 helft 2
Pers 3	X3 helft 1	X3 helft 3
...	...	...

$$r(X_{\text{helft 1}}, X_{\text{helft2}}) = \text{schatting voor betrouwbaarheid}$$

$$= S^2(T)/S^2(X)$$

$$= r_{xx}(\text{equivalent}) \quad (\rho_{XX(\text{equivalent})} = r(X_{\text{helft 1}}, X_{\text{helft 2}}))$$

Zo krijgt men een schatting van de betrouwbaarheid

- **Hoe splits je een test in 2 helften?**
 - Een splitsing in de eerste helft van de items en de laatste helft van de items is niet aan te raden

- Vermoeidheid, verveling, leereffecten in 2e helft waardoor 2e helft niet als parallelversie van 1e helft kan worden beschouwd
- Wanneer de test van makkelijke naar moeilijke items gaat kan de 1e helft van de test beroep doen op andere vaardigheden (-> andere T) dan 2e helft
- Een splitsing volgens odd-even methode is geschikter
- **Waarom moet men een correctie toepassen op de r tussen beide helften? Welke correctie?**
 - De betrouwbaarheid van een halve test is lager dan de betrouwbaarheid van de hele test omdat de somscores op een korte test minder stabiel zijn dan de scores op een langere test.
 - Als er zich een toevallige meetfout (bv. vergissing, gelukte gok) voordoet dan heeft dit een grotere impact op de somscore van kortere tests dan op langere tests.
- Spearman Brown Prophecy Formule

Betrouwbaarheidstest van hele test $\leftarrow r_{xx'} = \frac{2r_{yy'}}{1+r_{yy'}}$ $\rightarrow$ betrouwbaarheid van halve test (nl. r tss 2 testhelften)

- Algemene Spearman Brown Prophecy Formule:
 - Invloed van testlengte op betrouwbaarheid

Betrouwbaarheid 'gewenste' test $\leftarrow r_{xx'} = \frac{Kr_{yy'}}{1+(K-1)r_{yy'}}$ $\rightarrow$ 'feitelijke waarde' die je berekent

$\downarrow$
verlengings- of verkortingsfactor

- K
 - 2 bij correctie ivm halveringsbetrouwbaarheid
 - Hoeveel keer meer items we eraan moeten toevoegen
 - Als je weet wat K is dan kan je de gewenste test schatten

$\Rightarrow$ met hoeveel items moet een test verlengd worden om een gewenste betrouwbaarheid te krijgen?

$$K = \frac{r_{xx'}(1-r_{yy'})}{r_{yy'}(1-r_{xx'})}$$

$\nearrow$ betrouwbaarheid 'feitelijke' test
 $\searrow$ betrouwbaarheid 'gewenste' test

-
- Vb. een test met 5 items heeft een betrouwbaarheid van .40, maar men wil een betrouwbaarheid van .80. Hoeveel items moet men bijmaken?
- er zijn dus 6x meer items in gewenste test dan in feitelijke test; gewenste test bevat dus 30 items -> er moeten items bijgemaakt worden

$$K = \frac{.80(1-.40)}{.40(1-.80)} = 6$$

Interne consistentiemethode

- Uitgangspunt: items zijn onderling inwisselbaar (onderling parallel). Elk item kan als een deel van een test worden beschouwd. Analoog aan gedachte uit halveringsmethode
- Gevolg: de scores op de items worden gebruikt
- Men gaat na wat de samenhang (covariantie) is van alle items onderling
- **Formule van Cronbach (Cronbach's alpha)**

$$\rho_{XX'} \geq \frac{N}{n-1} 1 - \frac{\sum \sigma^2 Y_i}{\sigma^2 X} = \alpha$$

N = aantal items
in de test

Som van de varianties van de item scores

$$\alpha = \frac{N}{N-1} \left(1 - \frac{\sum_i S^2(Y_i)}{S^2(X)} \right)$$

verscheidenheid
van de items

verscheidenheid tussen de scores van de personen

Variantie van de totale ruwe score

- Bij relatieve kleine verscheidenheid (=homogeen) van de vragen zal alpha toenemen.
- Hoe meer items je toevoegt, hoe meer jegaat stijgen
 - Je verzamelt meer en meer info => nauwkeurigheid

• Meerdere formules voor interne consistentie

- **Cronbach's Alpha**
Formule voor gebruik bij niet-dichotome items
- **Kuder-Richardson 20 (KR20)**
Formule voor gebruik bij dichotome items

$$KR20 = \frac{N}{N-1} \left(1 - \frac{\sum_i p_i q_i}{S^2(X)} \right)$$

- P : proportie item correct
- van dichotome items
- q : proportie van item fout
- proportie personen die item goed hebben geantwoord x proportie personen die item fout hebben geantwoord
- Beide waarden (Cronbach's alpha en KR20) worden gebruikt als schatter van de betrouwbaarheid van een test
- Betrouwbaarheid ≥ alpha
- deze maat voor interne consistentie van de test geeft een onderschatting voor de "echte" betrouwbaarheid van de test
- ondergrens voor de betrouwbaarheid want de waarde is gelijk of groter dan alpha

Opmerking: Vaak gebruikt als maat voor homogeniteit of uni-dimensionaliteit → Fout (om betrouwbaarheid te gaan schatten)

Cronbach's alpha is een toenemende functie van het aantal items

Bij testverlenging nadert alpha naar 1

Vb. alpha = 0.90

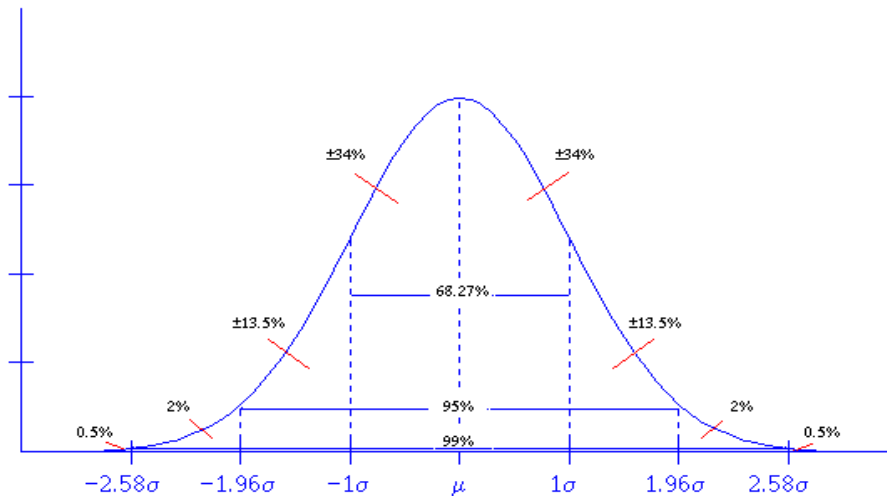
- Lange test met lage interne consistentie
- Korte test met hoge interne consistentie

Gebruik van de betrouwbaarheid p. 131-132 , 190-191

- Ware score zal normaal rond het gemiddelde vallen => 50
- Meet een oneindig aantal maal de ruwe score, dan zij de scores normaal verdeeld en is het gemiddelde van deze verdeling de True score (ware testscore)
- ...de standaarddeviatie van de scores rondom het gemiddelde = standaardmeetfout = standard error – S(E)
- True score
 - Soms beperkte fouten
 - Kleine afwijkingen van de ware score
 - Kleine SE
 - Soms grotere fouten
 - Grotere afwijkingen van ware score
 - Grotere SE (standaardmeetfout)

Betrouwbaarheidsinterval

95% van alle ruwe scores (uit de oneindige testing) valt tussen -1.96 en $+1.96$ maal de standaardmeetfout (SE)
rond de WARE score



Maar ook: In 95% van de gevallen valt de WARE score ergens tussen -1.96 en $+1.96$ maal de standaardmeetfout rond de ruwe score

Verband tussen betrouwbaarheid en SE(X)

Hoe hoger de betrouwbaarheid, hoe kleiner de meetfout

Hoe lager de betrouwbaarheid, hoe groter de meetfout

Rxx	SD	SE	Betr. int.	Marge
1	15	0	90-100	0
0.90	15	4.74	90-110	20
0.80	15	6.71	86-114	28
0.70	15	8.22	83-117	34
0.60	15	9.49	81-119	38

←
Zeer onnauwkeurige meting

→ uitrekenen

→ betrouwbaarheid daalt => SE stijgt

→ foutenmargen worden altijd groter als interne uitbreidt

Betrouwbaarheid van verschilsscores

- Gebruikt bij:

- Voor – na meting van één persoon
- Verschil tussen twee personen
- Verschil tussen twee verschillende tests

standaardmeetfout van verschilscore

$$SE_{diff} = \sqrt{(SE_1)^2 + (SE_2)^2}$$

$$SE_{diff} = SD\sqrt{2 - r_{11} - r_{22}}$$

- We moeten tenminste 9 deviatie IQ punten verschil hebben voor we kunnen spreken van een werkelijk verschil (met 95% zekerheid)
-

Problemen van betrouwbaarheid bij KTT

- De schatting van de betrouwbaarheid is gevoelig voor de variantie in de steekproef
- Regressie naar het gemiddelde
- Niet alle betrouwbaarheidsindexen zijn zinvol bij alle soorten tests (bv geen splitsingsmethode of interne consistentie bij speedtests)
- De schatting van de betrouwbaarheid is gevoelig voor de variantie in de steekproef, bijv de homogeniteit van steekproef (gecontextualiseerd, restriction of range)
- Voorbeeld: Probleem bij correlatieberekening, als gevolg van te homogene selectie van personen
- Wat is het verband met betrouwbaarheid?
 - De variabiliteit (spreiding) van de ware score
 - Indien variantie van T afneemt
 - Betrouwbaarheid is ook een correlatie

$$r_{xx} = \frac{S^2(T)}{S^2(T) + S^2(E)}$$

Problemen van betrouwbaarheid bij KTT: Regressie naar het gemiddelde

- een statistisch fenomeen, opeenvolgende metingen in een populatie neigen meer naar de gemiddelde waarde
- als een score fluctueert, dwz niet perfect betrouwbaar is, gaat bij een herhaalde meting de score van een groep hoge scores per definitie omlaag en bij een groep lage scoorders omhoog
- deze verandering is een gevolg van het feit dat een test niet perfect betrouwbaar is.
- illustratie: hertesten met gelijkwaardige IQ-tests

Vuistregels voor de beoordeling van betrouwbaarheid

- **Hoe hoog moet een betrouwbaarheidscoëfficiënt zijn?** (p 192-193)
 - Vuistregels bij vraagstellingen die betrekking hebben op
 - belangrijke beslissingen op individueel niveau
 - minder belangrijke beslissingen
 - Wetenschappelijk onderzoek

5. Validiteit

Validiteit als alledaags begrip p. 193-197

- Doen van 'geldige' uitspraken over gedragingen van personen, gebeurtenissen en zaken
- In psychodiagnostiek

- Geldige, ware of valide uitspraken over (probleem)gedrag, cognities en gevoelens van cliënten
- validiteit heeft betrekking op tests en instrumenten, uitspraken en beslissingen
- Wat verstaan leken onder 'geldigheid' van uitspraken? Geldigheid voldoet aan vier criteria:
 - Correspondentie
 - Coherentie
 - Pragmatisch nut
 - Intersubjectiviteit
- Correspondentie
 - Heeft te maken met waarheidsgehalte
 - Is er een overeenstemming : ja / neen
 - Overeenstemming van wat iemand beweert met wat er werkelijk gebeurd is, met de buitentalige werkelijkheid
 - of Overeenstemming tussen een uitspraak (kenner) en de werkelijkheid (wereld)
 - Goede operationalisatie (heeft te maken met)
- Coherentie
 - Het passen van een uitspraak in dat wat iemand al eerder vertelde over zaken of personen
 - of Nieuwe uitspraak is niet in strijd met de bestaande uitspraak
 - Consistentie
- Pragmatisch nut
 - Het feit dat er door te praten over een persoon, gebeurtenis een probleem wordt opgelost
 - Uitspraak moet nuttig, functioneel zijn
- Intersubjectiviteit
 - Overeenstemming tussen beoordelaars
 - Als verschillende beoordelaars dezelfde uitkomsten geven

Validiteit als instrument

Betrouwbaarheid versus validiteit

In hoeverre zijn de beslissingen die we maken juist → betekenis/nuttig/geschikt

- **Betrouwbaarheid**
 - Exactheid bij meten
 - Toevallige fouten
 - Meetfouten en KTT
 - Error variabiliteit
 - Te berekenen
 - Term gebruikt voor: metingen
- **Validiteit**
 - Kwaliteit van de testen
 - Juistheid van beslissing
 - Systematische fouten : fouten die ontstaan door systematisch iets anders te meten (operationalisering niet perfect)
 - Validiteit heeft ook te maken met meetfouten → hier wel de systematische (dus los van de persoon, omgeving,...)
 - Bias
 - Bij resultaat een vertekening → bias (validiteit)
 - Bij resultaat een vertekening → error var. (betrouwbaarheid)
 - Geen formules

- → de validiteit bestaat niet
 - Onderzoek naar validiteit → meer en meer onderzoek die meer en meer aangeven over de validiteit → niet 1 getal
- → de betrouwbaarheid kan je gaan schatten
- Term gebruikt voor: conclusies, afleidingen, generalisaties en besluiten
 - Validiteit heeft te maken met AL deze termen → niet uitsluitend voor metingen (betrouwbaarheid)

Validiteit als psychometrisch begrip

- Waarheid: juistheid, nuttigheid, accuraatheid, interpreteerbaarheid, voorspellend vermogen en bijdrage aan beslissingen, ... (p 199)
 - Termen die verwant zijn aan validiteit
 - Validiteit speelt niet alleen een rol bij tests
 - Test validiteit verwijst naar:
 - “de mate waarin de vertaling van een theoretisch, abstract begrip naar een meetbare eenheid geslaagd is.” Hoe goed beantwoordt de test aan zijn doel?
 - maw “validiteit is in hoeverre een test meet wat het beweert te meten, in hoeverre het mogelijk is geldige gevolgtrekkingen uit de testcores te maken”
 - Validering van een test is een proces zonder einde
 - Validiteit is een verzamelterm: meerdere vormen van validiteit
 - Men maakt een onderscheid tussen verschillende soorten validiteit die elk op een andere manier onderzocht worden
 - Centraal bij validiteit staan twee aspecten
 - Test als voorspeller van ander gedrag: de mate waarin een test toelaat om een correcte uitspraak te doen over gedrag buiten de testsituatie (criterium, predictieve validiteit) → criteriumvaliditeit
 - Test als operationalisering van een psychologisch begrip: de mate waarin een testscore daadwerkelijk een goede indicator is van het construct dat de test beoogt te meten (**begripsvaliditeit, constructvaliditeit**) → de mate waarin de test geoperationaliseerd is
 - **Operationalisatie van psychologische constructen**
 - Meetbaar maken van *constructen*
 - Constructvaliditeit(*begripsvaliditeit*)
 - Angst, Ego-sterkte, Motivatie, locus-of-control, Prestatie-motivatie, zelf-actualisatie,...
- ⇒ Zijn niet direct waarneembaar

Het moderne ‘Unified’ validiteitsbegrip

- **Unified’ validiteitsbegrip: vier bronnen van evidentie**
 - 4 bronnen die meer aanwijzingen geven over de test => 4 bronnen van evidentie
 - Hierdoor kunnen we argumenten aanvoeren om iets te zeggen over de validiteit van een test

Bronnen

- Testinhoud
- Analyse van antwoordprocessen
- Interne structuur van de test
- Relatie test met externe variabelen

Klassieke validiteitssoorten

- inhoudsvaliditeit (1)
- constructvaliditeit (2)
- constructvaliditeit (2)
- criteriumvaliditeit (3)

• Testinhoud

- de test moet die bepaalde kennisinhoud, rekenvaardigheid of karaktertrek dekken

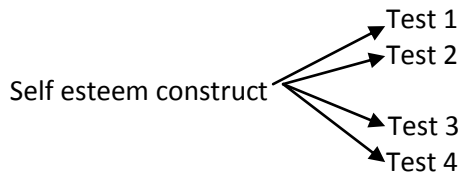
- inhoud verwijst naar de volledigheid waarmee iets wordt gemeten. De representativiteit van de testinhoud.
 - **Analyse van antwoordprocessen**
 - de test moet een goede representatie zijn van de kennis, vaardigheid of karaktertrek die we willen meten. Antwoord en responsprocessen nagaan.
 - de test meet het volledige construct en niets anders.
 - bv. vaststellen van rekenvaardigheid zonder beroep te doen op leesvaardigheid
 - **Interne structuur**
 - meten alle items, te weten komen door de interne structuur na te gaan van een test
 - relatie tussen items, tussen subgroepen van items, tussen items en de volledige test, tussen verschillende deelconstructen in een test...
 - bv. nagaan of verschillende groepen van mensen verschillende presteren
 - **Relatie van de test met andere tests en externe variabelen**
 - externe variabele: ander construct waarmee je een samenhang of juist geen samenhang verwacht bv. schoolsucces, partnerkeuze, succes,
- de 4 zaken die we moeten onderzoeken van een test i.v.m. de validiteit

Verskillende vormen van validiteit

Constructvaliditeit

- de mate waarin een testscore een concrete weergave is van wat men wenst te meten bv. intelligentie en niet schoolse kennis
 - het gaat om het vinden van de aanwijzingen die aantonen dat de test werkelijk het meetinstrument is van datgene wat de test bedoelt te meten
 - Probleem:
 - betekenis van een begrip ligt niet altijd vast
 - betekenis kan verband houden met andere zaken
 - andere termen en begrippen kunnen een min of meer gelijke betekenislading dekken
 - grote overlap tussen verschillende termen en begrippen
- ⇒ meerdere operationalisaties mogelijk. Meestal is geen enkele volledig.
- **Meet het volledige construct en niets anders**
 - **De analogie met de rechtspraak** (nut van dit te herhalen)
 - De waarheid, de volle waarheid, niets anders dan de waarheid
 - Omdat je zaken kan vertellen over iemand anders – niet relevant
 - Jury kan verkeerde beslissingen nemen
 - Deze situatie transfereren naar het construct = de waarheid
 - In termen van het construct
 - Het construct, het volledige construct, niets dan het construct
 - Te maken met het juist gaan operationaliseren
 - Constructen zijn niet wederzijds uitsluitend
 - Ze bestaan in een net van overlappende betekenissen
 - De betekenis van een construct
 - relatie (verschil/gelijkenis) aangeven met gelijkaardige constructen
 - met welke constructen gaat het? samenhang / geen samenhang
 - Constructvaliditeit:
 - **Convergente validiteit**: mate waarin twee begrippen, die theoretisch met elkaar zijn verbonden, ook in de werkelijkheid met elkaar samenhangen of correleren
 - **Discriminante validiteit**: mate waarin twee begrippen, die theoretisch geen relatie hebben, ook in de werkelijkheid geen correlatie vertonen
 - **Constructvaliditeit: convergente validiteit**
 - Metingen van gelijkaardige constructen moeten hoog correleren

- Theorie



- Je theoretiseert dat alle tests hetzelfde meten
- Correlaties moeten het bewijs leveren

- **Constructvaliditeit: congruente validiteit**

- **Congruent**

- gelijkheid in vorm

- **Congruente validiteit**

- correlatie tussen twee tests die precies hetzelfde onderwerp meten met gelijke methode

- **Speciale situatie van convergente validiteit**

- Vb. *Amsterdamse Biografische Vragenlijst* (ABV) (schaal neuroticisme) versus *Eysenck Personality Inventory* (schaal neuroticisme)

- **Constructvaliditeit: discriminante validiteit**

- Nagaan in welk opzicht 2 testen verschillend zijn
- Je theoretiseert dat je 2 verschillende constructs hebt
- Deze correlaties moeten het bewijs leveren dat de items uit de 2 tests discrimineren
- Locus of control construct
 - Locus of control : gebeurtenis => gevolg toeschrijven aan
 - Jezelf = intern
 - Anderen = extern

- **Constructvaliditeit: Multitrait-Multimethod matrix**

Constructvaliditeit nagaan. Een correlatiematrix om construct na te gaan. Nieuw aspect : houd rekening met methode

- Wat is de MTMM Matrix?

- Een benadering ontwikkeld door Campbell, D. en Fiske, D. (1959). *Convergent and discriminant validation by the multitrait-multimethod matrix*. 56, 2, 81-105.
- Een matrix (tabel) van correlaties zodanig opgesteld dat de construct validiteit van een instrument kan worden onderzocht
- Integreert zowel convergente als discriminante validiteit
- Neemt aan dat je meerdere kenmerken meet met meerdere methodes

- Principes uit de MTMM matrix

- Convergentie: Dingen die theoretisch moeten samengaan gaan ook samen bij meting
- Divergentie/Discriminatie: Dingen die theoretisch niet mogen samengaan gaan ook niet samen bij meting

- Validiteit en betrouwbaarheid in de MTMM-matrix

	Constructen	Methode	Gewenst
Betrouwbaarheid	=	=	hoog
Convergente validiteit	=	≠	hoog
Discriminerende validiteit	≠	=	laag
Nonsens	≠	≠	laag

- | | | | | | | | | | |
|-----|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| | (.89) | | | | | | | | |
| .51 | | (.89) | | | | | | | |
| .38 | .37 | | (.76) | | | | | | |
| .57 | .22 | .09 | | (.93) | | | | | |
| .22 | .57 | .10 | .68 | | (.94) | | | | |
| .11 | .11 | .46 | .59 | .58 | | (.84) | | | |
| .56 | .22 | .11 | .67 | .42 | .33 | | (.94) | | |
| .23 | .58 | .12 | .43 | .66 | .34 | .67 | | (.92) | |
| .11 | .11 | .45 | .34 | .32 | .58 | .58 | .60 | | (.85) |
- Betrouwbaarheidsdiagonaal
 - Convergente validiteit
 - Discriminante validiteit
 - Betrouwbaarheid moet de hoogste coëfficiënten bevatten
 - Convergente validiteitsdiagonalen moeten hoge waarden (r') vertonen
 - De overige getallen = nonsenscorrelatie => moeten laag zijn: geeft argumenten voor discriminante val.

• Voordelen

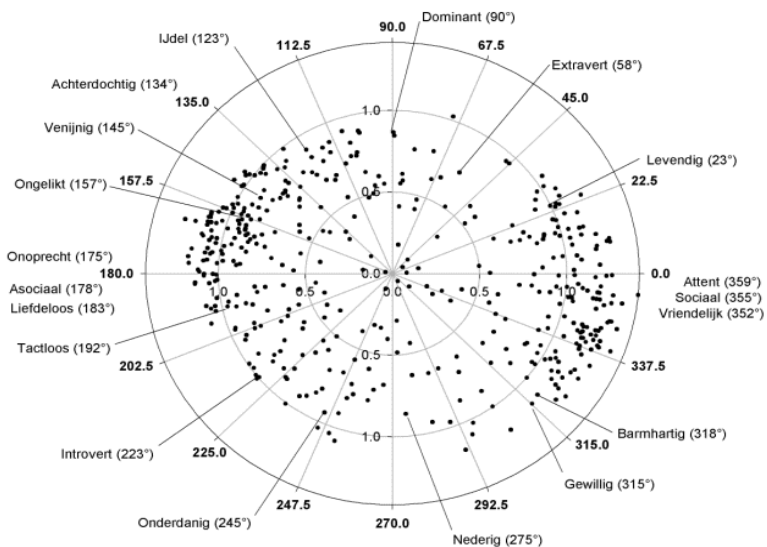
- Beschouwt simultaan convergente en discriminante validiteit
- Benadrukt het belang van de methode bij onderzoek en meten (methode als foutenbron)

• Nadelen

- Moeilijk om praktisch volledig uit te werken
- Geen algemene validiteitscoëfficiënt beschikbaar

Inhoudsvaliditeit p217

- **Mate waarin een itemverzameling representatief is voor een (specifiek) omschreven domein**
 - Volledig en omvattend het construct in kaart brengen
 - **Inhoud** verwijst naar de volledigheid waarmee iets wordt gemeten
 - **Representativiteit** van de test
 - analoog aan de relatie steekproef-populatie
 - mate waarin de inhoud van een test een representatieve steekproef vormt uit het geheel aan kennis, vaardigheid of gedrag dat de test beoogt te meten
 - itemverzameling moet alle niveau's dekken met goede vragen
 - zowel de hoogste en laagste regionen van je construct (vb. IQ test : zowel hoog als laag begaafd meten)
- ⇒ mate waarin de test volledig en omvattend het construct in kaart brengt
 - **Toepassingen** op werkgebied
 - Er wordt meer nadruk gelegd op de inhoudsvaliditeit bij testen die bepaalde vaardigheden en kennis toetsen (bv. arbeid of onderwijscontext)
 - **Bepaling** van de inhoudsvaliditeit: kwalitatief
 - is vooral een werk van **oordelen** en zorgvuldig construeren en niet van rekenen
 - bij de bepaling van de inhoud van een test kan een theoretisch model zeer nuttig zijn



Vb. van interpersoonlijk gedrag

2-dimensionaal

Linguïstisch paradigma:
Als je iets wil zeggen over
persoonlijkheid => gebruik
maken van taal (woorden)

Interpersoonlijk gedrag : 450
woorden

- **Verschil inhoudsvaliditeit en constructvaliditeit?**

inhoudsvaliditeit richt zich alleen op de volledigheid waarmee een test een construct meet, zonder enige uitspraak te doen over wat het construct precies is

- **Inhoudsvaliditeit**

- gaat uit van de rationale van waaruit de test is ontstaan, uitgaande van de kennis over het betreffende domein

- **Constructvaliditeit**

- aanwijzingen die aantonen dat de test verband houdt met het construct dat deze beoogt te meten

Criteriumvaliditeit p218

Criterium = waarin je geïnteresseerd bent, wat je wil voorspellen,...

Test = hoe je dit criterium gaat onderzoeken

Hangen prestaties op een test samen met prestaties in real-live (externe) situaties?

Twee methoden

In hoeverre komen testresultaten overeen met criteriumgegevens die op die moment voor handen zijn?

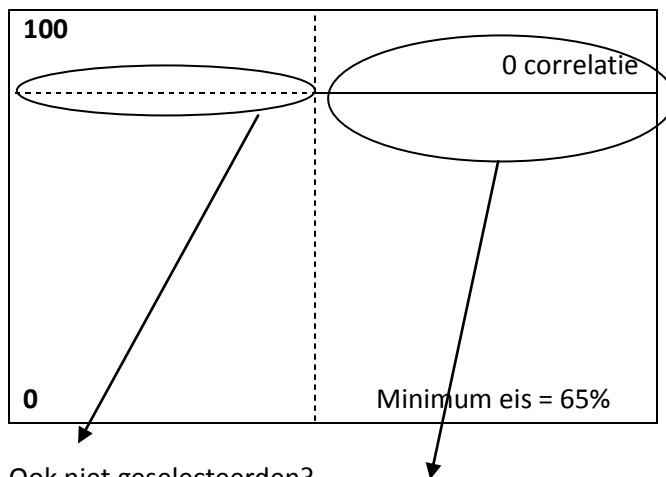
- **Concurrent validiteit**

- criterium wordt op zelfde moment gemeten als testprestatie (predictor)
 - vb1 correlatie tussen testresultaten werknemers en prestatie op werkvloer
 - vb2 correlatie tussen testresultaten leerlingen en oordeel leraar
 - Nut?
 - Wat is het nut van een extra test als je het criterium al hebt?
 - ✓ Het is niet altijd goed om enkel op een criterium voort te gaan, want het is veel subjectiever en minder gestandaardiseerd
 - ✓ Test = efficiënter

- **Predictieve validiteit**

- criterium wordt op later moment gemeten als testprestatie (predictor)

- vb1 toelatingsproef arts met als doel selectie van studenten met grotere slaagkans. Verwachting: hoge cijfers op toegangsproof → hogere cijfers op examen → hoger slagingspercentage
- vb2 screening selectiekandidaten voor bepaalde job
- Vb3 screening patiënten met hoge succesratio voor bepaalde therapie
- Toelatingsproef prestatie versus eindresultaat
- Absolute eis = een hoog positieve correlatie (zelden aanwezig, zelden onderzocht vooraf)
- Correlatie tussen beide metingen op *alle personen die afstuderen* is onvoldoende



Samenvattend

- **Validiteit: begrijpen en verklaren van samenhangen tussen gedragingen** (p 218 Box 6.8)
- De vraag of we met een valide meting te maken hebben van gedragingen, cognitie, motivatie en emotie, beantwoorden we meestal door een vermoeden uit te spreken **wat een construct** is, bijv angst.
- We maken er items (vragen) bij en analyseren deze.
- Daarna zeggen we bijv.
 - dat we iets over angst weten (constructvaliditeit),
 - dat we geen belangrijke dingen zijn vergeten (inhoudsvaliditeit),
 - en dat we er iets aan hebben (criteriumvaliditeit) want we kunnen er zinvol mee voorspellen, bijv iemands gedrag bij vreemden

Problemen bij criteriumvaliditeit

Criteriumvaliditeit: streven om bij selectie en plaatsing zo weinig mogelijk fouten te maken

- Echter enkele **problemen** bij criteriumvaliditeit
 - **Soorten criteria:** Het ene criterium is moeilijker te voorspellen dan het andere
 - Specifiek versus globaal criterium
 - Onmiddellijk - Intermediate – ultimate criterium
 - **Betrouwbaarheid van het criterium:** meting van criteriumprestaties in de praktijk
 - Criteria worden vaak minder betrouwbaar gemeten dan constructen
 - Risico's aan menselijk oordeel (oordeel **expert**) als criterium
 - **Test-criterium (cor)relaties**
 - Correlaties zijn gevoelig voor de omvang en variantie van de steekproef (restriction of range)
 - Testcriteriumrelatie wordt berekend door twee continue variabelen.
 - **Face validity**
 - ~ “liefde op het eerste gezicht”

- Heeft betrekking op wat een test op het eerste gezicht meet, mn correlatie tussen bepaalde predictor en criterium maar hiervoor weinig empirische steun

De test-criterium correlaties zijn gevoelig voor de omvang en variantie van de steekproef ('restriction of range') Voorbeeld: Probleem als gevolg van een kleine homogene steekproef.

Vuistregels voor beoordeling validiteit

Validiteitscoëfficiënt

- **Belangrijk om interpretaties te maken obv testcores**
 - correlatie tussen de testcore en één of ander extern criterium
 - tussen 1 en -1. Hoe dichterbij de waarde 1, hoe beter
- **Hoe hoog moet een validiteitscoëfficiënt zijn?**
 - Hangt af van de onderzoeksvraag
- **Vuistregels Cohen (1977)**
 - $r = .10 \rightarrow$ laag
 - $r = .30 \rightarrow$ gemiddeld
 - $r = .50 \rightarrow$ hoog

Predictieve validiteit en interpretatie

- **Meta analyse van Meyer (2001)**
 - 125 meta-studies: criteriumvaliditeit van psychologische en medische meetinstrumenten (predictieve tests en criteria)
 - 800 steekproeven
 - 144 correlaties
- **Conclusie**
 - Gemiddelde predictie: $r = .32$
 - Standaard deviatie: $.19$
 - Zelden boven $r = .50$
- **Meyer et al. (2001): Vuistregel ~ regels van Cohen**
 - Correlaties tussen $.00$ en $.15 \rightarrow$ verwaarloosbaar
 - Correlaties tussen $.15$ en $.31 \rightarrow$ klein, gering
 - Correlaties tussen $.32$ en $.50 \rightarrow$ gemiddeld, bescheiden
 - Correlaties groter dan $.50 \rightarrow$ hoog

Predictieve validiteitscoëfficiënten in vergelijking met

- **Betrouwbaarheidscoëfficiënten?**
 - Wenselijke minimummaat: $.70$ à $.80$ (zie vuistregels)
 - Mate dat test gelijk is aan zichzelf
- **Absolute grens van correlatie 1?**
 - predictieve validiteitscoëfficiënt zegt iets over de mate van invloed die een aspect heeft op een ander
 - waarde 1 zal nooit bereikt worden in validiteitsonderzoek
- **Convergente validiteitscoëfficiënten (monomethode coëfficiënten)?**
 - coëfficiënten worden verkregen door twee tests die hetzelfde kenmerk meten en dit doen op dezelfde manier, met elkaar te correleren
 - waarden liggen hier meestal op 0.50 of hoger
 - Verhoogde correlatie door gedeelde methode én kenmerk
- **Waarom zoveel lage validiteitscoëfficiënten?**
 - Lage criterium betrouwbaarheid
 - Restriction of range

6. Het nemen van beslissingen ogv tests

Klinische versus statistische predictie

Na diagnostisch proces

- Klinisch of statistisch beslissen?

Oorsprong van de controverse (p 56)

Windelband (1894)

- **Nomothetische wetenschappen**
 - zoeken naar wetmatigheden in een groep van individuen
- **Idiografische wetenschappen**
 - begrijpen/beschrijven van het individu

Opvatting

- **Nomothetische werkwijze**
 - elk individu kan beschreven worden als elk ander individu
 - individuen zijn onderling uitwisselbaar
 - de kenmerken van een persoon, die hem uniek maken, zijn niet noodzakelijk om het gedrag te begrijpen of te voorspellen
- **Idiografische werkwijze**
 - elk individu is uniek
 - algemeen geldende (gedrags)wetmatigheden zijn onvoldoende om gedrag te beschrijven
 - unieke, individuele beschrijving van een persoon

Methoden

- **Nomothetische werkwijze**
 - Statistische analyses van groepsdata
 - vb. Big Five model. *Exploratie van algemene persoonlijkheidsdimensies*
 - Vergelijken van personen tegen de waarden van een groep
 - vb. standaardscores bij testing
- **Idiografische werkwijze**
 - Niveau van het individu
 - Geen vergelijking met anderen
 - Bv. (semi)-ongestructureerde interviews, observaties en gevalsstudies
- **In psychologische diagnostiek:**
 - verschil mbt
 - belang van instrumenten en methodes
 - wijze waarop de data worden geïntegreerd om te beslissen
- **Idiografisch → klinisch beslissen**
 - Persoon heeft unieke eigenschappen
 - Beslissing op grond van klinische kennis en ervaring
Antwoord: categorisch (ja, nee)
- **Nomothetisch → statistisch beslissen**
 - Voorkeur voor formele, goed gevalideerde tests,
 - Persoon is een combinatie van scores op tests
 - Informatie wordt geïnterpreteerd mbv normgegevens of vergeleken met een criterium en/of ingebracht in een beslissingsformule.
 - Antwoord: probabilistisch (...% kans dat...)

Verloop van de controverse (p 59)

1954: Paul Meehl

- Clinical versus Statistical Prediction: a Theoretical Analysis and a Review of the Evidence”
- 20 vergelijkende onderzoeken: ‘19 – 1’- conclusie van Meehl

1965: **Paul Meehl**

- 50% statistisch beter
- 2% klinisch beter
- 48% geen verschil

2000: **Grove en medewerkers**

- 47% statistisch beter
- 6% klinisch beter
- 47% geen verschil

Welk type predictie verdient de voorkeur?

- **Statistische predictie beter in:**
 - Schoolsucces
 - Fraude in bedrijven
 - Levensduur
 - Succes in militaire training
 - Hartinfarcten
 - Neuropsychologische aandoeningen
 - Voorwaardelijke invrijheidstelling
 - Afmaken van opleiding politieagent
 - Differentiatie neurose – psychose
 - Geweld
- Ondanks overweldigende bewijsvoering voor de statistische beslissingsmethode zijn er nog steeds clinici die dit in vraag stellen
- **Meehl (1998):**
 - *“When statistical and non-statistical predictions collide, as they often do, we rarely have a “compromise option” lying between them. Suppose the equation says Patient X will benefit from shock therapy, but the psychiatrist thinks not. Do we compromise by using half as many volts A.C. as the potential needed to induce a cerebral storm?”*
- **Conclusie:**
 - statistische predictie verdient de voorkeur, maar de twee methoden blijken regelmatig gelijkwaardig.
- Gebrek aan mathematisch, methodologische kennis en soms aversie tegen het gebruik van formele methoden
- Validiteitsillusie: het eigen oordeel wordt accurater ingeschat dan het werkelijk is.

Evaluatie van meetinstrumenten

- **Engelstalige tekst**
 - “Validity for Decisions: Criterion-Related Validity”
 - Toledo
 - Engelstalig hoofdstuk uit Murphy & Davidshofer (2005). Psychological Testing: Principles & Applications.
 - Inhoud:
 - Criterium validiteit (cfr les 14) pg 178- 185
 - Uitgezonderd p 182 -184 (‘Practical Objections’ tem. ‘advantages’)
 - Evaluatie van meetinstrumenten (cfr les 16) pg 190-195
- **Twee soorten van evaluatie**
 - Is de test OK? ➔ Instrumentele evaluatie

		Test zegt	
Werkelijkheid (criterium) zegt		Goed	Slecht
	Goed	Correcte beslissing (sensitiviteit)	Vals Negatieven
	Slecht	Vals Positieven	Correcte beslissing (specificiteit)

- **Beïnvloedende factoren**
 - **Base rate ligt vast**
 - **Andere cut-off** (vb. andere selectieratio)
 - Andere sensitiviteit
 - Andere specificiteit
 - Andere vals negatieven/vals positieven
 - **Andere kritische criteriumscore** (vb. lagere/hogere eisen)
 - Andere sensitiviteit
 - Andere specificiteit
 - Andere vals negatieven/vals positieven

Sensitiviteit en specificiteit: terminologie

- **Base rate** (toevalskans)
 - Basisverhouding: mate van voorkomen van een kenmerk in een random populatie.
 - In voorbeeld: verhouding van succesvolle kandidaten (in een job = criterium) tot het totale aantal (100) zonder het gebruik van een test.
Base rate = $76/100 = .76$ of 76%. De kans op een succesvolle kandidaat is 76%.
 - Belang van Base rates (p 212)
 - Hoge base rate (90% - 100%): zinloos om te testen. Elke kandidaat is goed.
 - Extreem lage base rate (witte raven): zinloos om te testen, zelfs niet met de allerhoogste validiteitscoëfficiënt.
- **Selectieratio**
 - Testselectieverhouding: Verhouding van het aantal testpositieven (= vraag) tot het totale aantal (100) (= aanbod) of % kandidaten dat wordt aangenomen
 - In voorbeeld:
$$\frac{B+D}{A+B+C+D}$$
selectieratio = $84/100 = .84$ of 84%. We gebruiken een cutoff van 6 zodat we de beste 84 personen selecteren (obv test).
- **Succesratio**
 - Succesverhouding: Verhouding in de positieve concordantiegroep (geschikte kandidaten) tot het aantal dat een positief testresultaat behaalde of % succesvolle kandidaten binnen de aangenomen groep
 - In voorbeeld:
$$\frac{B}{B+D}$$
De kans op job succes met gebruik van de test is dus $72/84 = 0.86$
- **Vals Negatieven (False Rejections)**
 - personen die voldoen aan het criterium, maar niet worden weerhouden door de test
- **Vals Positieven (False Acceptances)**
 - personen die geselecteerd worden met de test maar niet slagen voor het criterium (maw. niet voldoen in de praktijk)
- **Sensitiviteit van een test**
 - proportie personen die werkelijk goed zijn en die ook als zodanig door de test worden herkend
- **Specificiteit van een test**
 - proportie slechte kandidaten die ook als zodanig door de test werden herkend