

Hoofdstuk 4: Consequentie als basis voor testinterpretatie: waardedenimplicaties (elina margashvili)

1. Waarden eigen aan validiteit

Waardengeladenheid van theorieën: de vraag is niet langer of maar hoe waarden in rekening moeten worden gebracht

Als we aan assessment doen zijn we nooit waardenneutraal bezig, aan onze assessment aan de uitspraken, de interpretaties die we maken hangen waarde oordelen aan vast. Hoe moeten we daarmee omgaan met waardegeladenheid van assessment, interpretaties en theorieën die er achterzitten?

"Valid" en "value" afgeleid van zelfde Latijnse stam, "valere" wat betekent sterk zijn, en houden beiden verband met Franse "valoir" wat betekent "waard zijn"

Bij validiteit geven we wat de waarde is van het instrument maar de interpretatie, gebruik van instrument maar er hangen waardeconnotaties aan vast

Waarden als vertekening en als bron van betekenis

Waardevertekening (bias): "intrusion of ordinarily tacit extrascientific motives of beliefs into the fulfillment of scientific purposes" Intrusie van gewoonlijk stilzwijgende buitenwetenschappelijke motieven van overtuigingen in de vervulling van wetenschappelijke doeleinden. ⇒ bemoeilijkt het proces

Wanneer spreken we van waardevertekening? **impliciete stilzwijgende** overtuigingen hebben, motieven die impact gaan hebben op hoe we wetenschap gaan doen

Maar waarden ook intrinsiek relevant, vb beroepsethiek

Wanneer +/-: "Values make for bias, not when they dictate problems, but when they prejudge solutions" (Kaplan) Waarden zorgen voor vertekening, niet wanneer ze problemen dicteren, maar wanneer ze vooruitlopen op oplossingen. Als we bepaalde antwoorden niet willen erkennen leiden ze tot bias

Conclusie: Hoe mee omgaan?: expliciteren en meenemen in ganse valideringsproces; expliciet aan de orde stellen in het onderzoek

Waar kunnen waarden overal naar voren komen? Bronnen van waardenimplicaties bij metingen en constructen

De waarde van constructnamen: waarden kunnen in een laag niveau naar voor komen in hoe je je construct benoemt

vb test met cognitieve taken waarvoor je voor verschillende mentale processen kan gebruiken en je noemt resultaat of de score ga je interpreteren itv

"Flexibiliteit vs rigiditeit" : het is een waardegeladen interpretatie, wanneer je flexibiliteit gebruikt dan zeg je dat is positief en rigiditeit heeft een negatieve connotatie; hoe meer iemand aan rigiditeits kant zit heb je allerlei nadelen, mensen zijn rigide moeilijker oplossing vinden, moeilijk vinden om nieuwe situaties om te gaan,...

"Verwarring vs consistentie": beschreven iets gelijkaardigs, wat aan de ene kant van dimensie zit gaat niet zelfde mentale processen gebruiken en aan de andere kant wel consistentie meten van mentale processen; maar door verwarring vs consistentie te noemen geef je heel andere implicatie, als iemand

aan de verwarrings kant zit dat negatief is (dimensie) wijst op gebrek aan tekorten = probleem en wie aan de consistentie kant zit positief is, consistentie als benaming is inherent positief geconnoteerd; dit moet je aantonen dat mensen verwarrings kant zitten dat die het net moeilijker hebben

"Geïnhibeerd vs impulsief" beide uiteinde van dimensies zijn negatief: vanuit persoonlijkheidspsychologie je stelt mensen aantal vrage en zeggen mijn dimensie is geïnhibeerd dus impulsief, waar zal iemand die de "gezonde" persoonlijkheidskenmerk hoe zou die moeten scoren op de dimensies. Je zou zeggen dat iemand die gezond is is eerder geïnhibeerd maar geïnhibeerd heeft al negatieve connotatie dus als je het zo verwoordt dan zou je eigenlijk moeten verwachten dat mensen die normaal gezond functioneren meer in het midden zitten van die schaal, die zijn niet geïnhibeerd en niet impulsief

"zelfcontrole vs expressiviteit" descriptief zit daar een grote overlap tussen, beide polen zijn verschillen maar positief, altijd positief gewoon anders

"Stress" vs "uitdaging" zelfde construct maar stress heeft negatieve connotatie en uitdaging positief, "stress" kan zijn dat die persoon niet slaagt "uitdaging" geeft de kans om verder te ontwikkelen en te slagen en verder gaan in het leven

Waarden komen onmiddellijk om de hoek hoe we onze constructen gaan benoemen; neutrale namen bestaan er niet

Belangrijk: streven naar trekimplicaties en evaluatieve implicaties van de naam, of uitklaren van conflicterende waardenimplicaties

dat namen overeenkomen met ons construct

vb ontwikkeling van een nieuwe schaal van Inschattingen voor morele distress, mag je die resultaten van items "ik word dingen te doen waarvan ik vind dat ze niet horen op mijn werk" Het zijn maar cognitieve items, mag je over distress spreken? het zijn items die te maken hebben met interpretatie van werkomgeving. Antwoord: ja. items zijn puur cognitief die verwijzen alleen naar in welke maten vind je dat je werksituatie, dat er morele aspecten. Blijkt als je die vragen afneemt en mensen scoren hoger op hun werkplek blijken ze hoger te scoren op burn-out, intentie omval,

minder plezier in hun werk. Cognitieve items impliceren stress, impliceren negatieve uitkomst, we mogen ze moreel distress upreasels noemen en niet moreel upreasels

Constructen verwijzen zowel naar mogelijke als actuele relaties

De benaming impliceert de toekomstige resultaten en niet enkel wat je verwacht van je construct

Naam geven aan je construct is gewoon een beginfase

Moet zorgvuldig gebeuren want door naam op te plakken veronderstel je heel veel dingen, ook nadenken wat de implicaties zijn van je naamgeving

vb als je denkt tis een positief kenmerk is van een persoon een "deugd" dan verwacht je allerlei positieve effecten van dat construct, het gaat niet enkel over het positief effect dat je NU hebt gevonden maar ook in mogelijke toekomstige onderzoeken

De waarde van theorieën: onze theorieën impliceren ook bepaalde waarden

Ook bredere theorieën bepalen hoe observaties worden geïnterpreteerd: theorieën gebruiken om onze scores te interpreteren dus waarden die vasthangen aan theorie kunnen impact hebben op onze score interpretatie

Voorbeeld: intellectuele vaardigheid:

Oorsprong en ontwikkeling

Beperkingen en mogelijkheden van leren

Favoriete onderzoeksbenaderingen: als je dit meet heb je twee tegengestelde benaderingen, ene mens is onbeschreven blad en vanaf de geboorte wordt door allerlei ervaringen ingevuld, de andere de mens is biologisch wezen, tis genetisch gegeven; twee benaderingen impliceren verschillende waarden; nurture vs nature

Impliciete waardeovertuigingen voer aard van mens-zijn en maakbaarheid van de mens; ideologische implicaties

De waarden van ideologieën

Ook ideologieën geven perspectief en richting

Ideologie = een complexe configuratie van gedeelde waarden, affecten en overtuigingen die een interpretatiekader van de wereld bieden

Waarden hebben is niet verkeerd, wat verkeerd is dat we ons daar niet bewust van zijn

vb institutioneel-universalistisch paradigma vs participatorisch-selectivistisch paradigma. Sluit aan moeten we ijkingstoetsen of niet

institutioneel-universalistisch paradigma: kennis is universeel gegeven, op systematische wijze kennis kan aanleren meer dan sommige mensen, met toegangsproeven organiseren, zoveel mogelijk onderwijs voor zoveel mogelijk mensen

participatorisch-selectivistisch paradigma: gaat over ontwikkeling van individuele mensen en elk persoon op zich uniek en eigen ontwikkelingspad, en onderwijs zo vormen dat die persoon zijn eigen persoon wordt

Methodologieën voor identificeren van waarden en waardenvertekeningen

Expliciteren in onze diagnostisch en wetenschappelijk werk

Hoe subtiele impact van namen, theorieën en ideologieën op het spoor komen?

Door Kantiaanse, Hegeliaanse en Singerianse benaderingen te gebruiken en deze verschillende benaderingen in verband brengen met elkaar en te onderzoeken

Tegengestelde verklaringen met elkaar confronteren en kijken waar beide verwijzen (Kantiaans en Hegeliaans)

⇒ niet consensus maar conflict: we gaan kijken vanuit conflictbenadering kijken, wie is het niet eens met mij itv waar we naartoe moeten

Ook waardenimplicaties moeten empirisch ondersteund en rationeel onderbouwd worden

Messick daarom vind dat waardeimplicaties wel degelijk tot begrip validiteit behoren is dat waardeimplicaties empirisch moeten worden ondersteund; dus wanneer de implicaties is dat je

iets aan het meten zou zijn wel biologisch wordt bepaald, dat moet je gewoon moeten kunnen aantonen

Dit is probleem want we hebben evidentie dat genetische componentie van intelligentie maar we weten niet om welke genen het gaat. We kunnen obv intelligentiescore geen uitspraak doen over de vraag 'in welke mate is van dit niveau van cognitieve ontwikkeling van persoon doet het toe aan de genen of omgeving' = geen evidentie voor!

We moeten wel waarde implicaties aantonen en welke mate kunnen we die hard maken

2. Evidentie als basis voor testgebruik: construct validiteit + relevantie + utiliteit

Welke evidentie hebben we nodig om test te kunnen gebruiken in de praktijk?

Construct validiteit: oordeel over relevantie, utiliteit en passendheid alleen in licht van scorebetekenis:

Zelfs bij diplomering of voorspellen van jobsucces: achterliggende theorie van curriculum of van job

Correlatie (correlatie tussen psychologisch test instrument en toekomstig criterium is op zich onvoldoende op te gebruiken) zonder score-interpretatie op zich onvoldoende: vb predictie van schoolsucces op basis van prestatiemotivatie of volgzzaamheid ipv vaardigheden en inzicht

vb schoolsucces voorspellen: test meet hoe conformistisch leerlingen zijn en mss voorspellen hoe die leerlingen gaan doen op school = dit is niet de bedoeling! weten wat die test meet en waarom zou die voorspellen, voorspelt voor goede reden

vb selectieproef op psychologie en we steken in die proef wiskundige vragen en ruimtelijke vragen (taakjes), ijkingsstoetsen zoals bij ingenieurs wetenschappen, zou dat werken voor psycholoog? studenten die meer scoren zal meer kans hebben om te slagen, ja dit gaat werken, als we de proef hier toepassen gaan we correlatie hebben van slaagkansen na eerste bach. Intelligentie subtests zijn positief gecorreleerd, met ruimtelijke testen heb je een stukje te pakken van algemene intelligentie MAAR slecht idee om dat

te doen want zal negatieve directe consequenties daarvan zijn, proportie van jongens en meisjes zal helemaal anders liggen daarom heeft het te maken met consequenties met u testgebruik.

Het feit dat je een correlatie vindt tussen test en een criterium is onvoldoende! weten waarom dat voorspellend is

⇒ we mogen alleen testen gebruiken waarvan we weten wat ze meten, waarvan we dus de scoreinterpretatie mooi wetenschappelijk kunnen beargumenteren, en als we niet kunnen aantonende dat scores op bepaalde manier interpreteren dan moeten we ons vraag stellen of de test toe te passe in de praktijk vb we kunnen die test gebruiken in onderwijs, mensen aandvies geven over een opleiding, werkcontext evalueren welke trainingen ze kunnen gebruiken en hoe ze zich kunnen verbeteren of in klinische context welke problematiek waarmee persoon voorstelt

Als we iets willen oorspellen dan hebben we een beeld een hypothese waaruit dat bestaat

Relevantie en representativiteit

De inhoud of processen van de testen aantonen dat ze relevant zijn vb ijkingstoets van ingenieurs toepassen in psychologie, wiskunde is relevant maar ruimtelijk inzicht is niet relevant, dus we hebben irrelevantie in ijkingstest

Ten overstaan van

Curriculum /onderwezen stof

Prestatie: kennis als resultaat van instructie (examen, inhoud van de lessenreeks is bepalend) (belangrijk hoe kennis en vaardigheid tot stand is gekomen).

Prestatie verwijst je naar opleiding

Competentie: kennis en vaardigheid in inhoudelijk gebied ongeacht hoe ze tot stand gekomen is (niet belangrijk hoe kennis en vaardigheid tot stand gekomen is)

Competentie verwijst je naar een domein

Job

Test-criterium relaties zijn ook belangrijk

We moeten kijken naar het kwaliteit van criterium

Er kunnen twee problemen zijn met criterium

Ook betrouwbaarheid van de criteriummeting

niet betrouwbaar gemeten vb in jobcontext wanneer gekeken wordt naar voorspellende kenmerken van jobprestaties, veel metingen zijn niet erg betrouwbaar

vb aan leerkrachten vragen hoe goed leerling doet op testen algemene beoordeling, niet zo betrouwbaar .60

Ook constructvaliditeit van criteriummeting: er kunnen 2 problemen zijn bij

Constructirrelevantie (meet ook andere dingen die niet relevant zijn aan het construct): criteriumcontaminatie

vb je wilt test ontwikkelen vb van hoe goed leerling doet op school en je criterium wordt gemeten als het oordeel van de leerkracht, beoordeling proberen voorspellen. Leerkracht kan rekening houden met de relatie met leerling, als goed klikt dan heeft de leerling meer kans om goed beoordeeld te worden ook itv zijn academisch presteren, ook bij jobs

Constructonderrepresentaite: criteriumdeficiëntie

dat je criterium te weinig zit

vb discussie over dat top ceo's teveel verdienen omdat ze uitzonderlijke kwaliteiten hebben, maar hoe kan je dat dan aantonen? ⇒ moeilijk om aan te tonen omdat er teveel criteria is om te voorspellen

Je wordt hier geconfronteerd met criterium

Test-criterium relaties meestal georiënteerd op convergerende verbanden, zelden op *discriminante evidentie* (andere constructen, methodevariantie) ⇒ dit is het probleem met test-criterium correlaties

We moeten ook kijken naar verbanden die niet mogen zijn (divergente verbanden) mogen ook niet vaststellen maar men kijkt dikwijls naar convergerende evidentie ipv divergente evidentie

Selectie- versus classificatiebeslissingen

Selectiebeslissing: ja/nee beslissing

Beste persoon kiezen voor job positie uit beste kandidaten

Classificatiebeslissing : toewijzing aan behandeling

Rekrutering bij leger, iedereen werd toegewezen aan bepaalde positie. Hoe kan je efficiënt taken in leger verdelen dat doenbaar is in het leger?

Concurrente vs toekomstige criterie: kijken naar hoe goed voorspeld schoolsfunctioneren nu en hoe kind zal doen binnen 12 jaar

Beslissingsutiliteit

Niet alleen sterkte van **test-criterium** relatie maar ook over bruikbaarheid

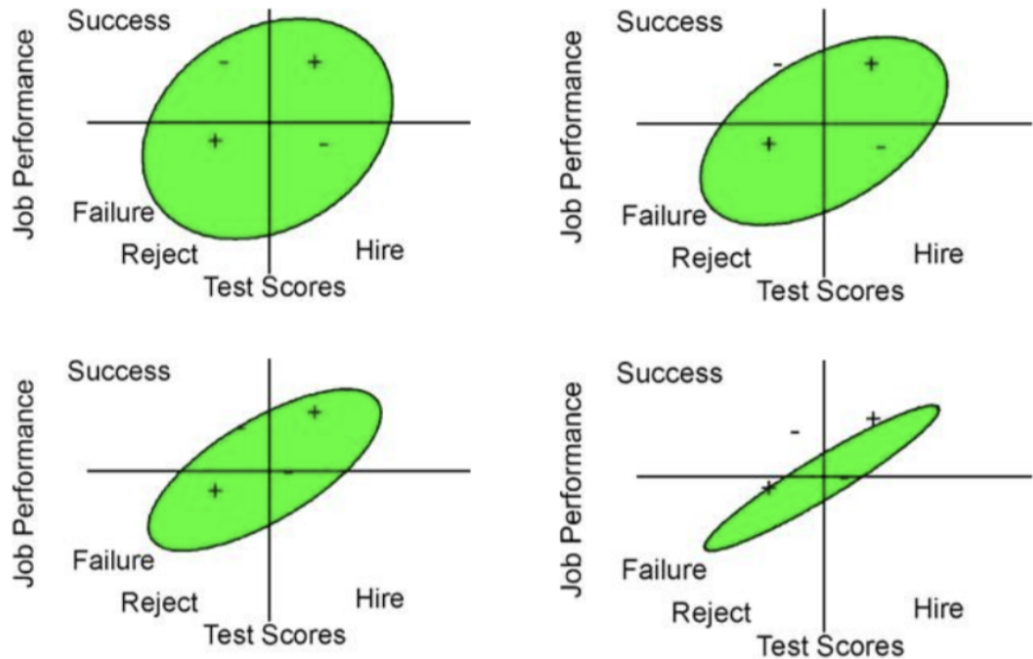
Base rate: proportie individuen die potentieel zouden slagen in programma of in job

Selectie ratio: proportie individuen die boven de predictor cutoff scoren

Kosten die de testing met zich mee brengt

Rekening houden met fairheid!

Voorbeeld: jobperformance (failure en succes) en testcores (reject en hire)



Er zijn 2 type fouten bij selectie: we kunnen onterecht iemand aanwerven en iemand onterecht niet aanwerven

4 situaties verschillen van in testcriterium correlaties: hoe sterk hangt de test met criterium samen?

- 1) lage correlatie: we gaan ten onrechte veel mensen weigeren en ten onrechte veel mensen aanwerven (linksboven)
- 2) hoge correlatie: ten onrechte veel mensen weigeren en ten onrechte veel mensen aanwerven (rechtsboven)

Base rate: wat is de kans dat je job goed of niet goed gaat doen is ongeveer 50/50

Selectieratio: de kans dat je geselecteerd word obv van die test is ook 50/50

Taylor-Russell tabel:

Base Rate of Success (BRS)	Level of Validity	Selection Ratio (SR)										
		.05	.10	.20	.30	.40	.50	.60	.70	.80	.90	.95
.30	r	.05	.10	.20	.30	.40	.50	.60	.70	.80	.90	.95
	.00	.30	.30	.30	.30	.30	.30	.30	.30	.30	.30	.30
	.25	.50	.47	.43	.41	.39	.37	.36	.34	.33	.32	.31
	.50	.72	.65	.58	.52	.48	.44	.41	.38	.35	.33	.31
	.75	.93	.86	.76	.67	.59	.52	.47	.42	.37	.33	.32
.50	r	.05	.10	.20	.30	.40	.50	.60	.70	.80	.90	.95
	.00	.50	.50	.50	.50	.50	.50	.50	.50	.50	.50	.50
	.25	.70	.67	.64	.62	.60	.58	.56	.55	.54	.52	.51
	.50	.88	.84	.78	.74	.70	.67	.63	.60	.57	.54	.52
	.75	.99	.97	.92	.87	.82	.77	.72	.66	.61	.55	.53
.70	r	.05	.10	.20	.30	.40	.50	.60	.70	.80	.90	.95
	.00	.70	.70	.70	.70	.70	.70	.70	.70	.70	.70	.70
	.25	.86	.84	.81	.80	.78	.77	.76	.75	.73	.72	.71
	.50	.96	.94	.91	.89	.87	.84	.82	.80	.77	.74	.72
	.75	1.00	1.00	.98	.97	.95	.92	.89	.86	.81	.76	.73

Source: Adapted from H. C. Taylor and J. T. Russell, "The Relationship of Validity Coefficients to the Practical Effectiveness of Tests in Selection," *Journal of Applied Psychology*, Vol. 23, 1939, pp. 565-578.

Base rate: Wat gebeurt er wanneer je een scenario hebt waarbij je een lage base rate hebt dus geen 50% maar 30% of base rate gaan variëren hoeveel % van je populatie zou die job goed doen 30% 50%, 70%

.30: iedereen aanwerven en je hebt 30% kans dat persoon goed functioneert op werk

.50: 50% kans dat iemand zou goed functioneren

.70: 70% kans

Selectie ratio: hoeveel % van je kandidaten ga je selecteren voor de job van 5% naar vrijwel iedereen 95%

Test-criteriumcorrelatie: van 0 geen enkel verband tot .75 heel sterk verband

Goede selectieproef voor hoger onderwijs correleert .50 hoe men doet op hoger onderwijs, .50 is een realistisch verband

Wanneer je iedereen aanneemt 95% maakt het niet uit hoe goed je test is, hoe goed je criteriumcorrelatie is

Wanneer je test-criteriumcorrelatie 0 is maakt het niet uit hoe streng is, je komt altijd terug naar base rate dus je wint niets

Maar naarmate je test meer correleert met criterium en je mag heel streng selecteren, ga je enorme winst boeken dus als .70 van populatie zou slagen/job goed doen en je test correleert .70 met

jobprestatie en je neemt die test af en je neemt 5%
hoogscorende op die test dan ben je zeker dat iedereen die je
aanwerft de job goed zal doen

Als je base rate hebt van .50 en testcriterium correlatie .75 zit je
aan 99% en als je .50 hebt dan zit je aan 88%

Utiliteit is dus niet enkel kijken naar testcriteriumcorrelatie die moet
zo hoog mogelijk zijn maar ook kijken naar de base rate en selectie
ratio

vb Stanford hebben hoge selectie ratio

EXTRA: Wat kun je afleiden uit een Taylor-Russell tabel? base-rate,
de validiteit en de selectie ratio.

Geven inzicht in toename in het percentage goed functionerende
mensen bij gebruik van een tests met verschillende predictieve
validiteiten en verschillende selectie ratios. Baserate is het
proportie studenten wat succesvol is in de hele klas. Naarmate
de validiteit en selectie ratio (aantal studenten aangenomen)
verandert veranderd het percentage wat succesvol wordt. Met
een grote validiteit zijn er meer mensen die aangenomen worden
ook daadwerkelijk succesvol. En hoe kleiner de selectieratio
wordt (hoe minder mensen aangenomen) hoe belangrijker de
validiteit voor het percentage.

Validiteitsgeneralisatie

Wanneer we aan predictie doen, in welke mate is er een sprake van
validiteitsgereneralisatie?

Specifieke validiteit vs generaliserende validiteit: grote variatie in
criterium-predictorcorrelatie omwille van meetfouten en artefacten

Administratieve fouten

Te kleine steekproeven: binnen 1 bedrijf of 1 afdeling

Correlaties met niet-continue variabelen: beoordeling van de baas of
beoordeling slagen of niet slagen, voldoen die personen of niet

Beperking van bereik: als je onderzoek doet in heel homogene
steekproef dan ga je minder sterke verbanden vinden en dat hangt
ook bij onbetrouwbaarheid van het criterium vb beoordeling van de
baas is relatief onbetrouwbaar .60

Onbetrouwbaarheid van het criterium

Schmidt & Hunter: niet situatie, maar vooral meetfouten en artefacten ($r = .50$ tussen 'general mental abilities' en 'work sample tests')

Als je bovenstaande (validiteitsgeneralisatie) allemaal samenneemt = metaanalyses

Er is een duidelijke verband tussen intelligentiescores en jobprestaties voor heel veel van de verschillende jobs

Dus er is duidelijke sprake van validiteitsgeneraliseerbaarheid: niet enkel voor 1 type van job maar voor heel veel types

Belang van meta-analysestudies voor cumulatie van kennis maar probleem met confirmation bias

vb Als ik studie doe met nieuwe instrument die populaire is of heel veel aandacht trekt en ik vind geen verband dan de kans is groot dat ik die studie niet gepubliceerd krijg maar als ik verwachte verband wel vind dan is de kans groot dat ik die studie gepubliceerd krijg

We moeten oppassen wat we terug vinden in gepubliceerde artikels daar heb je de neiging van meer verwachte verbanden te vinden

Bij goede studies moet je naar niet gepubliceerde studies kijken maar ook naar doctoraten die niet gepubliceerd geweest zijn, zo heb je alle mogelijke resultaten

⇒ 3 kwadrantjes gedaan van Messinck

3. Consequenties als basis voor testgebruik: sociale consequenties

Anticiperen van neveneffecten en bijproducten van het testgebruik

Nadelige effecten het gevolg van constructierelevantie of constructierelevante testvariantie (dit moet je nagaan)

Dus niet alleen bedoelde doelstelling, maar geheel van effecten (bijwerkingen) moeten in rekening worden gebracht

vb meerkeuze prestatietoetsen kunnen leiden tot meer nadruk op geheugen en analytische vaardigheden dan op divergent denken en synthetischst vaardigheden (essayvragen)

proffen groeit niet jaar na jaar maar wel studenten groeien jaar na jaar

vb selectieproef voor psychologie gebruiken toegangspreef van ingenieurs wetenschappen

Onbedoelde effecten: meer mannen

Als dit gevolg zou zijn van selectieprof is dit zeer problematisch want we zullen mensen geselecteerd hebben ifv hun algemene intelligentie (goed) maar ook ifv hun ruimtelijke vaardigheden (slecht): mensen zijn ten onrechte geselecteerd want ruimtelijke vaardigheden heb je niet nodig om psychologie te studeren

Als meer jongens zijn in STEM richtingen krijgt dit maatschappelijke aandacht maar als er weinig jongens zijn in de psychologie of andere richtingen krijgt dit geen enkele maatschappelijke aandacht

Meer mannen dan vrouwen in psychologie: irrelevante info voor test

Voor ingenieurs wel relevant

Voorgesteld gebruik van testmethode afzetten tegen alternatieven

Tegenvoorstellen: gebruik van andere assessmentprocedures, hoe?

Sociale consequenties van gebruik vs niet-gebruik:

wat gebeurt er als we niets doen? moeilijk maatschappij te overtuigen dat ijkingsstoetsen een goed idee zou zijn want we weggooien veel talent

Sommige 3x niet slagen voor een opleiding en uiteindelijk geen opleiding hebben, en gaan werken, terwijl die beter georiënteerde dat die waarschijnlijk heeft afgewerkt en zonder maatschappelijk kost

in VS op alles getest

We moeten voor-en nadelen van testen bekijken, alles samenbrengen

Implicaties bekijken vanuit verschillende waarden-perspectieven

Afhankelijk van waarden die je hebt van bepaalde vorm van test die goed is of niet goed

vb distributieve rechtvaardigheid: wat zijn de criteria die je gebruikt om middelen te verdelen?

Het kan zijn obv vaardigheden, inzet, realisaties, billijkheid, noden, gelijke mogelijkheden, vraag en aanbod op de markt, principe van reciprociteit, algemeen belang, boven een bepaald minimum, gelijkheid

Welke distributief rechtvaardigheid gaan we gebruiken? expliciteren waarom gaan we het doen en niet doen, en wat zijn de argumenten

Discussie rond ijkingsproeven: 1) je moet naar verschillende dingen gaan kijken enerzijds naar wat is de ideologische basis daarvan, wat zijn de waarden die mensen belangrijk vinden, waarom al dan niet tegen zijn 2) en wat weten we daarover, welke positie neemt men in zo'n debat

We maken fouten, elk systeem maakt fouten, de vraag is hoe groot is die fout?

⇒ dit laatste was de vierde kwadrant van Messick

Evidentie voor interpretatie, testgebruik, consequenties van testinterpretaties dat zijn de waarden, en consequenties van testgebruik. We moeten naar alles kijken, geheel bepaald validiteit