

Aanvulling op het COTAN Beoordelingssysteem m.b.t. Fairness

Fairness (ook wel onpartijdigheid) betreft de toepasbaarheid van psychologische tests bij verschillende groepen, onafhankelijk van hoe die groepen gedefinieerd zijn. Persoonlijke kenmerken die niet van belang zijn voor de te meten eigenschap mogen geen invloed hebben op de testresultaten of op de interpretatie daarvan. Hoewel er binnen het *COTAN Beoordelingssysteem voor de kwaliteit van tests (2010)* aandacht is voor de vraag of tests betrouwbaar gebruikt kunnen worden bij verschillende (sub)groepen, vormt de beoordeling daarvan geen apart criterium. Fairness is geen geïsoleerd concept en moet worden beschouwd in alle aspecten van een test. De werkgroep 'Fairness' van de COTAN heeft zich gebogen over de vraag hoe de beoordeling van fairness beter zichtbaar kan worden gemaakt in het COTAN Beoordelingssysteem. Er is gekozen voor een gestructureerde beschrijving van onderzoek dat is gedaan naar fairness met behulp van de zogeheten 'fairness matrijs'.

Op basis van de ervaring die is opgedaan met het gebruik van de (eerste versie van de) fairness matrijs in de periode van 1 juli 2015 tot en met 1 juli 2021 is de fairness matrijs herzien, dat wil zeggen de beoordelingsvragen uit het COTAN Beoordelingssysteem die betrokken (kunnen) worden bij het opstellen van de fairness matrijs zijn opnieuw bekeken en gewijzigd. Bij de selectie van beoordelingsvragen is gekeken naar in hoeverre de vraag daadwerkelijk informatie geeft over fairness, naast de nodig geachte kenmerken van een test (bijvoorbeeld representativiteit van de normgroepen). Op basis daarvan is besloten om de volgende vragen niet meer mee te nemen bij het opstellen van de fairness matrijs:

- *Kwaliteit van het testmateriaal*
Vragen 2.6 en/of 2.14: "Zijn de items correct geformuleerd?"
- *Kwaliteit van de handleiding*
Vraag 3.3: "Wordt er informatie gegeven over de gebruiksmogelijkheden en beperkingen van de test?"
Vraag 3.4: "Wordt er in de handleiding een samenvatting van de onderzoeksresultaten gegeven?"
Vraag 3.5: "Wordt er met behulp van voorbeelden aangegeven hoe testcores kunnen worden geïnterpreteerd?"
Vraag 3.6: "Wordt gewezen op soorten informatie die bij de interpretatie van belang kunnen zijn?"
- *Normen*
Basisvraag 4.3.b: "Zijn de normgroepen representatief?"
- *Betrouwbaarheid*
Vraag 5.3.b: "Zijn de steekproeven op basis waarvan de betrouwbaarheidsgegevens zijn berekend, in overeenstemming met het beoogde testgebruik?"
- *Begripsvaliditeit*
Vraag 6.3.b: "Komen de steekproeven die in het begripsvalideringsonderzoek zijn gebruikt, overeen met groepen waarvoor de test is bedoeld?"
- *Criteriumvaliditeit*
Vraag 7.3.b: "Zijn de steekproeven op grond waarvan de criteriumvaliditeitsgegevens zijn berekend in overeenstemming met het beoogde testgebruik?"

De volgende beoordelingsvragen uit het COTAN Beoordelingssysteem worden meegenomen in de herziene fairness matrix, inclusief een explicitering van de relatie van de vraag met fairness:

- *Uitgangspunten van de testconstructie:*
Basisvraag 1.1b: “Is er aangegeven wat de doelgroep(en) is (zijn) van de test?” (*toegevoegd*)
Bij deze vraag is het van belang dat zo specifiek mogelijk wordt aangegeven voor wie de test bedoeld is (en voor wie niet) ter voorkoming dat de test gebruikt wordt bij personen of groepen voor wie deze niet of minder geschikt is en daarom mogelijk bias kan opleveren.
- *Kwaliteit van het testmateriaal:*
Basisvragen 2.3 en/of 2.11: “Zijn de items vrij van racistische, ethnocentrische, seksistische en voor bepaalde bevolkingsgroepen kwetsende inhoud?”
Het gaat bij deze vraag om de a priori bruikbaarheid van een test voor een bepaalde groep.
- *Normen:*
Vraag 4.6: “Worden er gegevens verstrekt over mogelijke verschillen tussen subgroepen (bijvoorbeeld alloctonen-autoctonen, vrouwen-mannen)?”
De resultaten van dit onderzoek kunnen helpen bij het vaststellen van een mogelijk discriminerend effect (of *adverse impact*) en kunnen een extra reden zijn voor het uitvoeren van test- en/of itembiasonderzoek.
- *Begripsvaliditeit:*
Vraag 6.2.c: “Maken de resultaten voldoende aannemelijk wat wordt gemeten op basis van de gegevens over de invariantie van de factorstructuur en mogelijke itembias?”
Met dergelijk onderzoek kan inzicht verkregen worden in de (un)fairness van de items en/of de test als geheel.
Vraag 6.2.e: “Maken de resultaten voldoende aannemelijk wat wordt gemeten op basis van de gegevens over verschillen tussen relevante groepen?”
Hoewel men bij dit type onderzoek juist graag verschillen wil zien, kan het zo zijn dat men wat anders vindt dan men verwacht (bijv. geen verschillen of andere verschillen dan verwacht). Dergelijke resultaten zouden een indicatie van unfairness door testbias kunnen zijn.
- *Criteriumvaliditeit:*
Vraag 7.2: “Zijn de resultaten voldoende, gelet op het beoogde type beslissingen dat met de test moet worden genomen?” (*toegevoegd*)
Bij deze vraag is het van belang dat, indien daar aanleiding toe is (bijvoorbeeld als subgroepen verschillen in gemiddelde scores of als uit onderzoek bij vergelijkbare tests is gebleken dat de voorspellende waarde per subgroep kan verschillen), onderzoek uitgevoerd is naar mogelijke predictiebias voor de betreffende groepen.

Op basis van de door de beoordelaars gegeven toelichting bij deze beoordelingsvragen zal per kenmerk waarop doelgroepen van elkaar kunnen verschillen (zoals geslacht en herkomst) het uitgevoerde onderzoek worden beschreven in de fairness matrix. De vaste kenmerken ‘geslacht’, ‘herkomst (etnisch-cultureel)’, ‘leeftijd’ en ‘beperkingen’ worden altijd weergegeven, ongeacht of er onderzoek naar gedaan is en/of aandacht aan besteed is.

Een voorbeeld van een ingevulde fairness matrijs, zoals deze weergegeven wordt in de COTAN Documentatie, is te zien op de laatste pagina. Hierin worden in de kolommen de kenmerken gegeven waarop doelgroepen van elkaar kunnen verschillen en in de rijen staan de criteria uit het COTAN Beoordelingssysteem genoemd waarover onderzoek met betrekking tot fairness gerapporteerd kan worden, zoals onderzoek naar vraagpartijdigheid (*itembias*) bij het criterium *Begripsvaliditeit*.

N.B. De fairness matrijzen die voor de ingangsdatum van deze herziene fairness matrijs (d.w.z. 14 juli 2021) zijn gepubliceerd in de COTAN Documentatie blijven ongewijzigd.

Voorbeeld van een ingevulde fairness matrix zoals weergegeven in de COTAN Documentatie

Fairness Matrix		Tonen ▾			
Aandacht besteed aan fairness?	Geslacht	Herkomst (etnisch-cultureel)	Leeftijd	Beperkingen	Regio
Uitgangspunten testconstructie	Ja i			Ja i	
Testmateriaal			Ja i		
Normen			Ja i		Ja i
Begripsvaliditeit	Ja i				
Criteriumvaliditeit					

Er is onderzoek uitgevoerd naar itembias. Een X aantal items laat bias zien op basis van 'geslacht'.

N.B. De beschrijving van het uitgevoerde onderzoek wordt zichtbaar door de cursor op de 'i' te houden.