

6.2: Criteria van de COTAN: informatie

Bron: Inleiding van de brochure “COTAN beoordelingssysteem voor de kwaliteit van tests” van Evers, Lucassen, Meijer & Sijtsma (2010, www.cotan.nl).

De Commissie Testaangelegenheden Nederland (COTAN) beoordeelt psychodiagnostische instrumenten die in Nederland worden uitgegeven en bij de COTAN ter beoordeling aangeboden.

Inhoud van het beoordelingssysteem

Hoewel de COTAN zich al sinds haar oprichting in 1959 bezighoudt met het beschrijven van tests, werden pas in de Documentatie van Tests en Testresearch van 1969 voor het eerst beoordelingen van tests gepubliceerd. Het gehanteerde systeem leidde tot een globaal oordeel over tests. De beoordeling kon variëren van A (zeer goede test) tot F (slechte test of test in ontwikkeling). Mede vanuit de behoefte een meer gedifferentieerd oordeel over een test mogelijk te maken, werd in de Documentatie van 1982 een geheel nieuw systeem gebruikt. Elke test werd beoordeeld op vijf criteria: 1) uitgangspunten van de testconstructie, 2) kwaliteit van het testmateriaal en de handleiding, 3) normen, 4) betrouwbaarheid en 5) validiteit.

Bij de voorbereiding van de Documentatie van Tests en Testresearch van 2000 werd het systeem herzien, waarbij de vragen, de toelichtingen en de wegingsvoorschriften werden aangepast. Het ging hierbij vooral om relatief kleine aanpassingen en verbeteringen. Een wezenlijke verandering betrof echter de splitsing van de criteria 2 en 5 in aparte beoordelingen voor kwaliteit van het testmateriaal, kwaliteit van de handleiding, begripsvaliditeit en criteriumvaliditeit. Zonder veel moeite konden destijds de beoordelingen op vijf criteria worden omgezet naar de meer gedifferentieerde beoordelingen op de zeven nieuwe criteria.

Bij de huidige herziening zijn de wijzigingen echter ingrijpender, hoewel het aantal van zeven criteria ongewijzigd is gebleven. Ook het principe van verscheidene vragen per criterium, waaronder een of meer basisvragen, is onveranderd gebleven. Nieuwe inzichten op het gebied van testtheorie en testconstructie hebben er echter toe geleid dat er vragen zijn toegevoegd, dat sommige vragen zijn onderverdeeld in subvragen en dat andere vragen zijn vervallen. Ook zijn er aanwijzingen gespecificeerd en scoringsregels veranderd. Als men een test volgens het oude en het nieuwe systeem zou beoordelen, kan dat uiteraard tot verschillen in beoordeling leiden. Vanaf april 2009 worden tests volgens het nieuwe systeem beoordeeld. Er zullen geen herbeoordelingen plaatsvinden van de circa 600 tests die met de ‘oude’ versie zijn beoordeeld. Alleen met betrekking tot de veroudering van normgegevens wordt hierop een uitzondering gemaakt.

De zeven criteria

Hieronder wordt een globale omschrijving van het herziene beoordelingssysteem gegeven, inclusief de belangrijkste wijzigingen. De zeven criteria zijn:

1 Uitgangspunten van de testconstructie

Dit criterium wordt beoordeeld door middel van drie vragen, waarmee achtereenvolgens wordt vastgesteld of het gebruiksdoel, de theoretische achtergrond en de operationalisatie daarvan in de testinhoud zijn beschreven. De beoordeling op dit criterium is van invloed op de waardering van andere criteria, omdat de meetpretentie bepaalt welk type normerings-, betrouwbaarheids- en validiteitsonderzoek moet worden verricht. De grootste verandering ten opzichte van de vorige versie van het beoordelingssysteem is dat de eerste vraag is gesplitst in drie deelvragen waarin expliciet wordt gevraagd of de meetpretentie, de doelgroepen en de functie van de test worden beschreven.

2 Kwaliteit van het testmateriaal

Dit criterium wordt beoordeeld door middel van acht vragen. Bij dit criterium komt onder andere aan de orde of testopgaven, scoring en instructie zijn gestandaardiseerd en of er voldoende aanwijzingen voor de geteste worden gegeven. Ook wordt er een vraag gesteld over de voor specifieke bevolkingsgroepen mogelijk kwetsende inhoud van items. Nieuw bij dit criterium is dat er een vraag over de kwaliteit van de items wordt gesteld en dat er aparte series vragen zijn opgenomen voor afname met behulp van papier-en-potlood en afname via een computer.

3 Kwaliteit van de handleiding

Dit criterium wordt beoordeeld door middel van zeven (papier- en-potlood) of tien (computer) vragen. Bij dit criterium wordt gevraagd naar de informatie die wordt geboden ter ondersteuning van de testgebruiker bij afname en interpretatie van de test. De voornaamste wijziging is dat er voor afname via een computer drie additionele vragen worden gesteld.

4 Normen

Dit criterium wordt beoordeeld door middel van zeven vragen (normgerichte interpretatie) of vijf vragen (domeingerichte of criteriumgerichte interpretatie). Nieuw voor alle typen normen is dat bij de beoordeling rekening wordt gehouden met een verjaringstermijn. Bij normgerichte interpretatie wordt verder vastgesteld wat de kwaliteit is van de normen en van de erbij verstrekte informatie. Ook voor normen die via continue normering worden berekend, worden nu richtgetallen gegeven voor de gewenste grootte van de normgroepen. Nieuw zijn ook de vragen die betrekking hebben op domeingerichte en criteriumgerichte interpretatie.

5 Betrouwbaarheid

Dit criterium wordt beoordeeld door middel van drie vragen. Eerst wordt de hoogte

van de betrouwbaarheidscoëfficiënten beoordeeld en vervolgens de kwaliteit van het uitgevoerde onderzoek naar de betrouwbaarheid. Als gevolg van nieuwe ontwikkelingen wordt er nu naar zes mogelijke betrouwbaarheidsmaten gevraagd (in vergelijking met vier in de vorige versie van het beoordelingssysteem).

6 Begripsvaliditeit

Dit criterium wordt beoordeeld door middel van drie vragen. Eerst worden de uitkomsten beoordeeld en vervolgens de kwaliteit van het uitgevoerde onderzoek naar de begripsvaliditeit. Nieuw is dat explicieter wordt aangegeven welke typen onderzoeksgegevens ter ondersteuning van de begripsvaliditeit kunnen dienen en welke typen gegevens voor een bepaalde beoordeling zijn vereist.

7 Criteriumvaliditeit

Dit criterium wordt eveneens beoordeeld door middel van drie vragen. Net als bij de begripsvaliditeit wordt eerst de hoogte van de uitkomsten beoordeeld en vervolgens worden deze geëvalueerd in het licht van de kwaliteit van de onderzoeksprocedure. Bij dit criterium hebben geen belangrijke wijzigingen plaatsgevonden.

De beoordelingen

Het oordeel voor elk van deze criteria kan ‘onvoldoende’, ‘voldoende’ of ‘goed’ zijn. De schaalpunten op de vragen zijn 1, 2 en 3; deze komen eveneens overeen met de betekenissen ‘onvoldoende’, ‘voldoende’ of ‘goed’. Bij enkele basisvragen moet de score ‘1’ worden geïnterpreteerd als ‘nee’, de score ‘2’ als ‘n.v.t.’ en de score ‘3’ als ‘ja’. Een negatief oordeel op een basis (deel) vraag leidt direct tot het oordeel ‘onvoldoende’ voor het betreffende criterium.

De inhoud van de vragen, de toelichtingen en de wegingsvoorschriften worden gegeven in de volgende zeven hoofdstukken. Bij de wegingsvoorschriften van sommige criteria moeten somscores van vragen worden berekend.

Het doel van de toelichtingen is een houvast te bieden bij de beoordeling en waar nodig de statistische of psychometrische beweegredenen te verduidelijken. De toelichtingen hebben uiteraard niet de pretentie een statistisch of psychometrisch leerboek te zijn. Bij onduidelijkheden kan men de referenties of andere literatuur op de gebieden van testconstructie en psychometrie raadplegen.

Op de site: www.cotan.nl (online COTAN Documentatie, geraadpleegd op 22/3/2015) zijn de COTAN-beoordelingen en bijbehorende informatie te vinden.