



BFP

Testcommissie

TQ-Scan

Een Vlaams beoordelingsstelsiem
voor de kwaliteit van tests

www.bfp-fbp.be/testcommissie

WOORD VOORAF

Dit document beschrijft het **Test Quality Scan beoordelingssysteem (TQ-Scan)** gehanteerd bij de inschatting van de kwaliteit van psychodiagnostisch testmateriaal in Vlaanderen, in het bijzonder wat betreft 'cognitieve vaardigheidstests'. Deze versie is tot stand gekomen door de Testcommissie die deel uitmaakt van de Belgische Federatie van Psychologen (BFP - FBP).

De BFP Testcommissie beoogt met de publicatie van dit beoordelingssysteem de bruikbaarheid en toepasbaarheid van testbeoordelingen te vergemakkelijken en toegankelijker te maken voor alle testgebruikers. De complexiteit op het gebied van testtheorie, testconstructie en testtoepassing betekent immers dat er hoge eisen worden gesteld aan de kennis en vaardigheden van testontwikkelaars, alsook testleiders.

De bedoeling van de TQ-Scan is op een efficiënte en haalbare wijze vanuit de Testcommissie een zicht te krijgen op de kenmerken en kwaliteiten van een test of psychodiagnostisch instrument. In dit document worden verschillende criteria gehanteerd – bij ieder criterium worden de kwaliteitseisen beknopt toegelicht. Deze criteria worden echter niet vooropgesteld als een exhaustieve en alomvattende leidraad op vlak van testconstructie.

De TQ-Scan is een Vlaams beoordelingssysteem met een screenend karakter en is niet bedoeld om uitgebreide en meer diepgaande beoordelingssystemen zoals opgesteld door de [EFPA](#) of [COTAN](#) of andere organisaties te vervangen. De TQ-Scan kan door zijn opbouw wel geïntegreerd worden in zulke uitgebreide beoordelingen.

Veerle Briers, Steven Joris, Walter Magez, Mark Schittekatte (voorzitter) & Marlies Tierens

Contact: mark.schittekatte@ugent.be

BFP Testcommissie

Mei 2020

Overzicht van het beoordelingssysteem

Testen of psychodiagnostische instrumenten worden (klassiek) beoordeeld op de volgende 5 criteria:

- Relevantie
- Testmateriaal, instructies, scoring en gebruik
- Normen
- Betrouwbaarheid
- Validiteit

Beoordelingsschaal

Ieder criterium wordt afzonderlijk beoordeeld. Het oordeel voor elk van deze criteria wordt geplaatst op volgende beoordelingsschaal:

- A goed tot zeer goed ('goed')
- B behoorlijk tot goed ('voldoende')
- C "grijze zone", duidelijk positieve punten, maar (veel) tekorten ('met reserve') *
- D niet goed, teveel tekorten ('onvoldoende') of substantieel gebrek aan informatie in de handleiding

() voorlopig aanvaardbaar in uitzonderlijke gevallen, al is de kwaliteit niet voldoende of (nog) niet bekend of verouderd.*

Wanneer een test een beoordeling A, B of C krijgt kan eventueel nog een '+' bijgevoegd worden door de beoordelaars ter nuancering in positieve zin.

Eindoordeel en weging van de te beoordelen criteria

Na rijp beraad werd beslist om binnen de TQ-Scan een eindbeoordeling (A, B, C of D, eventueel aangevuld met een "+") te hanteren bij elke beoordeling van een instrument. Een beoordeling via het [Test Review Model van EFPA](#) of via het [COTAN beoordelingssysteem](#) leidt niet tot een dergelijke eindbeoordeling per instrument – wel wordt er uitgebreid geïnformeerd over de kwaliteiten op de onderscheiden criteria. Ook wij verkiezen dat de testgebruiker evaluerende informatie krijgt over de sterktes en zwaktes van een instrument, maar we gaan ook een globale eindscore toekennen. De expliciete vraag van verschillende instanties in Vlaanderen naar een globaal kwaliteitslabel op instrumenten was hierin doorslaggevend. Zo zijn er de [limitatieve testlijsten van het RIZIV](#) en [deze van het Vlaams Agentschap Zorg en Gezondheid](#) die "acceptabele en niet-acceptabele" instrumenten wil onderscheiden, zich baserend op TQ-Scan beoordelingen. Ook het Vlaams Agentschap voor Personen met een Handicap (VAPH) [vermeldt en hanteert](#) de BFP-kwalificaties van IQ-tests in het kader van hun beschrijving van een verstandelijke handicap. Bij PRODIA en het Vlaams Kwaliteitscentrum voor Diagnostiek worden in hun richtlijnen '[Zwak Cognitief Functioneren](#)' en '[CDP Verstandelijke Beperking](#)' eveneens de BFP TQ-Scan beoordelingen gebruikt.

De vijf klassieke criteria ter beoordeling nl. *Relevantie / Testmateriaal / Normen / Betrouwbaarheid / Validiteit* spelen in het eindoordeel van deze TQ-Scan niet allen een even belangrijke rol. De beoordelingen van de

kwaliteit van de *Normen, Betrouwbaarheid en Validiteit (NBV)* vormen de basis voor het bepalen van een eindbeoordeling. De kwaliteitsinschatting van *Relevantie en Testmateriaal* gebeurt doch is eerder 'informatief' en speelt dus geen rol in de beslisbomen die tot een eindbeoordeling leiden.

Deze eerste kwalificatie-categorie verwijst naar de **kenmerken en kwaliteiten** van de cognitieve vaardigheidstest. De kwalificatielabels A+, A, B, C en D (uitstekend, goed, voldoende, "met reserve(s)", onvoldoende) worden gehanteerd om deze categorie uit te drukken.

Voor cognitieve vaardigheidstests in het bijzonder bestaat er nog een tweede kwalificatie-categorie die verwijst naar de **meetpretentie** van de test. Cognitieve vaardigheidstests kunnen worden gekwalificeerd als een algemene intelligentietest (a), een specifieke cognitieve vaardigheidstest (s), of een instrument dat toelaat het cognitief ontwikkelingsniveau (c) te bepalen.

Een visualisering van een beoordeling van een instrument neemt dan bijvoorbeeld volgende vorm * aan:

N(ormen)	B(etrouwbaarheid)	V(aliditeit)	R(elevantie)	T(estmateriaal)	TQS eindlabel
A	B	C	A+	A	B(a)

*illustratief

i Verloop van de beoordelingsprocedure

De Testcommissie beoordeelt testmateriaal dat ter kwalificatie aan de commissie wordt voorgelegd. Daarnaast voert de Testcommissie ook een actief beleid en worden soms tests beoordeeld waarvan de commissie vindt dat deze relevant zijn voor de klinische praktijk of vanwege onderzoeksmatige belangen.

Beoordelingsprocedure

De TQ-Scan beoordelingsprocedure verloopt als volgt:

- Twee beoordelaars doorlopen voor het verzamelde testmateriaal de verschillende onderdelen van het beoordelingssysteem, en doen dit onafhankelijk van elkaar.
- Via overleg worden beide werkstukken samengevoegd tot één geheel voor elk van de onderdelen. Dit resulteert in een *intern voorstel* dat aan de voltallige commissie wordt voorgelegd.
- Via overleg met de verschillende onafhankelijke commissieleden komen er eventueel aanpassingen aan het intern voorstel. Dit resulteert in een *voorlopig Commissierapport*.
- Het voorlopige Commissierapport wordt door de voorzitter bezorgd aan de testauteurs. Deze kunnen hierover in overleg gaan met de commissie, dit in een tijdsperiode van maximum 30 dagen.
- Op basis van het eventuele overleg met de auteurs kan de Testcommissie beslissen eventuele wijzigingen en/of aanvullingen aan te brengen aan het voorlopige Commissierapport. Dit resulteert in een *definitief Commissierapport, inclusief eindbeoordeling* voor de betrokken test.
- Uit het (uitgebreide) *definitief Commissierapport*, wordt een synthese-versie gedistilleerd, met de belangrijke aandachtspunten én de evaluatiecodes (A, B, C,...) van de vijf klassieke criteria ter beoordeling, aangevuld met het eindlabel, bv. A(a).

Beoordelaars krijgen tests toegewezen op basis van zijn of haar deskundigheid binnen een bepaald domein. Bij discrepanties in de beoordeling zal de beoordelaars worden gevraagd om in onderling overleg tot consensus te komen. In uitzonderingsgevallen kan een derde beoordelaar worden ingeschakeld. Alle commissieleden kunnen fungeren als medebeoordelaar. In veel gevallen zullen externe deskundigen optreden als beoordelaar, op basis van aan te tonen deskundigheid op een bepaald gebied. Er wordt ernaar gestreefd dat steeds minstens één van de beoordelaars een lid is van de Testcommissie.

Belangrijke beoordelingskwesties kunnen steeds worden besproken op de gezamenlijke overlegmomenten van de Testcommissie. Wanneer nieuwe informatie over een reeds beoordeelde test verschijnt, kan een herbeoordeling van deze test plaatsvinden. Ook al blijkt er geen aanvullende informatie of is er geen nieuw instrument te beoordelen, adviseren we toch één maal (bij voorkeur twee maal) per jaar de kwalificaties te herbekijken.

Vertrouwelijkheid

De beoordelingsprocedure steunt op het principe van de vertrouwelijkheid. In de eerste plaats betreft dit de anonimiteit van de beoordelaars: voor de testontwikkelaar blijft onbekend wie de test heeft beoordeeld. In de tweede plaats hebben onbevoegden nooit toegang tot het aangeschafte of ter beschikking gestelde testmateriaal – deze zijn in het bezit van de Testcommissie onder conditie van strikte geheimhouding.

Standpunt van de Testcommissie i.v.m. pen-en-papier versus digitale testafname/scoring

Verschillende van de gekwalificeerde instrumenten worden door de testuitgevers ook aangeboden om digitaal af te nemen. Een digitale afnamemodus kan (minstens) op 3 manieren:

- Een individuele afname met de geteste aan een scherm en met een verantwoordelijke testleider in de buurt tijdens de testsessie.
- Een individuele afname van de geteste met aanbieding van het materiaal op een tablet, met ook de testleider met een tablet in de hand tijdens de testsessie.
- Een testafname op afstand (*'remote testing'*) waarbij de geteste en testleider via 'videobellen' een testsessie afwerken.

In de nabije toekomst kan men ook voorstellen van groepsgewijze digitale afnames verwachten. Verschillende delen van een digitale testafname zullen vermoedelijk in de toekomst een verrijking betekenen van het zo valide en betrouwbaar mogelijk meten van psychologische constructen. Tot op vandaag blijkt er echter onvoldoende wetenschappelijke evidentie om ook maar één van de hieronder gekwalificeerde instrumenten, ondanks wat testuitgevers in beperkte equivalentiestudies beweren, op één of andere wijze digitaal af te nemen met een voldoende validiteit en betrouwbaarheid en vooral met degelijke normen. Digitale testafnames worden in Vlaanderen dus voorlopig afgeraden, zolang er geen normen ontwikkeld worden die analoog zijn aan de afnamemodus. Het belangrijkste tekort van de ons bekende equivalentiestudies tussen pen-en-papier afnames en digitale afnames is dat, waar voldoende gelijkwaardige resultaten bekomen worden op groepsniveau, dit niet is aangetoond voor scores van individuen bij een vergelijking.

Verdere aandachtspunten bij digitale afnames kunnen teruggevonden worden in enkele algemene richtlijnen van het Kwaliteitscentrum voor Diagnostiek vzw (De Ganck & Van Gampelaere, 2020a, 2020b).

In tegenstelling tot de scepsis ten aanzien van digitale afnames, op welke manier ook, wordt het gebruik van digitale scoringsprogramma's in het algemeen toegejuicht. Als belangrijkste voordelen worden terecht minder foutenlast vermeld evenals tijdswinst. Voorwaarde voor een voldoende kwaliteit is wel dat de uitgever voldoende kan

aantonen dat de algoritmen achter het scoringsprogramma foutenvrij zijn, en daar bleek tot nu toe steeds aan voldaan. Een gevaar is dat soms te veel cijfermatige output aangeleverd wordt, dat door de gebruiker soms verkeerd geïnterpreteerd kan worden, dus voorzichtigheid met minder bekende psychometrische maten is een aandachtspunt.

i Inhoud van de beoordelingsprocedure

Normen / Betrouwbaarheid / Validiteit (NBV-beoordeling)

De NBV-beoordeling is van primair belang in de evaluatie van testinstrumenten. Op drie belangrijke elementen wordt telkens gefocust:

- Is er evidentie met Vlaamse data of equivalentie-onderzoek?
- Hoe is de kwaliteit van de data?
- Hoe recent gebeurde de data-verzameling?

Beoordeling van de normen

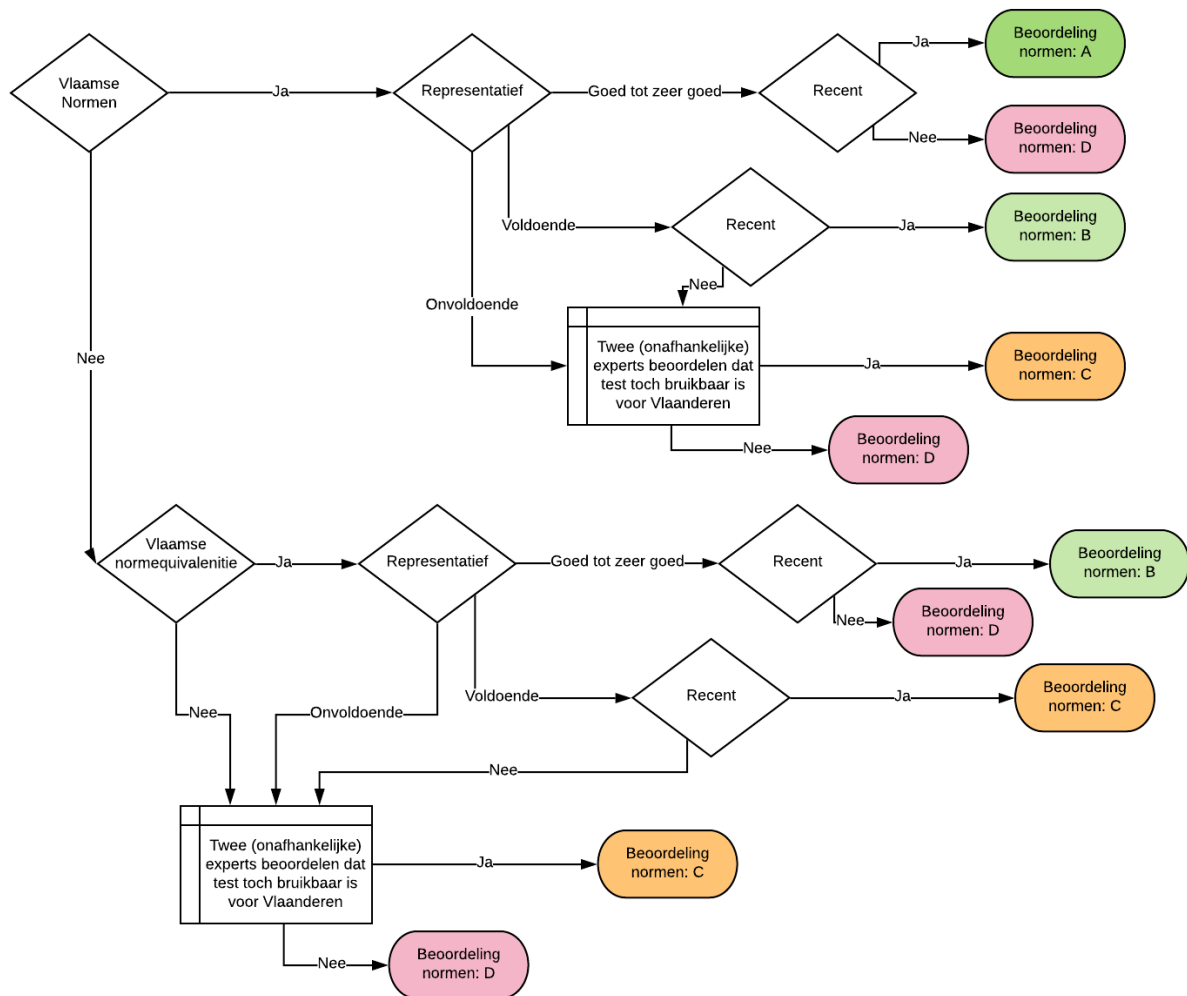
De beoordeling van de normen omvat volgende topics:

- Zijn de normen op Vlaamse data gebaseerd of is er equivalentie-onderzoek?
- Zijn de normgroepen voldoende representatief voor de algemene populatie en voldoende groot?
- Zijn de normgegevens voldoende recent?

De beoordeling van de normen omvat volgende criteria:

NBV-beoordeling	Inhoud
Normen	<ul style="list-style-type: none">▪ <i>Criterion 1.</i> Vlaamse normen/normreferentiekader aanwezig of normequivalentie wordt duidelijk aangetoond.▪ <i>Criterion 2.</i> Representatief voor Vlaanderen of representativiteits-equivalentie.▪ <i>Criterion 3.</i> Recente normen op de doelgroep beschreven in de handleiding.

Voor de beoordeling van de normen hanteren we een letterevaluatieschaal bestaande uit A, B, C en D (cf. supra) volgens onderstaande beslisboom. Bij de letters A, B en C kan eventueel een '+' worden toegevoegd ter nuancering.



Beoordeling van de betrouwbaarheid

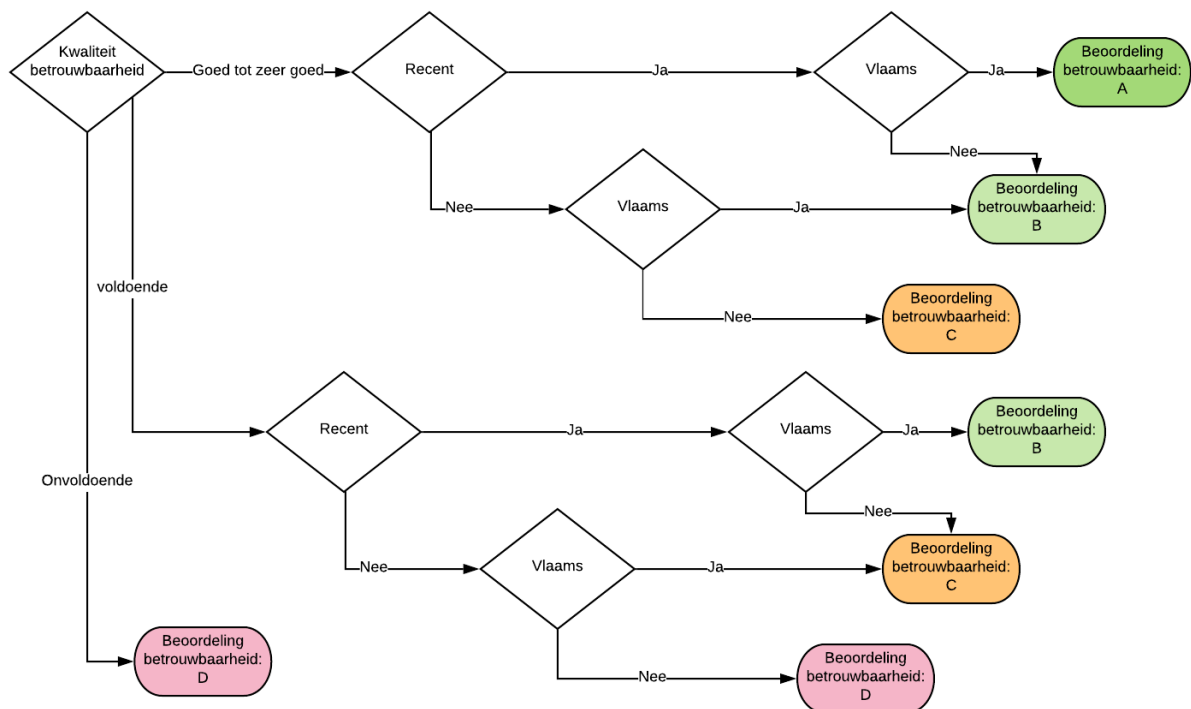
De beoordeling van de betrouwbaarheid omvat volgende topics:

- Interne betrouwbaarheid
- Test/hertestbetrouwbaarheid
- Interbeoordelaarsbetrouwbaarheid (eventueel)
- Paralleltestbetrouwbaarheid (eventueel)

De beoordeling van de betrouwbaarheid omvat volgende criteria:

NBV-beoordeling	Inhoud
Betrouwbaarheid	<ul style="list-style-type: none"> ▪ <i>Criterion 1.</i> Betrouwbaarheid is aangetoond op Vlaamse groepen, of Vlaamse equivalentie is aangetoond of kan afgeleid worden o.a. op basis van onderzoek op equivalente tests (literatuur, handleidingen, ...). ▪ <i>Criterion 2.</i> Betrouwbaarheid is in de normgroep (<i>handleiding</i>) onderbouwd en daar (minstens) voldoende. ▪ <i>Criterion 3.</i> Recent, op doelgroep beschreven in de handleiding.

Voor de beoordeling van de betrouwbaarheid hanteren we een letterevaluatieschaal bestaande uit A, B, C en D (cf. supra) volgens onderstaande beslisboom. Bij de letters A, B en C kan eventueel een '+' worden toegevoegd ter nuancering.



Beoordeling van de validiteit

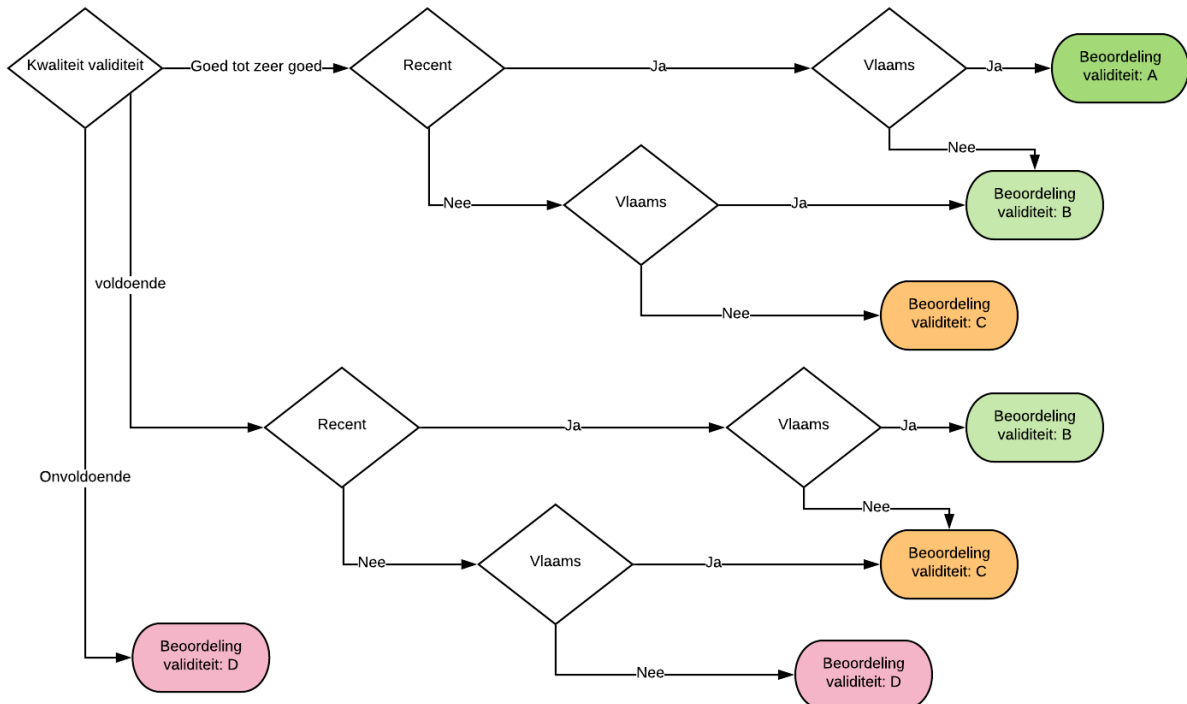
De beoordeling van de validiteit omvat o.a. volgende topics:

- Begripsvaliditeit
- Constructvaliditeit
- Externe criteria

De beoordeling van de validiteit omvat volgende criteria:

NBV-beoordeling	Inhoud
Validiteit	<ul style="list-style-type: none"> ▪ <i>Criterion 1.</i> Validiteit op Vlaamse groepen is aangetoond, of Vlaamse equivalentie is aangetoond of kan afgeleid worden o.a. op basis van onderzoek op equivalente tests (literatuur, handleidingen...). ▪ <i>Criterion 2.</i> Validiteit is in de normgroep (<i>handleiding</i>) onderbouwd en daar (minstens) voldoende. ▪ <i>Criterion 3.</i> Recent, op doelgroep beschreven in de handleiding.

Voor de beoordeling van de validiteit hanteren we een letterevaluatieschaal bestaande uit A, B, C en D (cf. supra) volgens onderstaande beslisboom. Bij de letters A, B en C kan eventueel een '+' worden toegevoegd ter nuancering.



Relevantie / Testmateriaal (RT-beoordeling)

Beoordeling van de relevantie

De beoordeling van de relevantie omvat volgende topics:

- In welke mate is het instrument relevant binnen het domein waarvoor het instrument is gepubliceerd?
- Speelt het in op actuele noden?
- Hoe breed inzetbaar is het instrument?

Ieder criterium wordt afzonderlijk beoordeeld. Het oordeel voor elk van deze criteria wordt geplaatst op volgende beoordelingsschaal (Bij de letters A, B en C kan eventueel een '+' worden toegevoegd ter nuancering):

- | | |
|---|-----------------------------------|
| A | duidelijk relevant |
| B | relevant |
| C | matig tot eerder beperkt relevant |
| D | niet relevant |

Beoordeling van het testmateriaal

De beoordeling van het testmateriaal omvat volgende topics:

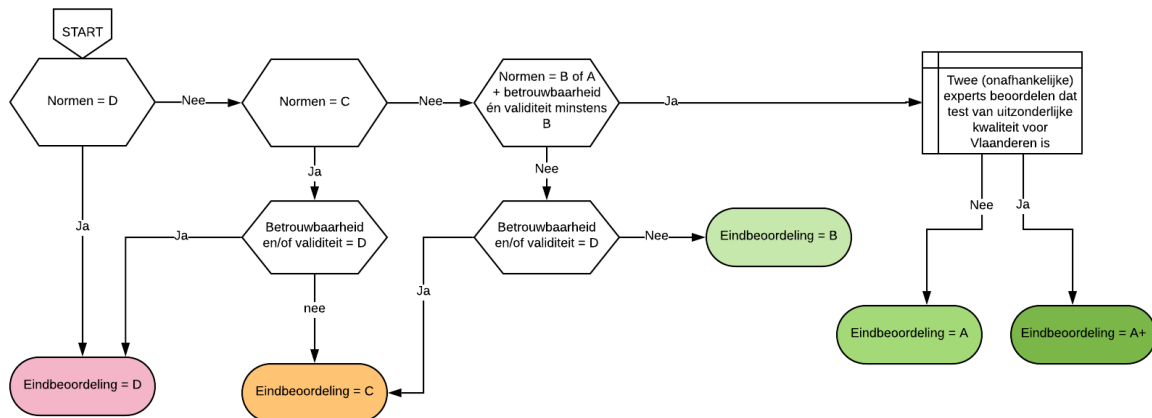
- Is het testmateriaal voldoende kwalitatief en verzorgd?
- Wat is het gebruiksgemak?
- Welke is de helderheid en kwaliteit wat betreft instructies, scoring en gebruik?

Ieder criterium wordt afzonderlijk beoordeeld. Het oordeel voor elk van deze criteria wordt geplaatst op volgende beoordelingsschaal (Bij de letters A, B en C kan eventueel een '+' worden toegevoegd ter nuancering):

- A zeer adequaat en helder
- B behoorlijk adequaat en helder
- C matig adequaat en helder, maar (veel) tekorten
- D onvoldoende

i Eindbeoordeling

Voor de eindbeoordeling hanteren we een letterevaluatieschaal bestaande uit A, B, C en D (cf. supra) volgens onderstaande beslisboom. Ook bij de eindbeoordeling kan bij de letters A, B en C eventueel een '+' worden toegevoegd ter nuancering. Zowel bij de beoordeling van de respectievelijke onderdelen als bij de eindbeoordeling kan en wordt bij voorkeur richtinggevend commentaar toegevoegd ter kadering en/ of ter informatie om de test door het toevoegen van een "+" naar een hogere kwalificatie te tillen.



De TQ-Scan omvat verschillende beoordelingsschalen die naar de psychometrische kwaliteiten van de test verwijzen. Onderstaande tabel geeft weer in welke beoordelingen een evaluatie kan resulteren.

Beoordeling	Inhoud
A+	Kwalificatie " zeer goed ": test van uitzonderlijke kwaliteit.
A	Kwalificatie " goed ": goede tot zeer goede test.
B	Kwalificatie " voldoende ": behoorlijke tot goede test.
C	Kwalificatie " met reserve ": "grijze zone", duidelijk positieve punten, maar (veel) tekorten m.a.w. voorlopig aanvaardbaar in uitzonderlijke gevallen, al is de kwaliteit niet voldoende of (nog) niet bekend of verouderd ¹ .
D	Kwalificatie " onvoldoende ", ook wegens gebrek aan informatie in de handleiding.

¹ Het moge duidelijk zijn dat als voor het betrokken testdomein A of B tests beschikbaar zijn, zij de voorkeur verdienen. Per uitzondering kan een test dit label ook krijgen in een "overgangperiode"; dit wordt dan expliciet aangeduid en is beperkt in de tijd.

Voor cognitieve vaardigheidstests in het bijzonder geldt echter nog een bijkomende kwalificatie:

- (a) Algemene intelligentietest (toetst de brede cognitieve vaardigheid – “BCV” – Gf en/of Gq én Gc én nog minstens 2 andere BCV).
- (s) Specifieke intelligentietest (toetst één of verschillende BCV, toetst daarbij eventueel Gf/Gq of Gc maar niet beide samen).
- (c) Cognitieve Ontwikkelingsschaal

Enkel algemene (a) intelligentietesten komen in aanmerking voor een IQ-bepaling. Specifieke (s) intelligentietesten komen hiervoor niet in aanmerking, zelfs indien zij dit pretenderen te doen.

Een visualisering van een beoordeling van een instrument neemt dan dus bijvoorbeeld volgende vorm * aan:

N(ormen)	B(etrouwbaarheid)	V(aliditeit)	R(elevantie)	T(estmateriaal)	TQS eindlabel
A	B	C	A+	A	B(a)

*illustratief

Hoe lang blijven de beoordelingen geldig?

In principe is de ‘*houdbaarheidsdatum*’ van elke kwalificatie geldig tot een nieuwe versie van kwalificaties gepubliceerd wordt. Minstens jaarlijks worden deze kwalificaties herzien. Wel wordt er bij een ‘uitdoofscenario’ van een bepaald testinstrument in principe 12 maanden tijd gegeven om zich een nieuwe, recentere versie (e.g. WPPSI-III-NL naar WPPSI-IV-NL) aan te schaffen, maar één jaar na publicatie zal de ‘oudere’ versie van een instrument, in de meeste gevallen een minder gunstig kwalificatie krijgen, te bepalen per geval door de Testcommissie.

i Toelichting en verslaggeving van een testbeoordeling

Toepassing van de TQ-Scan resulteert in een testbeoordeling met bijhorend kwalificatielabel. Deze testbeoordeling kan worden toegelicht in een:

- uitgebreid verslag (een **Commissierapport**)
- beknopt overzicht (een **Testfiche**).

Een testfiche zal zich steeds beperken tot een beoordeling op de 5 criteria met beknopte toelichting per criterium, alsook een eindoordeel. Testfiches uitgevaardigd door de Testcommissie zijn vrij raadpleegbaar op de website. Een Commissierapport is een uitgebreid verslag van de testbeoordeling en kenmerkt zich door een diepgaande analyse van het betrokken instrument. Commissierapporten uitgevaardigd door de Testcommissie zijn voorbehouden aan leden van BFP-FBP².

² Lidmaatschap van BFP-FBP kan worden bekomen via één van haar deelverenigingen. Meer informatie kan je vinden op www.bfp-fbp.be.

Referenties

De Ganck, J., & Van Gampelaere, C. (2020a). [Richtlijn tele-diagnostiek in Vlaanderen tijdens de Covid-19 crisis.](#) Kwaliteitscentrum voor Diagnostiek vzw.

De Ganck, J., & Van Gampelaere, C. (2020b). [Richtlijn Diagnostiek met Social Distancing. Diagnostiek tijdens COVID-19 – over de mogelijke impact van persoonlijke beschermingsmaatregelen tijdens fysieke consultaties.](#) Kwaliteitscentrum voor Diagnostiek vzw.

Goedgekeurd door: BFP Testcommissie

Datum: februari 2021



BFP

Testcommissie

www.bfp-fbp.be/testcommissie

© Copyright BFP-FBP

Niets in deze uitgave mag worden verveelvoudigd, overgenomen of gekopieerd zonder toestemming van de BFP Testcommissie.