

IQ-SCORE IS DRINGEND AAN MODERNISERING TOE

NAAR EEN NIEUWE INTERPRETATIE EN CLASSIFICATIE VAN DE GESCHATTE INTELLIGENTIE

Veel Nederlandse kinderen en jongeren ondergaan een intelligentietest. Het probleem bij de interpretatie is dat deze tests niet altijd hetzelfde meten. Afhankelijk van de gekozen test kan het IQ sterk variëren. De auteurs pleiten ervoor om de IQ-score niet als één punt, maar als een betrouwbaarheidsinterval weer te geven. De classificatie moet deze score beschrijven, niet de persoon.

SELMA A.J. RUITER, PETRA P.M. HURKS EN MARIEKE E. TIMMERMAN

Wat is intelligentie? Wechsler (1975) definieerde die als de cognitieve capaciteit waarmee iemand de wereld om zich heen begrijpt en de vindingrijkheid waarmee hij met de uitdagingen in die wereld omgaat. In de klinische praktijk is een intelligentietest in de regel onderdeel van een diagnostisch onderzoek dat antwoord moet bieden op hulpvragen als ‘welk type onderwijs past het beste bij dit kind?’

Voor het Nederlandse taalgebied zijn meer dan twintig tests gepubliceerd om iemands intelligentieniveau te onderzoeken (zie voor een overzicht Resing, 2015). Onderzoek heeft echter laten zien dat het IQ afhankelijk van de intelligentietest sterk kan verschillen. Geobserveerde correlaties tussen IQ's gemeten met verschillende tests variëren van circa 0,40 tot 0,80, zoals beschreven in diverse testhandleidingen (Tellegen & Laros, 2011; Van Baar e.a., 2014; Wechsler, 2002). We kunnen deze sterk variërende en soms relatief lage correlaties verklaren door onderlinge verschillen tussen intelligentietests in zowel de meetpretenties van de geïncludeerde subtests als de psychometrische kwaliteiten, zoals de betrouwbaarheid (Van Boxtel & Hurks, 2012). Maar wat verklaart nu precies de verschillen tussen de intelligentietests en wat zijn de implicaties voor de interpretatie van de resultaten? En daarbij: wat is de beste classificatiemethode voor een

individuele testscore? Dat laatste is een belangrijke stap in het diagnostisch proces. Er zijn verschillende systemen van intervalindelingen en de bijbehorende classificaties in omloop, bedoeld voor de communicatie met ouders en professionals in de praktijk. Maar die classificatiesystemen zijn niet allemaal even duidelijk en consistent.

DE ENE TEST IS DE ANDERE NIET

Een intelligentietest bij een kind of jongere levert twee soorten informatie op: kwalitatieve en kwantitatieve. De kwalitatieve informatie komt voort uit de observatie door de onderzoeker van het kind tijdens de afname. Hierbij gaat het bijvoorbeeld om het contact tussen de onderzoeker en het kind, diens taalgebruik, motivatie, werkhouding, doorzettingsvermogen en instructiegevoeligheid. Omdat deze informatie belangrijke sleutels vormen voor de beantwoording van de hulpvraag van bijvoorbeeld ouders of school, moet deze kwalitatieve informatie een belangrijk onderdeel zijn van een rapportage.

De kwantitatieve resultaten worden gebaseerd op de prestaties van het kind tijdens de test. Die geven een indruk van zijn prestaties ten opzichte van leeftijdgenoten. Deze informatie is van belang voor het beantwoorden van de hulpvraag. Bovendien is deze informatie vaak vereist voor de beoordeling van een



De foto's bij dit artikel zijn illustratief. De afgebeelde personen zijn niet dezelfde als die in het artikel.

Foto: Aleid Denter van der Gon

plaatsingsadvies, doorverwijzing of een aanvraag voor specifieke ondersteuning van het kind of de jongere op leergebied.

Om iemands intelligentieniveau te bepalen, worden meerdere (sub)tests afgenomen die pretenderen unieke, primair cognitieve functies, zoals perceptuele organisatie of verbaal begrip, te meten. Het intelligentieniveau van de testdeelnemer wordt vervolgens geschat op basis van de sommatie van de prestaties die zijn behaald op deze cognitieve (sub)tests, waarbij meestal iedere (sub)test hetzelfde gewicht krijgt. De score, die het intelligentieniveau moet reflecteren, wordt aange-

HET ONDERSCHIED 'VERBAAL-PERFORMAAL' IS NIET MEER LEIDEND

duid met de term Totaal IQ (TIQ) of IQ-score (IQ). Naast deze maat voor het intelligentieniveau bieden veel intelligentietests de testgebruiker ook de mogelijkheid om andere indices te berekenen, die een schatting geven van het niveau van meer specifieke cognitieve functies, zoals het verbaal begrip of het werkgeheugen.

Alle intelligentietests pretenderen iemands intelligentieniveau te meten. Het universele gebruik van de term IQ bij deze tests suggereert dat het individuele niveau van cognitief functioneren stabiel zou kunnen worden vastgesteld op basis van elke intelligentietest. Dit zou betekenen dat de IQ-score op de ene test uitstekend te voorspellen zou zijn op basis van de IQ-score op de andere test. Dit blijkt vaak niet zo te zijn: iemands IQ-score op de ene test verschilt in de regel van de IQ-score van diezelfde persoon op de andere test. Bij nadere beschouwing blijkt dat tests verschillen in de mate waarin zij een beroep doen op de schoolse vaardigheden en op de onderliggende cognitieve processen, zoals informatieverwerking. Het is van groot belang deze verschillen te onderkennen en er bij de interpretatie rekening mee te houden, want misinterpretatie kan grote gevolgen hebben voor het individu.

Ter illustratie: stelt u zich voor dat een school het beleid hanteert dat scholieren met maximale IQ-score X toegewezen worden aan een andere onderwijsvorm dan scholieren met een IQ-score groter dan X en dat de school enkel extra middelen beschikbaar stelt voor scholieren met een score X of lager. Omdat het specifieke IQ van een test afhankelijk is van de aard van de subtests, kan het zo zijn dat een kind met een disharmonisch profiel (een profiel met grote verschillen tussen relatieve sterktes en zwaktes) bij afname van de ene intelligentietest wel in aanmerking komt voor de specifieke onderwijsvorm (en de extra middelen), en bij afname van de andere test niet. Dit is bijvoorbeeld het geval als een kind de Nederlandse taal nog niet voldoende beheerst, omdat hij in Engeland is geboren en pas een half jaar in Nederland woont. Dit kind zal vermoedelijk een lagere IQ-score behalen op een

Nederlandstalige intelligentietest met subtests die verbaal begrip meten dan een IQ-score gebaseerd op een test zónder dergelijke subtests. Misschien komt hij daardoor in aanmerking voor een minder passend schooltype en/of financiering.

FLUID INTELLIGENCE IN ERE HERSTELD

Voor de klinische praktijk rijst dan de vraag welke intelligentietest een gedragsdeskundige moet gebruiken om het intelligentieniveau zo valide mogelijk te kunnen bepalen. Om de aard van intelligentietests algemeen te karakteriseren, is het nuttig het zogenaamde Cattell-Horn-Carroll (CHC)-model te bespreken (o.a. Horn & Blankson, 2012); zie voor een Nederlandse bespreking Resing, 2015. Het CHC-model is de grondslag voor een groot aantal moderne intelligentietests. Dit model is richtinggevend bij het bepalen van de cognitieve (deel)functies die zeker getest moeten worden alvorens een uitspraak te kunnen doen over iemands intelligentieniveau. Voorbeelden van intelligentietests die op het CHC-model zijn gebaseerd, zijn de WAIS-IV-NL en de RAKIT-2.

Voor de professional in de praktijk is het prettig om een intelligentietest te gebruiken die gebaseerd is op het CHC-model. Dergelijke tests geven namelijk inzicht in de ideeën over intelligentie die aan de test ten grondslag liggen. Ook is het bij tests die op dit model zijn gebaseerd duidelijk met welke subtest welke functie van intelligentie wordt onderzocht. Daarmee heeft de gebruiker houvast bij de keuze van de subtests die worden afgenomen, de interpretatie van de scores en de beschrijving van relatief sterke en zwakke punten in het cognitief functioneren van een persoon.

In 1938 postuleerde Thurstone zijn Model of primary mental abilities. Volgens dit model leveren diverse cognitieve functies (primary mental abilities), waaronder verbale vloeiendheid, geheugen en redeneren, een unieke bijdrage aan het intelligentieniveau. Dit betekent dat een gedragsdeskundige deze unieke cognitieve functies zou moeten onderzoeken om het intelligentieniveau van een individu te kunnen bepalen.

Sinds het verschijnen van Thurstones-model is er veel discussie geweest over welke specifieke cognitieve functies meegenomen moeten worden bij het vaststellen van het intelligentieniveau. Zo formuleerde Carroll (1993) circa 65 cognitieve functies. Het testen van al deze functies zou een zeer tijdrovend proces zijn en voor de geteste persoon te belastend. Binnen het zogenaamde CHC-model zijn deze 65 cognitieve functies (stratum I-functies genoemd) geclusterd in 8 tot 10 bredere, 'hogere-ordecategorieën' van cognitieve functies, waaronder vaardigheden die een beroep doen op crystallized of fluid intelligence. Deze hogere-ordefuncties worden in het model stratum II-functies genoemd (zie tabel 1 voor een specificatie). Gebruikers van het CHC-model stellen dat het essentieel is om zoveel mogelijk stratum II-functies te testen om het intelligentieniveau (het stratum III-niveau) van een individu te kunnen bepalen (Kaufman e.a., 2013).

Elke intelligentietest zou je kunnen ontleden in termen van deze stratum II-functies. De ene test zal

meer stratum II-functies dekken dan de andere.

Voor een overzicht van de intelligentietests die in Nederland beschikbaar zijn en de cognitieve deelfuncties die deze tests meten, verwijzen we de lezer naar Resing (2015).

Om een voorbeeld te geven, de huidige WISC-III-NL (Wechsler, 2002) dekt maar een klein aantal onderdelen van het CHC-model. Ontbrekende onderdelen zijn bijvoorbeeld subtests die specifiek gericht zijn op de visuele en auditieve informatieverwerking en het geheugen.

NIEUWE TESTS IN 2017

Momenteel zijn de uitgaven in voorbereiding van twee intelligentietests die naar verwachting in 2017 in Nederland worden gepubliceerd. In tegenstelling tot bijvoorbeeld de WISC-III-NL zijn deze tests (grotendeels) ontworpen op basis van het CHC-model. Het gaat om de Wechsler intelligentietest - vijfde editie (WISC-V, de opvolger van de WISC-III-NL, verwachte publicatie eind 2017) en de Intellectual and Developmental Scales - tweede editie (IDS-2-NL). De laatste is een nieuwe test voor Nederland (verwachte publicatie medio 2017).

De IDS-2-NL zal op het moment van uitgave de enige test in Nederland zijn die (vrijwel) alle stratum II-functies van het CHC-model in zich verenigt. De IDS-2-NL is niet alleen een intelligentietest, maar ook een test waarmee het executief functioneren, motorisch functioneren en sociaal-emotioneel functioneren apart kunnen worden onderzocht. Zelfs de schoolse vaardigheden maken deel uit van het instrument: spellen, lezen en rekenen. De makers van de test zijn expliciet uitgegaan van het CHC-model en hebben voor elk domein van het model twee subtests ontwikkeld. Eén daarvan doet voornamelijk een beroep op auditieve informatieverwerking, de ander op visuele informatieverwerking. De WISC-V dekt een groot deel

EEN IQ-SCORE ZEGT NIET ZOVEEL: NOEM ALTIJD DE NAAM VAN DE TEST

van de domeinen van het CHC-model, maar niet alle. Er ligt meer nadruk op onderzoek naar onderliggende cognitieve processen dan bij zijn voorganger het geval was. De test krijgt hierdoor een meer fluid karakter.

Voor de klinische praktijk is het van belang te weten dat het niet noodzakelijk is om alle stratum II-functies met één bestaande testbatterij, zoals de IDS-2-NL, te meten. De gedragsdeskundige kan ook gebruikmaken van de zogenaamde cross-battery assessment approach. Deze aanpak biedt gedragsdeskundigen een methode om de resultaten te interpreteren op een systematische, betrouwbare en theoretisch gestuurde manier en zo nodig aanvullend onderzoek te doen met cognitieve en neuropsychologische (sub)tests en toetsen voor schoolse vaardigheden uit andere batterijen.

Box 1: De theorie rond meetfouten

Om een betrouwbaarheidsinterval goed te kunnen interpreteren, is het goed om iets meer te weten over de theorie rond meetfouten. Het idee is dat iedere test een bepaalde meetonnauwkeurigheid heeft. Volgens de klassieke testtheorie is iedere geobserveerde testscore (X) opgebouwd uit een betrouwbaar deel, ook wel ware score (T) genoemd, en een deel dat te wijten is aan de meetfout (E). Dat wil zeggen dat iedere individuele score op een test beïnvloed wordt door zowel het niveau van de persoon (wat we willen meten, de ware score) als door een meetfout (wat we niet willen meten). De ware score is hierbij simpelweg gedefinieerd als de gemiddelde geobserveerde score over (vele) herhaalde testafnames van dezelfde persoon, waarbij het niveau van de persoon hetzelfde blijft. Dit laatste is in praktijk natuurlijk onmogelijk, want bij herhaalde testafnames zullen altijd herinnerings-, leer-, en vermoeidheidseffecten een rol gaan spelen. Dit betekent dat in de praktijk bij herhaalde afnames de ware score verandert, en dus dat de meetfout niet te scheiden is van de ware score. We kunnen dan ook nooit bepalen welk deel van een geobserveerde score te danken is aan het niveau van de persoon, en welk deel te wijten is aan de meetfout. Dat betekent dat we nooit in staat zijn om exact weer te geven wat het niveau van de persoon is. Wat we wel kunnen doen, is uitdrukking geven aan de mate waarin een geobserveerde score in het algemeen beïnvloed wordt door de meetfout. Deze invloed wordt uitgedrukt als de betrouwbaarheid van een test (in het Engels: reliability). Hoe hoger de betrouwbaarheid van de test, hoe kleiner de meetfouten in het algemeen zijn.

Dit geeft een completer beeld van iemands sterke en zwakkere kanten, dat ook psychometrisch beter is te verdedigen (Flanagan e.a., 2013, p.1).

Het gebruik van het CHC-model als basis voor de selectie van een testbatterij dan wel een combinatie van tests, geeft gedragsdeskundigen de middelen om kenmerken van een kind of jongere (zoals de relatieve sterktes en zwaktes) te beschrijven en te classificeren. Het overzicht van prestaties op stratum II-cognitieve functies geeft, in vergelijking tot enkel een IQ-score, een beter beeld van waarom een kind eventueel vastloopt in het dagelijks leven of op school. Ook kan zo'n overzicht concrete aanknopingspunten voor een behandeling bieden.

Het gebruik van het CHC-model als onderbouwing van een testbatterij of combinaties van tests ter bepaling van het intelligentieniveau (en de stratum II-cognitieve functies) betekent dat het vertrouwde en door veel gebruikers in de praktijk gebruikte onderscheid 'verbaal-performaal' niet meer leidend is. Dit komt simpelweg doordat de tests niet meer volgens deze structuur zijn opgebouwd. Het onderscheid verbaal-performaal maakt deel uit van een groter scala aan stratum II-cognitieve functies die samen de cognitieve intelligentie bepalen.

MEETFOUTEN

Hoewel bekend is dat de behaalde IQ-score van een individuele testdeelnemer sterk kan afhangen van de specifiek afgenomen intelligentietest, wordt in de praktijk nog altijd veel gewerkt met de verzamel score 'IQ'. Omdat specifieke intelligentietests verschillende combinaties van intelligentieaspecten dekken, pleiten



De foto's bij dit artikel zijn illustratief. De afgebeelde personen zijn niet dezelfde als die in het artikel.

wij ervoor om de rapportage van een IQ-score altijd samen met de naam van de test te rapporteren. Kraijer en Plas (2014) maken van dit punt ook melding in hun Handboek psychodiagnostiek en beperkte begaafdheid (p. 151). Dit zou betekenen dat 'het WISC-IQ' of het 'SON-IQ' wordt gerapporteerd. Op deze manier wordt expliciet aangegeven dat de gerapporteerde IQ-score gebaseerd is op een specifieke test en dat deze niet als vanzelf generaliserend kan worden gebruikt.

Na afname van een testbatterij moeten de ruwe scores van de testdeelnemer worden bewerkt tot genormeerde scores. Een genormeerde score drukt het niveau uit van een bepaald aspect van de intelligentie, of het Totaal IQ. Zo'n genormeerde score wordt geleverd in de vorm van een betrouwbaarheidsinterval (BI) (in het Engels: confidence interval) en een zogenaamde puntscore. De puntscore is de, op basis van de geleverde prestatie op de test, beste schatting van de relatieve prestatie van de persoon. Een betrouwbaarheidsinterval drukt uit dat de genormeerde testscore een meeton nauwkeurigheid heeft (box 1).

ZEKERHEID

Voor een individuele testscore kunnen we, als we de betrouwbaarheid van de test weten, aangeven met welke zekerheid de niveauscore valt in een bepaald interval, het C%-betrouwbaarheidsinterval. Hiermee wordt recht gedaan aan de meetonzekerheid van de test en daarmee aan allerlei toevallige, niet-structurele

invloeden op de score die een kind behaalt op een bepaald moment op een bepaalde test. Bijvoorbeeld, een 90%-betrouwbaarheidsinterval van 102 tot 120 voor een IQ-score geeft aan dat we er voor 90% zeker van zijn dat de ware IQ-score van een individu voor die betreffende test ligt tussen 102 en 120. De geobserveerde puntscore ligt altijd binnen het berekende betrouwbaarheidsinterval (soms precies in het midden, soms daar wat vanaf; dit is afhankelijk van de methode van berekening). Het is belangrijk dat u zich realiseert dat een dergelijk betrouwbaarheidsinterval alleen de onzekerheid als gevolg van meetfouten weergeeft. Dit betekent dat andere bronnen van onzekerheid, zoals verschillen in aard van de intelligentietests en onnauwkeurigheid in de afname doordat bijvoorbeeld wordt afgeweken van de handleiding, niet worden weergegeven. Hoe belangrijk het gebruik van een betrouwbaarheidsinterval is om een prestatie te karakteriseren, blijkt uit het feit dat zelfs intelligentietests met een hoge betrouwbaarheid (boven 0,90 en die door de Commissie Testaangelegenheden Nederland [COTAN] als 'goed' worden beoordeeld) nog vrij grote onnauwkeurigheidsmarges laten zien. Ter illustratie noemen we de Cito Intelligentietest (Eindtoets Basisonderwijs) die de zeer hoge betrouwbaarheid heeft van 0,95 voor de algemene intelligentie-index. Toch loopt bijvoorbeeld het 90%-betrouwbaarheidsinterval voor een persoon met een puntscore van 95 van 88,4 tot 101,6 (Van Boxtel & Hemker, 2009). Dit is dus een breedte van ruim dertien IQ-punten. Bedenk hierbij dat de IQ-scores zodanig genormeed zijn dat ze in de populatie van personen voor wie de test bedoeld is, een standaarddeviatie van 15 hebben. Dat betekent dat zelfs bij een test met een relatief hoge betrouwbaarheid, een 90%-betrouwbaarheidsinterval relatief breed is (bijna 1 standaarddeviatie). Er is dus een redelijke mate van onzekerheid over de relatieve positie van het onderzochte kind. Het gebruik van zo'n intervalscore drukt dan ook expliciet uit dat er meetonzekerheid is in testscore. Een puntscore geven zou de indruk wekken dat 'het IQ' met een enkele testafname te bepalen is. We kunnen een goede schatting maken wanneer we valide en betrouwbare instrumenten gebruiken, maar daarbij moet altijd bedacht worden dat we niet 'de ware score' aan het vaststellen zijn. Wij bevelen dan ook van harte aan, in lijn met de DSM-5 (APA, 2013), om een betrouwbaarheidsinterval te rapporteren en de puntschatting achterwege te laten. In de conclusie van een onderzoeksrapportage zou bijvoorbeeld de volgende zinsnede passend zijn: 'Testdeelnemer was gedurende de testafname voldoende coöperatief om een betrouwbare schatting te kunnen maken van zijn algemene intelligentieniveau. Het WISC-IQ van de testdeelnemer valt met 90% zekerheid in het interval 78 tot 91.'

HOEZO: 'LAAG BEGAAFD'?

De prestatie op een intelligentietest wordt uitgedrukt met een bepaald interval, waaraan klinici in de regel een classificatie verbinden over het algemeen cognitief functioneren van een persoon, gegeven dat het

Tabel 1. Overzicht stratum II cognitieve functies (in McGrew, 2009)

Stratum II cognitieve functies	Toelichting
<i>Fluid reasoning</i>	Het kunnen gebruiken van intentionele en gecontroleerde mentale processen om nieuwe problemen op te lossen, die niet 'op de automatische piloot' uitgevoerd kunnen worden. Voorbeelden van dergelijke processen zijn conceptformatie, classificatie en het herkennen van verbanden.
<i>Crystallized intelligence (comprehension-knowledge) (begrip-kennis)</i>	De kennis over een cultuur die een individu zich eigen heeft gemaakt middels acculturatie. Het gaat hier onder andere om de kennis over en de toepassing van een taal, begrippen of gebruiken die specifiek zijn voor een cultuur.
<i>Visual processing (visuele informatieverwerking)</i>	Het vermogen om visuele figuren en waarnemingen op te slaan, te activeren en te transformeren.
<i>Auditory processing (auditieve informatieverwerking)</i>	Het vermogen om geluiden te kunnen interpreteren en organiseren, zoals het kunnen discrimineren tussen tonen in een lied.
<i>Short-term memory (kortetermijngeheugen)</i>	Het vermogen om een beperkte hoeveelheid informatie (bijvoorbeeld een aantal cijfers of letters) kortdurend in het bewustzijn te kunnen (be)houden.
<i>Long-term storage and retrieval (langetermijnopslag en herinnering)</i>	Het vermogen om nieuwe informatie op te slaan en te consolideren in het langetermijngeheugen en om in een later stadium de opgeslagen informatie vloeiend op te halen.
<i>Cognitive processing speed (cognitieve verwerkings-snelheid)</i>	De vaardigheid om automatisch en vloeiend een relatief gemakkelijke en/of geautomatiseerde cognitieve taak uit te voeren, vooral in een situatie waarin hoge mentale efficiëntie nodig is.
<i>Decision and reaction speed (beslissen en reactie-snelheid)</i>	Het vermogen om een of meerdere elementaire beslissingen te nemen naar aanleiding van en of reacties te geven op eenvoudige stimuli.
<i>Quantitative knowledge (kwantitatieve kennis)</i>	Het betreft hier de verworven kennis over getallen, rekenen en wiskunde.
<i>Reading/writing (lezen en schrijven)</i>	Het betreft hier zowel de relatief basale vaardigheden met betrekking tot het spellen en lezen van een enkel woord, als de vaardigheid om (gerelateerde) tekst(en) te kunnen lezen, begrijpen en schrijven.

gebaseerd is op een betrouwbaar, correct genormeerd en valide intelligentieonderzoek. Dit geeft degene die de test scores moet interpreteren houvast in de rapportage en bij het schrijven van een heldere en bruikbare conclusie. Maar is dat eigenlijk wel zo? Ligt de indeling van de intervallen vast in hoe IQ-scores (zoals 115 of 80) moeten worden geclassificeerd, in termen van bijvoorbeeld 'gemiddeld' of 'laag in het gemiddelde gebied'? En wat zegt het precies wanneer een persoon een 'beneden gemiddelde IQ-score' behaalt of een score die als 'laag begaafd/moeilijk lerend' wordt omschreven?

In trainingen en onderwijs dat wij geven, is de IQ-score als intervalscore - en dan met name de op intervallen gebaseerde classificatie (het kwalitatief beschrijven van de testresultaten) - een terugkerend punt van discussie. De intervalverdeling, en daarmee de classificatie, berust op het statistisch model dat de basis vormt van de normering van alle hedendaagse intelligentietests: de normale verdeling met 100 als gemiddelde en een standaarddeviatie van 15 punten.

Als bij iedere persoon in een populatie (bijvoorbeeld bij alle personen tussen 5 en 20 jaar in Nederland) eenzelfde intelligentietest zou worden afgenomen, dan zou dat een grote variatie in geobserveerde IQ-scores opleveren. De verdeling van de geobserveerde IQ-punt scores (voor nu is het voldoende om te denken in termen van de punt scores) wordt in statistische termen de 'normaalverdeling' genoemd. Deze normaalverdeling is verdeeld in verschillende intervallen en het gebied onder de curve representeert de kans op een bepaalde score. Dit levert informatie op over hoe bijzonder een score is. Bijvoorbeeld, onder een normaalverdeling verwijst het gebied rondom het gemiddelde tot maximaal twee standaarddeviaties onder of boven dit gemiddelde naar ongeveer 95 % van populatiescores. Voor IQ-scores is dit gebied dus tussen $(100 \pm 2 \times 15 =) 70$ en 130. Dan blijft er nog zo'n 5 % over. Omdat de normaalverdeling symmetrisch is, betekent dit dat slechts zo'n 2,5 % van de populatie boven 130 scoort en 2,5 % onder de 70. Dit zijn dus vrij uitzonderlijke IQ-scores.

Tabel 2. IQ- intervalindeling volgens handleidingen van internationaal gepubliceerde intelligentietests met bijbehorende classificatie (labeling).

IQ- intervalindeling	Classificatie*
≥ 130	zeer hoog
120-129	hoog
110-119	hooggemiddeld
90-109	gemiddeld
80-89	laaggemiddeld
70-79	laag
< 70	zeer laag

*Rectificatie van publicatie (okt. 2018)

In de praktijk worden diverse classificatieschema's gebruikt om de mate van 'bijzonderheid' van IQ-scores te duiden. Deze schema's verschillen in interval-indelingen en de bijbehorende labels. Zo zal een puntscore die valt in het interval 78 – 89 volgens het ene schema een classificatie krijgen 'grensgebied laag en laag in het gemiddelde gebied' en in het andere schema 'benedengemiddeld'. In de handleidingen van verschillende tests worden eigen indelingen voorgeschreven en scoringsprogramma's vermelden naast de standaardcores ook de bijbehorende beschrijvende classificatie. Niet alleen de interval-indeling, maar, veel belangrijker nog, de classificatieterminologie kan dan weer afwijken van wat men gewend is. Dit is verwarrend en komt de informatieoverdracht van gedragsdeskundigen naar ouders en andere professionals niet ten goede. We pleiten daarom voor het gebruik van een eenduidig systeem in interval-indeling en in classificatie.

IQ ALS CONTINUÛM

We stellen voor ons te voegen naar de interval-indeling die wordt voorgeschreven door de meest actuele en veelgebruikte tests wereldwijd, waaronder de Wechsler tests (WISC-V, 2014; WAIS-IV, 2012 en WPPSI-IV, 2012), maar ook de oudere Differential Ability Scales – Second Edition en de Stanford-Binet Intelligence Scale – Fifth Edition. De interval-indeling bestaat eruit dat categorieën onder en boven het gemiddelde (IQ = 100) gelijk verdeeld zijn in intervallen van 10 punten. Zie tabel 2 voor de interval-indeling en labels die gebruikt worden in deze tests. We merken op dat de voorgestelde interval-indeling in zekere zin arbitrair is. Met elke andere redelijk gespreide verdeling met voldoende intervallen, als bijvoorbeeld +2SD, +1SD, 0SD, -1SD, of -2SD afwijking van het gemiddelde, kunnen wij ons ook verenigen, zolang professionals onderling dezelfde interval-indeling en labels gebruiken, opdat de communicatie tussen partijen zo optimaal mogelijk verloopt.

Het is belangrijk te bedenken hoe om te gaan met de behaalde testscore in relatie tot de voorgestelde interval-indeling en classificatie. Hierbij moet men uitgaan van een redelijk (bijv. 90%-) betrouwbaarheidsinterval en dus zeker niet van de puntscore. Scoort een testdeelnemer bijvoorbeeld een WISC-IQ met een 90%-betrouwbaarheidsinterval van 78 tot 89, dan hoort daar de beschrijving bij 'grensgebied

zeer laag en laag in het gemiddelde gebied'. Als men de testscore louter als puntscore zou behandelen, in dit geval bijvoorbeeld 84, dan zou de karakterisering alleen 'laag in het gemiddelde gebied' zijn. Hiermee zou men ten onrechte de rol van toevallige invloeden negeren, en daarmee de testdeelnemer mogelijk tekort doen. Dit houdt in dat scores vaker dan nu het geval is, omschreven worden als liggend in het grensgebied van twee intervallen. Op deze manier zal het schema meer als een continuüm worden gehanteerd dan als een schema met scherpe grenzen tussen intervallen. De benadering als continuüm doet volledig recht aan de aard van de IQ-scores, en heeft ook vanuit dat perspectief de voorkeur.

BEZORGDEID

In lijn met de DSM-5 zou bovenstaande betekenen dat we bijvoorbeeld binnen het reguliere basisonderwijs ons zorgen moeten maken om kinderen die een testresultaat behalen in overeenstemming met een (ware) IQ-score van 70 - 75 of lager. Rekening houdend met de meeton nauwkeurigheden van intelligentietests zoals hierboven beschreven, betekent dit dat als in het, zeg, 90%-betrouwbaarheidsinterval van een individuele testscore het getal 75 of lager valt, we

ONDERZOEKER MOET TIJDENS IQ-TEST OOK NAAR HET KIND ZÉLF KIJKEN

ons zorgen moeten maken om de geteste persoon. Ter illustratie: kind Z in het reguliere onderwijs behaalt een IQ-score van 79 met een 90%-betrouwbaarheidsinterval van 73 tot 86. Bovenstaande rationale volgend, betekent dit dat we ons zorgen moeten maken over dit kind – de cijfers 75 en lager vallen immers binnen het geschatte betrouwbaarheidsinterval. Het kind heeft een verhoogd risico op uitval binnen het reguliere basisonderwijs en heeft waarschijnlijk baat bij extra dan wel aangepast onderwijs. Deze conclusie kan overigens pas getrokken worden wanneer ook andere persoonskenmerken onderzocht zijn. Een resultaat op een intelligentietest is nooit een voldoende

onderbouwing wanneer in conclusies belangrijke uitspraken worden gedaan over de persoon en over het perspectief van deze persoon, bijvoorbeeld met betrekking tot onderwijs.

In het voorgestelde en internationaal reeds toegepaste classificatieschema (tabel 2) beschrijven de classificaties de testcores van de persoon en niet de persoon als geheel. Het schema vermijdt expliciet termen als ‘hoogbegaafd’, omdat een term als hoogbegaafdheid de gehele persoon karakteriseert. Een hoge mate van intelligentie is slechts een van de kenmerken van een hoogbegaafd persoon. Om iemand te kunnen karakteriseren (en eventueel diagnosticeren) is informatie nodig over zijn functioneren op verschillende domeinen. Cognitief functioneren is er slechts één van.

Dat geldt ook voor een diagnose ‘verstandelijke beperking’. Daarvoor is het noodzakelijk om niet alleen onderzoek te doen naar het algemeen intelligentieniveau, maar ook naar iemands adaptief functioneren (het toepassen van vaardigheden in het dagelijks leven of het sociaal aanpassingsvermogen). Pas wanneer deze beide domeinen in kaart zijn gebracht,

kan wel of niet de diagnose ‘verstandelijke beperking’ worden gesteld. De DSM-5 (APA, 2013) schrijft zelfs voor dat de ernst van de verstandelijke beperking overwegend gebaseerd moet zijn op het adaptief functioneren en niet op de IQ-score. De noodzakelijke ondersteuning is namelijk afhankelijk van iemands adaptief functioneren. Daarbij zijn de BI'en van scores in de lage regionen van de IQ-range in de regel erg breed, zodat er grote onzekerheid blijft bestaan over het relatieve intelligentieniveau van deze personen.

Tot slot merken we op dat het, hoe belangrijk het kwantitatieve resultaat van een test ook is, van minstens zo groot belang is om iemands handelingen en werkhouding tijdens de testafname te observeren. Een gedragsdeskundige moet in staat zijn om de behaalde testscore goed te kunnen plaatsen op basis van overige informatie en zijn persoonlijke observaties en indruk van het functioneren van degene die hij heeft getest. Daarbij functioneert een schema als hulpmiddel, het is geen uitgangspunt. Het zijn werkmodellen, waarmee we de resultaten van een onderzoek kunnen duiden en er de best passende actie op kunnen toepassen.

LITERATUUR

- > American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders (5th ed.)*. Arlington, VA: American Psychiatric Publishing.
- > Baar, A.L. van, Steenis, L.J.P., & Verhoeven, M. (2014). *Bayley-III-NL, administration manual*. Amsterdam: Pearson Assessment and Information B.V.
- > Boxtel, H.W. van, & Hemker, B. (2009). *Wetenschappelijk verantwoording van de Cito Intelligentie-test Eindtoets Basisonderwijs*. Arnhem: Cito.
- > Boxtel, H.W. van, & Hurks, P.P.M. (2012). *Intelligentiebepaling bij zeer lage niveaus*. Tijdschrift voor neuropsychologie, 7(1), 40-48.
- > Carroll, J.B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. Cambridge, MS: Cambridge University Press.
- > Flanagan, D.P., Ortiz, S.O., & Alfonso, V.C. (2013). *Essentials for cross-battery assessment (3e druk)*. Hoboken, NJ: Wiley.
- > Horn, J.L., & Blankson, A.N. (2012). Foundations for better understanding of cognitive abilities. In: D.P. Flanagan, J.L. Genshaft & P.L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests and issues* (pp. 53-91). New York: Guilford Press.
- > Kaufman, J.C., Kaufman, S.B., & Plucker, J.A. (2013). *Contemporary theories of intelligence*. In: D. Reisberg (Ed.), *Oxford handbook of cognitive psychology*. Oxford: Oxford University Press.
- > Kraijer, D.W. & Plas, J.J. (2014). *Psychodiagnostiek. Handboek psychodiagnostiek voor de hulpverlening aan kinderen*. Utrecht: Tijdstroom.
- > Resing, W.C.M. (2015). *Psychodiagnostiek. Handboek intelligentietheorie en testgebruik*. Amsterdam: Pearson Assessment and Information B.V.
- > Tellegen P.J. & Laros, J.A. (2011). *SON-R 6-40 Snijders-Oomen Niet-verbale Intelligentietest*. Amsterdam: Hogrefe.
- > Thurstone, L. L. (1938). *Primary mental abilities*. Chicago: University of Chicago Press.
- > Wechsler, D. (1975). Intelligence defined and undefined: A relativistic appraisal. *American Psychologist*, 30, 135-139.
- > Wechsler, D. (2002). *WISC-III-NL: Wechsler Intelligence Scale for Children (3rd ed.)*. Handleiding en verantwoording: Kort, W., Schittekatte, M., Dekker, P.H., Verhaeghe, P. Compaan, E.L., Bosmans, M. & Vermeir, G. Amsterdam: Pearson Assessment and Information B.V.