

Correspondentieadres:

Dr. H.W. van Boxtel

Universiteit Utrecht, Faculteit Sociale Wetenschappen
Afdeling Ontwikkelingspsychologie

Postbus 80140

3508 TC Utrecht

h.w.vanboxtel@uu.nl

Intelligentiebepaling bij zeer lage niveaus

Samenvatting

Primaire doelstelling van de Commissie Testaangelegenheden Nederland (COTAN-NIP) is het bevorderen van de kwaliteit van psychologische tests en het professioneel handelen bij testgebruik. Daarnaast informeert de cotan testgebruikers over de stand van zaken op testgebied. Vanuit laatstgenoemde doelgroep, ontvangt de COTAN met regelmaat vragen die betrekking hebben op de bepaling en de interpretatie van IQ-scores, als operationalisatie van het begrip intelligentie. De normering, betrouwbaarheid en de validiteit van IQ-scores staan hierbij veelal ter discussie, met name het gebruik van dergelijke scores ter classificatie van een groep individuen van wie vermoed wordt dat zij qua prestaties behoren tot de onderste (of juist bovenste) 2,5 tot 16% van de populatie. Desondanks lijken wetgevers blind te varen op scores verkregen met traditionele intelligentietests, bijvoorbeeld in het kader van criteria voor toelating tot het speciaal onderwijs of het toekennen van leerlinggebonden financiering. De auteurs zullen in dit artikel ingaan op de beperkingen van intelligentietests in algemenere zin. Daarnaast zullen zij specifiek ingaan op het gebruik van dergelijke instrumenten ter classificatie van lage tot zeer lage intelligentieniveaus en doen zij in deze context aanbevelingen ter optimalisering van de (psycho)diagnostiek.

Inleiding

Met enige regelmaat bereiken het NIP of de COTAN vragen zoals onderstaande:

Momenteel heb ik een jongen van ruim tien jaar in onderzoek - een leerling van een ZML-school. Hij is in een speciale groep geplaatst, maar ouders vinden dat hij daar te weinig opsteekt en willen hem in een reguliere groep geplaatst zien. De school hanteert daarvoor als voorwaarde dat de leerling tenminste beschikt over een IQ van 35. Ik heb de jongen nu psychologisch onderzocht met de Wechsler Nonverbal Scale of Ability (WNV) - Nederlandse bewerking. Wanneer ik deze test binnen de eigen leeftijdscategorie scoor, komt er een IQ < 55 uit (de test differentieert niet onder de 55). Wanneer ik op subtest-niveau de leeftijdsequivalenten bekijk, kan ik niet lager dan 7;1 jaar. Wat nu?

Mijn vragen zijn:

- Hoe kan ik een IQ-score rondom de 35 bepalen? Is er misschien een formule om de score om te zetten zodat in de vergelijking met leeftijdgenoten het zeer lage niveau van de jongen tot uitdrukking komt?
- Zijn er andere instrumenten bruikbaar om in deze lage range te kunnen differentiëren?

De specifieke casus die hier beschreven wordt en waarin het gaat om het operationaliseren van een toelatingscriterium $IQ > 35$ kan men opvatten als een voorbeeld van een brede klasse van problematische diagnostische vragen, namelijk vragen die betrekking hebben op het adequaat beschrijven van lage tot zeer lage intelligentieniveaus. In feite begint deze klasse al bij IQ-scores lager dan 70, waarbij men volgens het inmiddels aardig ingeburgerde classificatievoorstel van Resing en Blok (2002) spreekt van 'Verstandelijke beperking' of (traditioneler) 'Zwakzinnigheid'. Zelfs bij de relatief hogere IQ-scores in deze categorie (60 tot 70) zijn de problemen legio. Traditionele intelligentietests hebben ook bij deze groep al aanzienlijke beperkingen, zoals uit het vervolg van dit artikel zal blijken.

Intelligentie, IQ en intelligentietests: Algemene knelpunten

Een deel van de hierboven genoemde problemen hangt samen met het concept intelligentie en de operationalisatie daarvan in algemenere zin. Velen lijken in de praktijk uit te gaan van de veronderstelling dat aan een intelligentietest vrijwel dezelfde status kan worden toegekend als aan de operationalisatie van het begrip lengte, namelijk die van een goed geijkte duimstok. Niets is minder waar. Op de eerste plaats is er geen eenstemmigheid over wat men onder het concept intelligentie zou kunnen of moeten verstaan. In de literatuur treft men een lange reeks definities die weliswaar met elkaar gemeen hebben dat intelligentie iets van doen heeft met cognitieve vermogens, probleemoplossingsvaardigheden en capaciteiten om te leren, maar tegelijk ook vrij veel ruimte bieden om het concept op zeer verschillende manieren te operationaliseren. Daar komt nog bij dat veel van deze definities erg breed zijn en betrekking hebben op een 'overall, aggregate, or global capacity' (zoals de Wechsler-definitie van intelligentie ons vertelt; zie Kievit e.a., 2008). De testconstructeur heeft dus nogal wat armslag als hij zich voorneemt om een test of testbatterij te construeren die beoogt het concept intelligentie te meten. In veel gevallen

leidt zijn onderneming tot een verzameling taken die ieder afzonderlijk een bepaald aspect, component, deelvaardigheid of factor van het alomvattende concept operationaliseren. De keuzes in dezen zijn afhankelijk van het theoretische uitgangspunt van de constructeur: vertrekt hij bijvoorbeeld vanuit een componententheorie over intelligentie (als voorbeeld, zie Sternberg, 1985) of vanuit de klassieke psychometrische traditie waarin op verschillende niveaus factoren worden onderscheiden (als voorbeeld, zie Carroll, 1993)?

Hoe het ook zij, het resultaat is vaak een testbatterij waarbij een verzamel score, gebaseerd op de te onderscheiden subtest scores, via enkele sommeringen en transformaties leidt tot een IQ-score, op een zodanige manier dat het gemiddelde in de normpopulatie 100 bedraagt en de standaarddeviatie 15. Het is verleidelijk om bij deze schijnbare eenvormigheid in het resultaat in de praktijk te werken met de veronderstelling dat men met deze score hét IQ te pakken heeft, en dat men daarmee dé intelligentie van een individu heeft vastgesteld. In feite zal elke testbatterij bij een bepaald individu tot een ander resultaat leiden, afhankelijk van de match tussen enerzijds de specifieke vaardigheden die door de batterij worden afgetapt en anderzijds de specifieke sterktes en zwaktes in het profiel van cognitieve vermogens van dat individu. Gegeven de verre van perfecte intercorrelaties tussen de subtest scores is dit een onontkoombare conclusie. Ter illustratie: de intercorrelaties tussen de subtest scores van de NIO (Nederlandse Intelligentietest voor Onderwijsniveau; Van Dijk & Tellegen, 2004) variëren tussen 0,37 en 0,69, met een gemiddelde van 0,53.

Hieruit blijkt dat met de gekozen subtest taken steeds andere deelvaardigheden en aspecten worden geoperationaliseerd en het eindresultaat op de test dus sterk afhankelijk is van de samenstelling van de batterij. Correlaties tussen verschillende intelligentiematen, gebaseerd op verschillende testbatterijen liggen dan ook vaak in een range van (ongeveer) 0,50 tot (ongeveer) 0,80. Zo is de correlatie van de WPPSI-III-NL (Wechsler Preschool and Primary Scale of Intelligence, Nederlandse editie, versie III) met de WISC-III-NL (Wechsler Intelligence Scale for Children, Nederlandse editie, versie III; een test uit dezelfde Wechsler-traditie, vertrekkend vanuit een gemeenschappelijke definitie, met een zekere overlap in leeftijdsbereik) 0,80, maar is de correlatie met zijn voorganger (de WPPSI-R; Wechsler Preschool and Primary Scale of Intelligence Revised) al een stuk lager (0,66) en met een soortgenoot als de SON-R (Snijders-Oomen Niet-verbale Intelligentietest – Revisie) slechts 0,51 (Hendriksen & Hurks, 2009). Feit is dat zowel de WPPSI als de WISC en de SON bij een kind van een bepaalde leeftijd geaccepteerde instrumenten zijn om (de) intelligentie vast te stellen. Het behoeft geen verder betoog dat afname van deze verschillende instrumenten tot zeer verschillende IQ-scores kan leiden.

Een bijkomend probleem is dat de betrouwbaarheid van intelligentietests (onafhankelijk van type batterij) beperkt is. Dit leidt tot vrij grote onnauwkeurigheidsmarges, zelfs bij testbatterijen met betrouwbaarheden van boven de 0,90 die door de COTAN als ‘goed’ worden gekwalificeerd. Dit laatste impliceert dat ze geschikt zijn voor gebruiksdoelen die betrekking hebben op het nemen van belangrijke beslissingen, zoals toelating tot het speciaal onderwijs. Zo heeft de Cito Intelligentietest bij de Eindtoets Basisonderwijs een keurige betrouwbaarheid van omstreeks 0,95 (afhankelijk van de gehanteerde maat), maar dit getal resulteert niettemin in een 90%-betrouwbaarheidsinterval van plus of min 6,61 IQ-punten. Men weet dus met 90% zekerheid dat de gevonden IQ-waarde voor een individu ligt in een gebied met een breedte van ruim 13 IQ-punten (Van Boxtel & Hem-

ker, 2009). Dit is niet uniek voor deze test, maar kenmerkend voor elke intelligentietest met een soortgelijke hoge betrouwbaarheid. Vanzelfsprekend wordt het probleem groter naarmate een IQ-test gehanteerd wordt met een lagere betrouwbaarheid.

Daar komt nog een derde probleem bij en dat is dat bij de normering van dergelijke batterijen de representativiteit van de normsteekproeven nogal te wensen overlaat. Gezien de grote praktische en financiële problemen die in de regel verbonden zijn aan het normeren van met name individueel afneembare intelligentiebatterijen hoeft dit geen verwondering te wekken. Om een paar voorbeelden te noemen: de aantallen individuen in normgroepen zijn vaak wat aan de lage kant, de steekproef is net niet helemaal representatief naar bijvoorbeeld schooltype, mate van verstedelijking of percentage allochtonen, soms zijn kinderen uit het speciaal onderwijs niet meegenomen en/of soms zijn individuen niet aselekt maar via een sneeuwbalmethode geworven. Het zijn misschien kleine effecten die bij beoordeling door de COTAN niet altijd zullen leiden tot een onvoldoende, maar aangenomen mag worden dat de normering, en dus de vaststelling van het IQ bij het individu erdoor beïnvloed is en wordt. De lezer wordt dan ook geacht waakzaam te zijn, zelfs bij een COTAN-beoordeling 'voldoende/goed'.

Ten slotte kunnen we in deze context nog wijzen op het mogelijke effect van normveroudering. Veel onderzoek is inmiddels gedaan naar het zogenoemde Flynn-effect, zowel in vele westerse als in enkele niet-westerse landen (Flynn, 2007). Het Flynn-effect houdt in dat bij hernormering van een intelligentietest in de regel blijkt dat de gemiddelde score op de test in de normsteekproef is toegenomen. Grofweg kan men stellen dat in de jaren tussen 1930 en 2000 de gemiddelde IQ-score met 3 à 5 punten per decennium is toegenomen. Dat is een effect van 10 à 20 IQ-punten per generatie. Ook al is het Flynn-effect wellicht de laatste jaren in de westerse landen wat aan het stabiliseren, dergelijke effecten zijn met name een probleem daar waar absolute IQ-scores worden gehanteerd. Een voorbeeld zijn wetgevers (als de overheid, [zorg]verzekeraars en/of onderwijsorganisaties) die absolute IQ's hanteren als grenswaarden, op basis waarvan beslissingen genomen worden die in sommige gevallen zeer ingrijpende gevolgen hebben (zoals bij leerlinggebonden financiering of zoals bij eerdergenoemd voorbeeld omtrent criteria voor toelating tot het speciaal onderwijs).

Bovenstaande tekorten omtrent begripsafbakening, testsamenstelling en lage betrouwbaarheid en kanttekeningen bij de representativiteit van de steekproef zijn van toepassing op de bepaling en daaropvolgend het gebruik van IQ-scores binnen het gehele bereik. In het kader van intelligentiebepalingen bij lage tot zeer lage intelligentieniveaus komen daar nog andere problemen bij; deze zullen in de volgende paragrafen worden besproken.

Intelligentietests bij lage tot zeer lage intelligentieniveaus

Ten eerste: op basis van de hierboven aangevoerde argumenten kan men concluderen dat uitspraken als 'Joost heeft een IQ van 82 en is dus niet toelaatbaar tot het PrO (Praktijkonderwijs), maar Linda met haar IQ van 78 wel' in deze algemeenheid geformuleerd geen hout snijden. Want de kans is groot dat de 'ware' IQ-scores van Linda en Joost er misschien toch wat anders uitzien dan met de afgenomen test wordt aangegeven. Toch lijken de door de wetgever vastgelegde indicatiecriteria dit soort uitspraken in de hand te werken. Er zijn immers strikte afgestgrenzen geformuleerd met betrekking tot 'het' IQ

(dat bijvoorbeeld moet liggen tussen 55 en 80 in geval van indicatie praktijkonderwijs – let wel: deze criteria variëren enigszins per regio), terwijl daarnaast jaarlijks in een lijst wordt vastgelegd welke intelligentietests zijn toegelaten om dat IQ vast te stellen. Zo is er voor 2010–2011 vastgesteld dat de eerdergenoemde WISC-III, NIO, SON-R, en WNV, alsmede de Drempelest (2007), de Groninger Intelligentietest voor Voortgezet Onderwijs (GIVO, 1999), de Instaptoets Voortgezet Onderwijs (IVO, 2002) en de Nederlandse Differentiatie Testserie (1999, normen 2004) gebruikt mogen worden als screenings- en testinstrumenten ten behoeve van de indicatie voor (clusters) praktijkonderwijs (<http://wetten.overheid.nl/BWBRO026141>). De kans is echter aanzienlijk dat al deze tests, indien zij bij Joost en Linda zouden zijn afgenomen, tot een ander IQ zouden hebben geleid. Een voorbeeld: de bovengenoemde WNV en SON-R zijn schatters van primair het performale IQ van het kind, terwijl bijvoorbeeld de WISC zowel verbale als performale subtests hanteert om een IQ-score te bepalen. De praktijk leert echter dat veel verstandelijk beperkten hogere performale dan verbale scores laten zien, hetgeen bij gebruik van eerstgenoemde tests zou leiden tot een mogelijke overschatting van het intelligentieniveau (uiteraard in relatie tot iets anders, bijvoorbeeld het te volgen onderwijs). Kortom, het is lastig om de IQ-waarden van Linda en Joost onderling te vergelijken, zeker als zij met verschillende testbatterijen tot stand zijn gekomen.

Ten tweede gaat het bij de indicatiestelling vaak om aftestgrenzen die lage tot zeer lage IQ-scores betreffen. In de criteria voor de indicatiestelling cluster 3 (waar het onder meer gaat om verstandelijke beperkingen of meervoudige lichamelijke en verstandelijke beperkingen) komen immers waarden voor als < 70 , < 55 en zelfs < 35 . Zelfs bij de wat hogere waarden van 70 en 80 is er bijna per definitie sprake van bodemeffecten. De met deze bodemeffecten samenhangende extra onbetrouwbaarheid (bovenop de al eerder aangegeven algemene effecten van onbetrouwbaarheid) mag bekend verondersteld worden. Toch gaan we er hier graag nog eens op in, want de in testhandleidingen genoemde (en door de COTAN in algemene zin met een ‘voldoende’ of ‘goed’ beoordeelde) betrouwbaarheden zijn met grote waarschijnlijkheid misleidend te noemen voor deze scoreregionen. Een test of subtest kan nu eenmaal niet voor het hele scorebereik dezelfde meetnauwkeurigheid realiseren. Zo worden leervorderingstoetsen dikwijls zo geconstrueerd dat ze een optimale nauwkeurigheid hebben in de wat lagere en gemiddelde scorerange; de moeilijkheidsgraad van de toets wordt daarop afgestemd. Bij instrumenten die op basis van Item Respons Theorie (IRT) zijn geconstrueerd is het mogelijk deze zogenoemde lokale meetnauwkeurigheid te berekenen en zichtbaar te maken. Helaas is het bij de klassieke constructiemethoden die bij intelligentietests doorgaans worden gehanteerd, praktisch onmogelijk om lokale betrouwbaarheden te rapporteren. De gebruiker heeft dan aan de vermelde algemene betrouwbaarheden weinig houvast. Daardoor blijft het feit dat in het bijzonder bij intelligentietests voor kinderen vaak sprake is van grote bodem- en plafondeffecten verhuld.

Deze intelligentietests zijn immers in de regel geconstrueerd voor een bepaald doch zeer breed leeftijdsbereik, bijvoorbeeld zeven tot zestien jaar. Ook al wordt er gewerkt met instap- en afbreekregels, het is voor dergelijke tests moeilijk om het hele leeftijdsbereik qua betrouwbaarheid adequaat te bedienen. Plafondeffecten zullen zich dan ook vooral voordoen als men bijvoorbeeld hypothesen over hoogbegaafdheid wil toetsen bij de wat

oudere kinderen in het betreffende leeftijdsbereik (vijftien à zestien jaar), terwijl men vanwege bodemeffecten vooral in de problemen komt bij de jongere kinderen (zeven of acht jaar) wanneer men hypothesen over verstandelijke beperkingen wil onderzoeken. Ook al heeft men niet de beschikking over lokale betrouwbaarheden, toch valt meestal aan de normtabellen wel te zien waar de schoen wringt. Ter illustratie laten we in Tabel 1 de omzettingstabel van ruwe naar standaardcores voor een willekeurige subtest (getiteld 'Puzzels') uit een fictieve intelligentietest zien voor twee- tot zevenjarigen. De linkerkolom heeft betrekking op de jongste groep in het leeftijdsbereik, de rechterkolom op de oudste groep. Het voorbeeld is overigens representatief voor omzettingstabellen gehanteerd binnen bestaande intelligentietests.

TABEL 1 Bodem- en plafondeffecten geïllustreerd met een normtabel fictieve Intelligentietest.

RUWE SCORES SUBTEST PUZZELS LEEFTIJDSBEREIK 2 JR 3MND - 2 JR 5 MND	STANDAARDSCORE	RUWE SCORES SUBTEST PUZZELS LEEFTIJDSBEREIK 6 JR 9 MND - 6 JR 11MND
-	1	0-10
-	2	11-19
-	3	20-23
0	4	24-26
1	5	27-28
2	6	29-30
3	7	31-32
4	8	33
5-6	9	34-35
7	10	36
8-9	11	37
10-11	12	-
12-13	13	-
14-15	14	-
16-18	15	-
19-21	16	-
22-23	17	-
24-26	18	-
27-37	19	-

Bij de oudste groep is te zien dat het scorebereik van de ruwe scores (0-37) vrijwel in zijn geheel verdeeld is over de standaardcores 1 tot 10. Standaardcores hoger dan 11 kunnen voor deze subtest niet behaald worden. Er is dus sprake van een groot plafondeffect, zeker als een vergelijkbare situatie zich voordoet bij de andere subtests van deze intelligentietest (wat helaas meestal zo is). Voor de jongste groep zien we het omgekeerde, zij het (in dit fictieve geval) wat minder dramatisch. De standaardcores tot 4 kunnen in dit leeftijdsbereik niet behaald worden. Over de onderste helft van de standaardcores (1 tot en met 10) moest het scorebereik in de ruwe scores van 1 tot en met 7 verdeeld worden, terwijl

het resterende scorebereik (8 tot en met 37) over de bovenste helft van de standaardcores werd verdeeld (11 tot en met 19). Alle ruwe scores boven de 26 worden in dezelfde standaardcore van 19 omgezet. Het is eenvoudig te zien dat effecten zoals deze aanzienlijke consequenties hebben voor de betrouwbaarheid (nauwkeurigheid) bij een bepaald niveau van cognitief presteren. Zo zal een 'toevalligheidsfoutje' van een tweejarige peuter met een verstandelijke beperking bij deze puzzeltest (zijn aandacht dwaalt bijvoorbeeld even af) meteen gevolgen hebben voor zijn standaardcore en dus ook voor zijn uiteindelijke IQ, terwijl zelfs twee of drie van dit soort onbedoelde foutjes bij een slim kind van deze leeftijd geen consequenties hoeven te hebben voor de score. Helaas worden dit soort bodemeffecten dikwijls 'vermenigvuldigd', omdat ze bij meer dan een of zelfs alle subtests van een testbatterij aan de orde kunnen zijn.

Veel van dit soort problemen speelt, zoals hierboven al aangegeven, bij de toetsing van criteria voor toelating tot het speciaal onderwijs of het verkrijgen van leerlinggebonden financiering. Bij de uitwerking van deze criteria in de Tijdelijke Commissie Advisering Indicatiestelling Leerlinggebonden financiering (TCAI) heeft men zich uitgebreid laten adviseren door een werkgroep (WAI) die was samengesteld uit deskundigen van de NVO en het NIP (COTAN). In de rapportages die daaruit voortvloeiden (bijvoorbeeld TCAI, 2001) worden de problemen aangeduid. Naast een vermelding van sommige eerdergenoemde valkuilen, wordt hierbij ingegaan op het gebruik van intelligentietests die bestemd zijn voor een lager leeftijdsbereik, het daarmee samenhangende gebruik van referentieleeftijden en de problemen die daarbij optreden omdat referentieleeftijden vaak alleen op subtestniveau kunnen worden vastgesteld. Het is dan aan de diagnosticus om zelf een soort van gemiddelde of mediane referentieleeftijd te berekenen (zie ook Kaldenbach, 2007), die dan op zijn beurt weer overeen zou moeten komen met een bepaald deviatie-IQ voor de leeftijdsgroep waarin de geteste zich bevindt.

De genoemde problemen worden ernstiger naarmate de ernst van de verstandelijke beperking toeneemt. We hebben het hier dan over de categorieën matig (35-49), ernstig (20-34) en diep (< 20) verstandelijk beperkt, respectievelijk zwakzinnig (Resing & Blok, 2002). Men kan zich bij deze categorieën van cognitief functioneren überhaupt afvragen of een intelligentietest (op zichzelf) wel kan dienen als middel om het functioneren adequaat te beschrijven. Aan de assumptie van een normale verdeling die aan de basis ligt van het klassieke deviatie-IQ lijkt niet of nauwelijks te kunnen worden voldaan; wijlen A.D. de Groot zou waarschijnlijk eerder van een ziekte- dan van een normaal verdeeld model hebben gesproken (vergelijk De Groot, 1985). Het aantal verstandelijk beperkte mensen met een IQ van 35 of lager dat in Nederland jaarlijks wordt geboren is immers aanzienlijk hoger dan de 1 à 2 per circa 200.000 die men op basis van de normale verdeling zou verwachten.

Verschillen in IQ in deze regionen lijken ook nauwelijks te relateren aan de vraag welk onderwijs of onderwijsniveau iemand aankan. Binnen deze groepen verstandelijk beperkten lijkt de predictieve validiteit gering. Daar komt bij dat intelligentietests in de regel niet ontworpen zijn om IQ-waarden lager dan 55 (betrouwbaar) te meten (zie ook hierboven over bodemeffecten). Kinderen en volwassenen uit deze groepen zijn in de regel niet testbaar en maken over het algemeen ook geen deel uit van de normgroepen (terwijl dit overigens strikt genomen eigenlijk wel zou moeten om van een landelijk representatieve normeringssteekproef te kunnen spreken). Resing en Blok (2002) noemen het uitsluitend beschrijven

van mentale retardatie in termen van labels voor de genoemde gradaties van IQ dan ook niet zinvol. Men zou kunnen spreken van een virtueel of 'fictief' (Kaldenbach, 2007) IQ.

Zijn er oplossingen voor dit probleem?

Op de eerste plaats kan men (let wel: enkel bij oudere kinderen en volwassenen) denken aan het gebruik van intelligentietests voor een lager leeftijdsbereik en het bepalen van referentieleeftijden (zie boven). De kans is echter groot dat men bij zeer lage niveaus ook hiermee in de problemen komt. Gesteld dat iemand met een bepaalde test ook echt testbaar is, kan men eventueel uitwijken naar de methode die Kraijer en Plas (2006) aangeven, waarbij de geschatte ontwikkelingsleeftijd richtinggevend is en niet de kalenderleeftijd. Zij geven als uitbreiding op de tabel met leeftijdsequivalenten in de handleiding van de WISC-III-NL een tabel met equivalenten voor leeftijden tussen drie jaar twee maanden en vijf jaar tien maanden, waarbij zij gebruik hebben gemaakt van de methodiek van Smulders (1969). Kaldenbach (2007) bespreekt de procedure en komt tot de conclusie dat er op basis daarvan ook bij IQ-waarden van onder de 45 differentiatie kan plaatsvinden, maar stelt zich verder nogal terughoudend op: dergelijke IQ-bepalingen hebben hooguit de status van een vrij ruwe schatting.

Beter is het daarom wellicht om bij deze (verwachte) niveaus maar helemaal af te zien van het gebruik van intelligentietests en terug te grijpen op algemene ontwikkelingstests, zoals de BOS (Bayley Ontwikkelings Schalen) 2-30, of de nieuwere versie daarvan (BSID-II-NL; Bayley Scales of Infant Development, Second Edition, Nederlandse versie). De nieuwste versie van de Bayley wordt momenteel voor Nederland genormeerd. Soortgelijke tests zijn de DOS (Denver Ontwikkelings Screenings test), de GOS (Groningse Ontwikkelingsschalen) en de Kent Infant Development Scale (KID-N). Nadelen ervan zijn dat deze instrumenten soms niet op alle fronten aan de COTAN-eisen voldoen, dat niet altijd een indicatie kan worden verkregen van de mentale leeftijd of dat normen verouderd zijn. Vergelijk de inventarisatie van test- en schaalgebruik bij personen met ernstige meervoudige beperkingen van Penne e.a. (2007), die zich richten op IQ-scores lager dan 20 à 25. Het is illustratief dat zij in de categorie 'Intelligentie & algemene ontwikkeling' geen enkele intelligentietest bespreken. De richtlijnen van de TCAI (zie boven) hielden een dergelijk gebruik van ontwikkelingstests bij diepe en ernstige verstandelijke beperking nadrukkelijk open (in combinatie met gedragsobservaties).

Met het gebruik van deze tests komen we al dicht in de buurt van instrumenten die sociale redzaamheid pretenderen te meten, een criterium dat op de lichte en matige niveaus van verstandelijke beperking als additioneel wordt gehanteerd. Goede kandidaten zijn hier de Sociale Redzaamheidsschaal-Z (meer specifiek de SRZ-I en de SRZ-P; laatstgenoemde is geschikt voor de wat hogere niveaus binnen de range die hier aan de orde is en biedt mogelijkheden om een globaal mentaal niveau te bepalen) en de Vineland Adaptive Behavior Scales (meer specifiek de Vineland-Z en de Vineland-Screener). De gebruiker wordt in dit opzicht verwezen naar de overzichten die aangeven welke instrumenten in het wettelijk kader als geschikt worden aangemerkt (Resing e.a., 2008).

We komen hier tot slot op het punt dat het in alle opzichten beter is 'om te beschrijven wat het individu, ondanks zijn intellectuele handicap, nog wel kan' (Resing & Blok, 2002, p. 247). De auteurs geven daarom in een tabel aan wat van een individu in een

bepaalde categorie van verstandelijke beperking, gegeven zijn of haar leeftijd, verwacht mag worden. Voor het in kaart brengen daarvan bieden ontwikkelingstests of -schalen, schalen voor (sociale) redzaamheid en gedragsobservaties veel meer aanknopingspunten dan traditionele intelligentietests.

Noot

- 1 Beide auteurs zijn lid van de COTAN, de Commissie Testaangelegenheden Nederland van het Nederlands Instituut van Psychologen (NIP).

Literatuur

- Boxtel, H.W. van & Hemker, B.T. (2009). *Wetenschappelijke verantwoording van de Intelligentietest Eindtoets Basisonderwijs*. Arnhem: Cito.
- Carroll, J.B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. New York: Cambridge University Press.
- Dijk, H. van & Tellegen, P. (2004). *Nederlandse Intelligentietest voor Onderwijsniveau (NIO)*. Amsterdam: Boom test uitgevers.
- Flynn, R.J. (2007). *What is intelligence? Beyond the Flynn effect*. New York: Cambridge University Press.
- Groot, A.D. de (1985). Over algemene begaafdheid: Begrip, manifestaties, verdeling. In: F.J. Mönks, & P. Span (red.), *Hoogbegaafden in de samenleving*. Assen: Van Gorcum.
- Hendriksen, J.G.M. & Hurks, P.P.M. (2009). *WPPSI-III-NL Nederlandstalige bewerking: Technische handleiding*. Amsterdam: Pearson Assessment and Information bv.
- Kaldenbach, Y. (2007). Hiërarchische analyse van de wisc-III nader toegelicht: Vragen en antwoorden. *Kind en Adolescent Praktijk*, 2, 1-15.

Woord van dank

De auteurs bedanken de COTAN-leden, in het bijzonder Dr. J.B. Blok, voor hun waardevolle opmerkingen en bijdragen aan de totstandkoming van het manuscript.

- Kievit, Th., Tak, J.A., & Bosch, J.D. (red.), (2008). *Handboek psychodiagnostiek voor de hulpverlening aan kinderen*. Utrecht: de Tijdstroom.
- Kraijer, D.W. & Plas, J.J. (2006). *Handboek psychodiagnostiek en beperkte begaafdheid*. Amsterdam: Harcourt.
- Penne, A., Petry, K. & Maes, B. (2007). *Inventarisatie test- en schaalgebruik bij personen met ernstige meervoudige beperkingen*. Leuven: K.U., Multipius.
- Resing, W. & Blok, J. (2002). De classificatie van intelligentiescores: Voorstel voor een eenduidig systeem. *De Psycholoog*, 37, 244-249.
- Resing, W.C.M., Evers, A., Koomen, H.M.Y., Pameijer, N.K. & Bleichrodt, N. (2008). *Indicatiestelling speciaal onderwijs en leerlinggebonden financiering: Conditie en instrumentarium*. Amsterdam: Boom.
- Smulders, F.J.H. (1969). Het transporteren van testuitslagen. *De Psycholoog*, 4, 65-66/251-256.
- Sternberg, R.J. (1985). *Beyond IQ: A triarchic theory of human intelligence*. New York: Cambridge University Press.
- TCAI (2001). *Criteria voor indicatiestelling: In het kader van het wetsvoorstel Regeling leerlinggebonden financiering*. Zoetermeer: TCAI.