

Grote verschillen tussen
uitslagen van intelligentietests

Meten is geen weten

Hoogbegaafdheid is meer dan een hoge intelligentie, daar is iedereen wel van overtuigd. Maar voor toelating – of niet – tot voorzieningen voor hoogbegaafde leerlingen wordt wel degelijk goed gekeken naar de score op een IQ-test. Sommige voltijds-hb-klassen hanteren de 130-punten-grens behoorlijk strikt. Met 125 kom je niet snel binnen. De ontdekking die Judith Kertai deed tijdens haar diepgravende masteronderzoek voor de studie orthopedagogiek, roept grote vraagtekens op over dit toelatingsbeleid – en over de bruikbaarheid van intelligentietesten voor (vermoedelijk) hoogbegaafde kinderen. Wat blijkt: de RAKIT-2 ziet een kind (veel) eerder als hoogbegaafd dan de WISC-V.

TEKST: CARLA DESAIN

Onderzoek bij vermoedelijk meer- en hoogbegaafde kinderen; dat was het thema van de masterstage van Judith Kertai. Die stage liep zij bij Geke Maes, orthopedagoog en gespecialiseerd in dit soort onderzoek. Voor de intelligentie-component zet Geke de in Nederland meestgebruikte IQ-testen in: de (Nederlandstalige versie van de) WISC-V (spreek uit: WISC-5) en de RAKIT-2. Judith: 'Geke uitte geregeld haar zorgen over de uitslagen van beide testen. Bij de RAKIT-2 kwam er opvallend vaak de maximumscore 145 uit. Dat is vreemd, want een IQ van 145 wordt gekarakteriseerd als 'extreem hoogbegaafd' en zou voorkomen bij minder dan 1 op de 1000 mensen. Nou komen in Gekes praktijk kinderen bij wie

hoogbegaafdheid wordt vermoed, dus het is niet gek dat ze die score wat vaker krijgt. Maar zelfs met die slag om de arm kwam het veel vaker voor dan je zou verwachten. Die hoge RAKIT-score kwam niet altijd overeen met haar beeld van het betreffende kind. Bij de WISC-V was de afwijking andersom: kinderen die Geke 'typisch hoogbegaafd' leken, bij wie ze een IQ-score van 130 of meer zou verwachten, scoorden vaak maar 120-125.'

Onderzoeksvraag

Ook andere orthopedagogen en psychologen hadden dergelijke vraagtekens: 'Meten die twee testen wel hetzelfde, met name bij meer- en hoogbegaafde kinderen?'

'Dit leek me een interessante kluif voor mijn mas-



Judith Kertai (links) en Geke Maes
foto: Carla Desain

teronderzoek' vertelt Judith. Het is ook een relevante vraag, omdat veel scholen voor hun aanbod aan meer- en hoogbegaafde kinderen de lat leggen bij een IQ-score van 130. Als het klopt dat IQ-scores enorm verschillen, afhankelijk van welke test je gebruikt, zou je dus kinderen tekort kunnen doen die wel van dat aanbod zouden kunnen profiteren.'

Data verzamelen

'Ik heb er lang mee gepuzzeld hoe ik de die twee testen goed en betrouwbaar tegen elkaar kon afzetten', gaat Judith verder. 'Het liefst zou ik natuurlijk bij elk kind beide testen afnemen (in

willekeurige volgorde) en dan vergelijken hoe dat kind scoort. Maar dat was niet haalbaar: ik was nog niet afgestudeerd, dus nog niet bevoegd om die testen zelfstandig uit te voeren. Het zou anderen erg veel tijd kosten dat voor mij te doen. Daarom heb ik ervoor gekozen om bestaande testdata te gebruiken. Met een groep kinderen die met de RAKIT-2 zijn getest en een andere groep, getest met de WISC-V. Ik kreeg van 14

orthopedagogen en klinisch psychologen de uitslagen van de laatste (vermoedelijk hoogbegaafde) kinderen die ze hadden getest. Ik kreeg volledige – anoniem gemaakte – onderzoeksverslagen van 131 kinderen met een testleeftijd tussen 6 en 12

Judith Kertai begon aan verschillende studies, die ze niet afmaakte. Ze runde een webhost- en webdesign-bedrijf. Dat was redelijk succesvol, maar niet iets om de rest van haar leven mee bezig te blijven. Het viel haar op hoe weinig kennis over hoogbegaafdheid er is op scholen. 'Ik zag (en zie) daarin wel een rol voor mezelf. Dus ben ik alsnog orthopedagogiek gaan studeren aan de UvA. Mijn specialisatie in hoogbegaafdheid regelde ik via stage en onderzoek in de praktijk van Geke Maes, 'Gekend Talent' in Meppel.'

Judith studeerde op 30 september af als master in de orthopedagogiek. Ze kreeg ze een 8,6 voor haar scriptie *Intelligentie gemeten met verschillende testen; Verschil in Intelligentiequotiënt wanneer deze gemeten wordt met de WISC-V en de RAKIT-2*. Deze is binnenkort te downloaden vanaf www.scriptiesonline.uva.uva.nl

Judith Kertais eigen praktijk 'Almeerbegaafd' is in de opstartfase.

We doen al jaren alsof diverse intelligentietests hetzelfde meten, wat nu dus niet zo blijkt te zijn. Het lijkt het erop dat groepen kinderen structureel te hoog of te laag worden ingeschat - met alle gevolgen van dien.'

jaar. Keurig verdeeld over beide testen: 65-66.

Ook de verdeling in leeftijd en jongen-meisje was redelijk netjes. Om uit te sluiten dat bijvoorbeeld alle verlegen kinderen de (wat minder talige) RAKIT krijgen, heb ik gecontroleerd of de aanmeldreken al voorselecteert in de keus voor een van de twee testen. Dat bleek niet zo te zijn.'

Testen vergelijken

'Nu wordt het verhaal een beetje ingewikkeld; maar het is wel interessant om mijn onderzoeksrapport stap-voor-stap te beschrijven. Beide testen onderzoeken in grote lijnen dezelfde intelligentiefactoren. Daarnaast heeft elke test iets dat de ander niet heeft: de RAKIT onderzoekt het langetermijngeheugen en de WISC kwantitatieve kennis en verwerkingssnelheid', legt Judith uit.

'Die vier overlappende factoren heb ik bekeken. Bijvoorbeeld of je met de subtests die de visuele informatieverwerking meten, op dezelfde score uitkomt – of je nu de ene test gebruikt of de andere. Ik heb nauwelijks verschil gevonden als ik de naar

de opgetelde scores (somscores) keek. Maar als ik die somscores volgens de bijbehorende tabellen omzette naar een IQ-score, bleek de WISC beduidend lager uit te komen dan de RAKIT – heel vreemd.’

Omzettingstabellen analyseren

‘Het zou natuurlijk zo kunnen zijn dat er toch iets raars aan de hand was met de dataset van mijn onderzoeksgroep’, geeft Judith toe, ‘al was die nog zo zorgvuldig samengesteld. Mijn begeleider bracht me op het idee om de omzettingstabellen uit de scoringshandleidingen van beide testen te analyseren. Daar heb ik hetzelfde gedaan: weer somscores omgezet naar IQ-scores en die met elkaar vergeleken. Hier zie je hetzelfde rare patroon. Bij kinderen die lager dan gemiddeld scoren, gaan beide testen gelijk op. Maar vanaf het gemiddelde (= een IQ van 100) lopen de testuitslagen uiteen. Bij een identieke prestatie, bij dezelfde afwijking

verwerkingssnelheid. Of wellicht had het verschil iets te maken met de persoon van de testafnemer... Maar het verschil blijkt dus met name te zitten in de omzettingstabellen. Heel vreemd. Kennelijk zijn die tabellen niet eerder grondig vergeleken of geijkt. Terwijl beide testen – en beide scoringshandleidingen met daarin de bewuste omzettingstabellen – op de markt worden gebracht door dezelfde uitgever.’

Verstreckende gevolgen

Geke Maes mengt zich in het gesprek: ‘Deze onderbouwing van Judith is heel belangrijk voor het werkveld. Het begon bij het gevoel van mij en verschillende collega’s dat er iets niet klopte. Meerdere kleuters achter elkaar die de plafondscore van 145 haalden met de RAKIT-2; en ook ‘typisch hoogbegaafde’ kinderen die niet boven de 120-125 uitkwamen met de WISC-V. Je kunt dan als orthopedagoog wel zeggen: ‘Dat geloof ik niet’, maar het zijn genormeerde tests. Je moet het daarmee doen. Ik heb in 2016 mijn zorgen over die hoge scores van de RAKIT geuit bij Pearson, de uitgever. Ik kreeg een mail terug dat ze dit geluid niet eerder gehoord hadden, maar het wel serieus namen. Maar tot nu toe is er niets veranderd.’

De IQ-scores wijken vooral sterk af in hogere regionen. Vergelijkbare prestaties kunnen leiden tot een verschil van 7 tot wel 13 IQ-punten. Waarbij de RAKIT-2 telkens hogere IQ-scores berekent.

ten opzichte van het gemiddelde, kennen de beide omzettingstabellen verschillende IQ-scores toe. De WISC-V berekent een lager IQ dan de RAKIT-2. En niet een luttel puntje verschil... Een kind dat op de WISC uitkomt op een IQ-score van 130, krijgt van de RAKIT een score van 143. Waar de RAKIT 130 zegt, zegt de WISC 123. Nogmaals: het gaat hier om kinderen die presteren op eenzelfde afwijking van het gemiddelde.’

Conclusie

Judith concludeert: ‘Het verschil dat Geke en haar collega’s al vermoedden, is er dus echt: een kind moet – door iets raars in de omzettingstabellen – meer van het gemiddelde afwijken om ook door de WISC-V als hoogbegaafd te worden gezien. Bij de start van mijn onderzoek zochten we de verklaring van het in de praktijk gevoelde verschil vooral in inhoudelijke verschillen tussen de tests. Bijvoorbeeld in het verbalere karakter van de WISC of in het gewicht dat deze test toekent aan

‘Het verschil in uitslagen blijkt nu echt niet alleen te bestaan in mijn waarneming, maar verklaard te kunnen worden uit de omzettingstabellen. Als dit inderdaad klopt, vind ik dat een schokkende ontdekking. Als tester moet je ervanuit kunnen gaan dat de testen die in omloop zijn, vergelijkbare uitkomsten geven. Het mag voor het resultaat niet uitmaken welke je kiest. Judith vergelijkt het wel eens met thermometers: als je er twee hebt, ga je ervanuit dat ze in dezelfde situatie allebei aangeven dat een kind koorts heeft – of juist allebei niet. Anders heb je er toch niets aan? Maar als bij twee even intelligente kinderen verschillende testen worden afgenomen, krijgt het ene kind een andere IQ-score toebedeeld dan het andere. Dat verschil kan verstrekkende gevolgen hebben: op basis van die IQ-score wordt bijvoorbeeld het ene kind wel toegelaten tot het voltijd hb-onderwijs en het andere niet. ‘ Judith: ‘Als vervolg zou ik graag alsnog mijn eerste onderzoeksidee uitvoeren: een grote groep vermoedelijk hoogbegaafde kinderen testen met zowel de WISC-V als de RAKIT-2. Ik ben benieuwd of dat onderzoek vergelijkbare resultaten laat zien.’

Maatschappelijke onrust

Geke: ‘Ik weet dat niet iedereen het ermee eens is dat dit artikel nu geschreven wordt. ‘Waarom maatschappelijke onrust creëren als dit nog niet voldoende onderbouwd is?’ hoor ik dan. Deze

bezorgdheid deel ik; voorzichtigheid en nader onderzoek vind ik dan ook absoluut op zijn plaats. Anderzijds doen we al jaren alsof diverse intelligentietests hetzelfde meten, wat nu dus niet zo blijkt te zijn. Het lijkt het erop dat groepen kinderen structureel te hoog of te laag worden ingeschat. Dit heeft maatschappelijke, maar zeker ook persoonlijke gevolgen die niet altijd goed zijn. En met deze gevolgen hebben wij in de testpraktijk dagelijks te maken. Deze kinderen kunnen gewoonweg niet een paar jaar wachten – tot nader wetenschappelijk onderzoek is afgerond. Zij hebben NU een passend onderwijsaanbod en begeleiding nodig.’

Boodschap voor scholen

Zowel Judith Kertai als Geke Maes benadrukken hoe belangrijk het is om verder te kijken dan de IQ-score. Intelligentie bepaalt maar voor een deel of een kind hoogbegaafd is. Ook de manier van denken is belangrijk, net als de zijnskenmerken als gevoeligheid, intensiteit, perfectionisme en rechtvaardigheidsgevoel. Nu blijkt uit dit onderzoek bovendien dat een IQ-score niet zo eenduidig is als je zou verwachten, je weet niet precies wat een score inhoudt. De beslissing over het onderwijsaanbod aan een kind kun je dus beter niet te strikt af laten hangen van de IQ-score.

Foto: Jeroen Oerlemans



REACTIE VAN PETER DE JONG, HOGLERAAR ORTHOPEDAGOGIEK, MET SPECIALISATIE ONDERWIJSLEERPROBLEMEN, AAN DE UvA; BEGELEIDER VAN DE MASTERSCRIPTIE EN HET -ONDERZOEK VAN JUDITH KERTAI

Nader onderzoek vereist
‘Ik voel me wat ongemakkelijk bij deze publicatie. Eerlijk gezegd vind ik het te vroeg voor een artikel als dit. Het heeft er alle schijn van dat we in dit onderzoek wel degelijk een probleem te pakken hebben. Maar vóór we daar echt stellige uitspraken over kunnen doen, moeten er nog wat stappen ondernomen worden. Ik had liever

eerst de discussie over deze onderzoeksaanpak en de implicaties ervan willen voeren met wetenschappelijke vakgenoten en mensen die gepokt en gemazeld zijn in testnormering. Ik vind het nog te vroeg om opschudding te veroorzaken.’

Verschillende uitslagen

‘We weten al langer dat bij ongeveer een derde van de mensen de uitslagen van verschillende IQ-testen wel 5 à 10 punten kunnen verschillen. We dachten tot nu toe dat die verschillen min of meer toevallig

tot stand kwamen: of iemand wel of niet z’n dag heeft, of de test beter bij de persoon past, of dat er een betere klik is met de afnemer van de test. Uit dit onderzoek blijken de verschillen systematisch te zijn. Ik denk dat de wortels ervan liggen in het normeringsproces. Dat wil niet zeggen dat de normering van de ene test of de andere fout is – al is het wel raar en onwenselijk dat bij RAKIT-2 al zo snel de plafondscoor van 145 bereikt wordt. Binnen het proces van normering moeten beslissingen genomen worden, daar zit een aantal vrijheden in. Bijvoorbeeld of je eerst de subtests normeert en dan optelt, of eerst alles optelt en dan normeert. Daardoor kan het zo zijn dat – ondanks dat het proces van normeren op zich netjes is gedaan – die twee normeringen toch tot andere uitkomsten leiden. Met andere testuitslagen als gevolg. Dat betekent niet per se dat er echt fouten zijn gemaakt. Wat ik wel verwijtbaar vind aan de testuitgever is dat in de technische handleiding niet beschreven is welke beslissingen er genomen zijn in dat normeringsproces. Zo is dit proces niet goed controleerbaar.’

Hoe nu verder?

‘Ik vind dat er nu snel een wetenschappelijke discussie gevoerd moet worden over het normeren van dergelijke intelligentietests. Met name op de extremen, de onderste en bovenste 2%, is de normering vanzelfsprekend maar op weinig kinderen gebaseerd. Dat kan vertekening geven. Pearson, de uitgever van zowel de RAKIT-2 als de WISC-V, moet mijns inziens de roep uit het veld en de uitslag van dit onderzoek heel serieus nemen. De onderzoeksteams van beide testen zouden grondig moeten duiken in de mogelijke verschillen in hun normeringsprocessen. En kijken of die gelijk getrokken kunnen – en moeten – worden.’

Professor Peter de Jong: ‘Het heeft er alle schijn van dat we in dit onderzoek wel degelijk een probleem te pakken hebben.’



HANS BUFFART, WAS HOGLERAAR INFORMATIEKUNDE VOOR DE SOCIALE WETENSCHAPPEN IN GRONINGEN EN GASTHOGLERAAR ALGEMENE PSYCHOLOGIE IN LEIPZIG

Voorzichtig omgaan met testcores

'Ik vind het een zeer boeiende en lovenswaardige scriptie. Voor zover ik kan beoordelen heeft mevrouw Kertai methodisch keurig gewerkt. Ik deel de mening van professor De Jong dat een publicatie vooralsnog voorzichtig moet worden aangegaan. De teneur en de titel van deze publicatie zouden vooral moeten zijn dat nader onderzoek nodig is. En dat er een indicatie is dat het verschil tussen beide testen – zoals dat in de praktijk wordt ervaren – ligt in het vertalen van de resultaten van de onderliggende testen naar het intelligentiequotiënt. Het onderzoek maakt duidelijk dat er een verschil is, maar niet waardoor het veroorzaakt

REACTIE VAN PEARSON, UITGEVER VAN BEIDE IQ-TESTEN UIT HET ONDERZOEK

Een IQ-test is een diagnostisch hulpmiddel

'We vinden het onderzoek van Judith Kertai interessant. Wel is het volgens ons lastig om de resultaten te generaliseren – vanwege een aantal methodologische keuzes en praktische beperkingen. Wij zijn er van overtuigd dat beide tests een betrouwbare en valide mogelijkheid bieden om het IQ van kinderen binnen de beoogde IQ-bereiken in kaart te brengen en dat de door ons bepaalde normen juist zijn. De betrouwbaarheid en validiteit van beide tests wordt ondersteund door onderzoek en wordt bevestigd door de positieve COTAN-beoordeling.'

Normeringsmodellen

'De normeringsmodellen voor beide tests zijn gebaseerd op de verzamelde representatieve steekproeven en berekend op basis van een continue normeringsmodel. Bij de WISC-V-NL is ook gekeken naar de normen van de test in andere landen. Bij het normeren van een test wordt (historisch gezien) nooit gekeken of vergeleken met de normen van andere tests. Zover ons bekend is hier ook nog geen goede methodologie voor.'

RAKIT-2 is vooral geschikt voor de lagere regionen

'Bij de ontwikkeling van de RAKIT-2 is gefocust op makkelijkere items en op uitbreiding van de geschiktheid voor de lagere IQ's. Hierdoor is de RAKIT-2 bij uitstek geschikt voor het testen van kinderen met IQ's onder het gemiddelde. Deze informatie wordt beschreven in de handleiding, is gedeeld op LinkedIn en wordt besproken in onze RAKIT-2 trainingen en in ons RAKIT-2 / WISC-V webinar dat voor iedereen toegankelijk is. Wij vinden het raadzaam om bij een vermoeden van hoge intelligentie of hoogbegaafdheid (en zeker vanaf de leeftijd van 11 jaar) te kiezen voor een andere Nederlandstalige IQ-test zoals de WISC-V. De RAKIT-2-test is wel in staat om kinderen met een hoog IQ te identificeren, maar de differentiatie binnen de hoge IQ-range is hierin minder goed. Bovendien heeft de RAKIT-2 een bovengrens van 145, de WISC-V van 160.'

Alleen een IQ-score is niet voldoende

'Discussies over de meetpretentie en normen van intelligentietests zijn niet nieuw en hebben eerder ook plaatsgevonden. Een IQ-test is een diagnostisch hulpmiddel en het is daarom ook

Slim!

Pas op voor het woord 'slim'.

Natuurlijk, iedereen vindt het fijn om slim genoemd te worden. Het is iets om trots op te zijn, toch? Maar stel nou dat je voortdurend hoort dat je slim bent? Vooral wanneer je bijzondere prestaties neerzet, het liefst zonder al teveel moeite? Dan zou je weleens het idee kunnen krijgen dat je moet bewijzen hoe slim je wel niet bent.

En hoe bewijs je dat je slim bent? Door als eerste klaar te zijn met een opdracht, door altijd het antwoord te weten, door de beste te zijn, door foutloos werk af te leveren. Dat klinkt op zich vrij onschuldig, maar wanneer je dit soort verwachtingen ontwikkelt ten opzichte van jezelf, zou dat je weleens kunnen gaan belemmeren.

Het kan er namelijk toe leiden dat je kansen om te groeien laat liggen uit angst om het label 'slim' kwijt te raken. Dat je gaat kiezen voor de veilige weg, die bestaat uit situaties en opdrachten waarin je zonder twijfel zult uitblinken.

Laten we het woord 'slim' daarom eens wat slimmer gaan gebruiken. Niet gerelateerd aan een eigenschap die allerlei verwachtingen met zich meebrengt, maar gekoppeld aan acties die je kunt ondernemen wanneer je uitgedaagd wordt. Je kunt problemen analyseren, een plan van aanpak maken of hulp vragen als dat nodig is. Je kunt even opnieuw beginnen, iets doen wat je nog niet eerder geprobeerd hebt of leren van een tactiek die niet werkt. Kortom, je kunt je richten op aanpak, vaardigheden en strategieën.

Je hoeft niet te bewijzen dat je slim bent, maar je kunt wel slimme dingen doen om iets (beter) te laten lukken. Zo. Die kan op een tegeltje.

wordt. Nieuw onderzoek moet uitwijzen waardoor het verschil in omzetting veroorzaakt wordt en of men beide testen kan ijken, dan wel onderling vergelijkbaar kan maken.'

'In ieder geval is een oproep op zijn plaats aan degenen die testcores gebruiken om daarmee voorzichtig om te gaan. Nogmaals, wetenschappelijke resultaten komen tegenwoordig vaak te vroeg in de algemene publiciteit. Dat heeft maatschappelijke gevolgen die niet altijd goed zijn. Het is niet uitgesloten dat nadere studie andere onderscheidende factoren aan het licht zal brengen.'

altijd noodzakelijk dat een volledig geïntegreerd beeld wordt gevormd. Alleen de uitslag op een IQ-test is onvoldoende om een hypothese te ontcrachten of bevestigen. Het is goed dat het werkveld zich bezighoudt met dit soort vraagstukken. Het opstellen van richtlijnen over hoe om te gaan met grenswaarden voor hb-onderwijsaanbod is – naar ons idee – aan het praktisch werkveld, zoals dat ook in het buitenland het geval is. We zijn uiteraard bereid om daarover mee te denken.'

Vervolgonderzoek zou informatief zijn

'Er staat nog geen vervolgonderzoek gepland, maar we gaan kijken of het noodzakelijk is om dat op te zetten. Wij staan open voor een discussie met wetenschappelijke vakgenoten, met name statistici, om te zien of de manier van vergelijken van normen in het onderzoek van Judith Kertai een goed beeld geeft en wat de implicaties hiervan zouden kunnen zijn. We hopen hiermee spoedig te kunnen starten.'

Het genoemde webinar is hier te vinden:
www.pearsonacademy.nl/intelligentiemeting-met-de-wisc-v-nl-en-rakit-2-nl

Floor Raeijmaekers

Floor Raeijmaekers is Specialist in Gifted Education (ECHA) en eigenaar van Het TalentenLab. Ze geeft trainingen over hoogbegaafdheid en mindset en ze coacht hoogbegaafde kinderen. Floor is de bedenker van Fixie & Growie en samen met haar dochter schreef ze het boek 'Hartstikke Hoogbegaafd!'.

