

## RAPPORTERING VAN TESTPRESTATIES MET EENDUIDIGE KWALITATIEVE LABELS Consensus statement van het Vlaams Forum voor Diagnostiek (VFD)

*Jules, een jongen van 8 jaar, wordt aangemeld voor een multidisciplinair diagnostisch traject. Op basis van de aanmelding en anamnestiche gegevens, vermoedt de arts de aanwezigheid van een complexe ontwikkelingsstoornis. Binnen dit kader, vindt er een uitgebreid diagnostisch protocol plaats, waarbij verschillende disciplines deelnemen. Elke discipline maakt gebruik van diagnostische testen om de prestaties van Jules op verschillende ontwikkelingsdomeinen in te schatten.*

*De psycholoog brengt een aantal cognitieve functies in kaart, met inbegrip van intelligentie aan de hand van de WISC-V-NL. De logopediste maakt gebruik van de CELF-5-NL om de taal in te schatten en de kinesitherapeut neemt de Movement Assessment Battery for Children-2 af om te screenen op motorische problemen.*

*Jules behaalde volgende resultaten: WISC-V-NL Totaal IQ = 91 (gemiddeld), Verbaal Begrip Index = 82 (laag gemiddeld), Visueel Ruimtelijk Index = 98 (gemiddeld), Fluïde Redeneer Index = 92 (gemiddeld); CELF-5 kernscore = 78 (laag), Movement Assessment Battery totale standaardscore 5, pc 5 (rode zone, duidelijke motorische problemen).*

Bovenstaande metingen gebeuren in kader van diagnostiek, om in kaart te brengen wat er bij Jules precies aan de hand is. Op deze manier kunnen tijdens het begeleidingstraject de behandeldoelen geïdentificeerd worden, evenals een indicatiestelling voor de intensiteit van een bepaalde therapie (nl. wat is prioritair in het multidisciplinair behandelaanbod en hoe intensief moet de therapie zijn?), en om te zien of de behandelingen het gewenste effect hebben.

Naast psychometrische kwaliteiten van diagnostisch instrumenten (vb. betrouwbaarheid, validiteit, normgroep), verdient ook de beschrijving van de resultaten de nodige aandacht. Resultaten worden gecommuniceerd naar de cliënt plus zijn omgeving en naar collega's, hetgeen zowel mondeling als schriftelijk kan. **Hierbij is het belang van uniforme en heldere communicatie evident.**

Deze evidentie zien we in de praktijk helaas niet terug. Er is namelijk nog geen consensus over de kwalitatieve beschrijving van testresultaten. Momenteel worden gelijke prestaties niet altijd op dezelfde manier verwoord. Deze verschillen merken we niet enkel op tussen de verschillende disciplines, maar zelfs binnen één en dezelfde discipline hanteren collega's andere termen voor dezelfde prestatieniveaus op een bepaalde taak. We zien zelfs dat binnen een ziekenhuis of dienst de termen niet door iedereen op dezelfde wijze worden gebruikt:

*De psycholoog op de diagnostische dienst omschrijft een percentielscore van 7 als "benedengemiddeld", terwijl een collega van het revalidatieteam die deze client 6 maanden later ook test, een percentielscore van 7 als "laag" omschrijft. Hierdoor lijkt het voor de andere teamleden alsof er een verandering heeft plaatsgevonden, terwijl de prestatie eigenlijk gelijk is gebleven.*

Daarnaast komt het voor dat soms testcores niet bij het verslag worden toegevoegd. In dit geval is een gelijke kwalitatieve omschrijving extreem belangrijk om bijvoorbeeld verschillen over de tijd te kunnen detecteren (bijv. ikv dementie-onderzoek).

Kortom, het zou **voor de klinische praktijk een meerwaarde zijn om een set descriptieve (beschrijvende) labels te hanteren om kwantitatieve gegevens, kwalitatief te omschrijven en die door collega's binnen én over disciplines heen hetzelfde zijn.** Dit opdat we allemaal dezelfde taal hanteren. Een taal die eenduidig, toegankelijk en voor eenieder begrijpelijk is. Dat is belangrijk voor de cliënten, verwijzers en alle andere professionals die bij de diagnostiek en behandeling betrokken zijn.

Het is natuurlijk absoluut geen nieuwe kwestie die hier aangehaald wordt. Ook is het niet enkel een probleem in Vlaanderen. Zo formuleerde de American Academy of Clinical Neuropsychology in 2018



specifieke aanbevelingen in het 'Consensus Conference Statement' voor het toepassen van uniforme labels voor prestatietests. Daarnaast was er in Nederland parallel aan onze werkgroep, een discussie lopende of ze de reeds in 2012 geformuleerde labels door Bouma en collega's (2012) of de Amerikaanse labels die de American Academy of Clinical Neuropsychology in 2018 hebben aanbevolen, die Hendriks en collega's vertaalden naar het Nederlands, zouden gebruiken.

Vanuit het **Vlaams Forum voor Diagnostiek (VFD)** is er een **werkgroep "Descriptieve labels"** samengesteld om deze kwestie voor Vlaanderen tegen het licht te houden en een advies te formuleren aan het werkveld. In deze werkgroep hebben diagnostici uit verschillende disciplines (logopedie, kinesitherapie en psychologie) en verschillende settings (zowel onderzoekers, docenten als klinici) samen gezocht naar het classificatiesysteem dat zowel statistisch als taalkundig correct is, en vooral ook in de klinische praktijk zo werkbaar en helder mogelijk is, zonder dat we nog een classificatiesysteem toevoegden aan de lijst van reeds bestaande. Uniformiteit tussen werkvelden en onderwijsinstellingen speelde bijgevolg mee bij deze keuze, gebaseerd op input uit het veld van logopedie, (neuro)psychologie & kinesitherapie. Om dit te bekomen, werd eerst een oplistings gemaakt van de meest voorkomende manieren om normaalverdeelde scores te beschrijven in de verschillende werkvelden, gecombineerd met wat in de (internationale) literatuur aan bod komt. Er werd vervolgens gekeken naar labels die eventueel tot misverstanden konden leiden omdat ze in verschillende systemen voor andere categorieën werden gebruikt, om daarna het taalkundige aspect vanuit Vlaamse bril te bekijken. Als resultaat ontstond een consensus over een **uniforme set van labels voor de kwalitatieve beschrijving van kwantitatieve prestaties** op prestatietaken, waarvan de **normen een normaalverdeling** benaderen, die gebruikt kunnen worden door de verschillende disciplines. Dit is een advies van het VFD op basis van de huidige informatie.

*Let op, dit zijn beschrijvende labels op het niveau van de geleverde prestatie op de taak, het zijn **geen klinische labels**. Het beschrijven van testprestaties (bijvoorbeeld is dit een (beschrijvende) lage of hoge score vergeleken met de normgroep waarvan de scores nagenoeg normaal verdeeld zijn) is niet hetzelfde als het interpreteren van onderzoeksresultaten (bijvoorbeeld dit resultaat geeft een subklinische, klinische, afwijkende, of significant lagere of hogere score weer in vergelijking met de normgroep). In het gebied van intelligentie bijvoorbeeld, is een score van 87 in het laaggemiddeld gebied (zie verder) maar is deze score niet significant lager (-1SD) dan gemiddeld omdat het niet lager is dan 85 (gemiddeld IQ van 100 – standaarddeviatie van 15). Het behalen van een slechte prestatie op een taak staat daarbij niet gelijk aan een volledig juiste inschatting van de te meten eigenschap of eventueel stoornis. Dit betekent dat onderstaand advies niet van toepassing is voor klinische interpretaties, vaak gebaseerd op verschillen in standaarddeviaties ten opzichte van het gemiddelde van de normgroep of cutoff waarden in de praktijk. Het doortrekken van uniformiteit naar klinische interpretatie is de volgende stap waar de werkgroep nu nog mee aan de slag is, en waar er dus nog adviezen over zullen volgen. **De huidige richtlijn is enkel en alleen gericht op het hanteren van de beschrijvende labels van de testresultaten.***



**Consensus statement van de werkgroep “descriptieve labels” van het Vlaams Forum voor Diagnostiek voor beschrijvende labels voor scores op prestatietests bij instrumenten waarop een normaalverdeling van toepassing is.**

Deze is een **advies** van het VFD op het moment dat **protocollen en/of handleidingen geen, onvoldoende, onjuiste of tegenstrijdige informatie bieden, of** wanneer het gaat om **verschillende diagnostische instrumenten naast elkaar** te bespreken, ook over disciplines heen.

*Opgelet: de beschrijvingen in Tabel 1, zijn enkel te hanteren bij instrumenten waarop een normaalverdeling van toepassing is.*

Tabel 1. Consensus statement beschrijvende labels bij normaalverdeling

Gebaseerd op Bouma e.a. (2012); Ruiters e.a. (2017) *	AU (Afgeleide uitslag)					%in categorie
	T-score	IQ/Indexscore	z-score	Percentiel		
Zeer hoge score	70 -...	130 -...	16 - ...	2.00 tem...	≥98 (98-...)	2,33
Hoge score	64 - 69	120 - 129	14 - 15	1.33 tem 1.99	<98 (91 - 97)	6,96
Hooggemiddelde score	57 - 63	110 - 119	12 - 13	0.67 tem 1.32	<91 (75-90)	16,28
Gemiddelde score	44 - 56	90 - 109	8 - 11	-0.67 tem 0.66	<75 (25-74)	49,50
Laaggemiddelde score	37 - 43	80 - 89	6 - 7	-1.33 tem -0.68	<25 (9-24)	15,97
Lage score	30 - 36	70 - 79	4 - 5	-2.00 tem -1.34	<9 (3-8)	6,74
Zeer lage score	... - 29	... - 69	... - 3	... tem -2.01	<3 (...-2)	2,22
M	50	100	10	0		
SD	10	15	3	1		

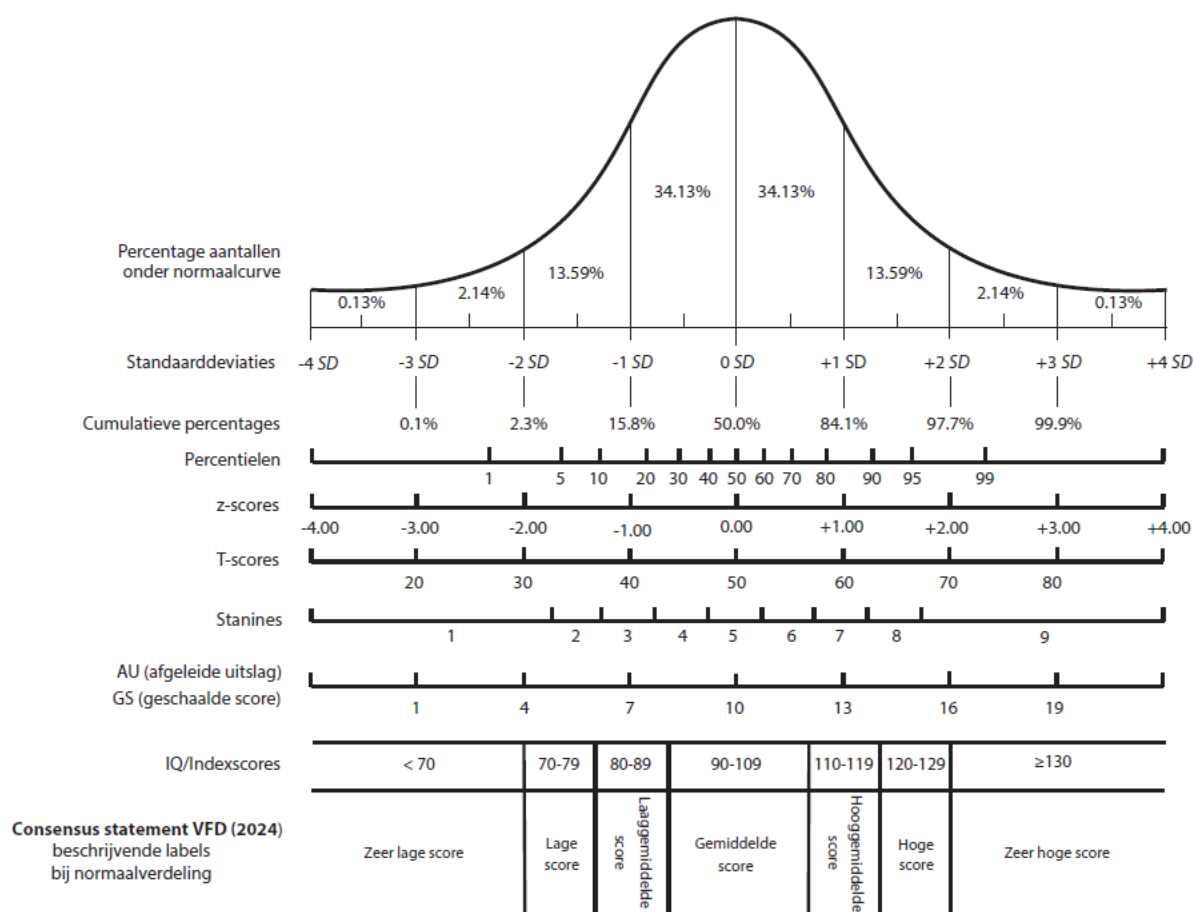
*\*Deze consensus statement sluit aan bij en is gebaseerd op het voorstel van Bouma (2012). Zoals reeds geduïd, gaat het over een **beschrijving van scores of prestatie** op een taak, we kiezen er daarom voor om het woord score expliciet in de labeling op te nemen. **Deze labels zijn niet gericht op het doen van klinische uitspraken.***

Het classificatiemodel is opgedeeld in **7 intervallen** van ongeveer een standaarddeviatie van 0.7 zoals we ook kennen van de zevenpuntschaal die gehanteerd wordt in de Wechslerschalen. Gezien het wereldwijde gebruik van de Wechslerschalen stellen we voor deze als uitgangspunt te nemen. Naast het feit dat deze verdeling mooi aansluit bij de reeds gangbare praktijk, zijn deze 7 intervallen verkozen boven het hanteren van 5 intervallen zodat er meer verfijning kan worden aangebracht in de classificatie binnen de normaalverdeling. *De 5 categorieën worden wel frequent gehanteerd bij het aanduiden van klinische interpretaties die dus niet binnen deze huidige beschrijving vallen. Deze kunnen bij het verder uitwerken van uniformiteit binnen de klinische interpretatie, in acht worden genomen.* In het huidige model met 7 intervallen komen de percentielen overeen met de effectieve waarden onder de Gauss-curve, de gemiddelde zone representeert ongeveer 50% van de normgroep (dit is minder dan 1 SD afwijkend van gemiddeld, en anders dan bij vele klinische interpretaties, bijv. een IQ van 85 tot 115).

De grenzen van de intervallen zijn steeds zo vastgelegd dat de percentielwaarden overeenkomen met de effectieve waarde onder de normaalverdeling van de Gauss-curve. Zo werd bij IQ/indexscores het interval “90 - 109” verkozen als gemiddelde score boven “90 - 110”, waarbij 110 nog binnen dit interval opgenomen wordt, omdat hierdoor de Gausscurve het best benaderd wordt. In feite omvat “90 - 109”, alle personen uit de normgroep die omgezet een normscore vanaf 90 tot 109.99 scoorden, wat afgerond 50% van de normgroep is. Bij “90 - 110” zouden alle personen die een 110.99 halen ook in deze categorie vallen waardoor deze categorie te groot wordt, namelijk



afgerond 52% van de normgroep. Deze redenering geldt voor alle categorieën die hoger dan gemiddeld zijn of aan de rechterkant van de Gausscurve vallen (110 - 119; 120 - 129; 130 - ...).



In de consensus statement is er gekozen voor de termen

- **“Zeër lage score”/ “Zeër hoge score”** boven de termen “Uitzonderlijk laag”/ “Uitzonderlijk hoog” omdat deze als minder veroordelend of als een klinische vaststelling worden gezien. Zoals Schmand en Bouma (2020) ook aanhalen worden in de Amerikaanse cultuur waaruit deze termen (“Uitzonderlijk laag”/ “Uitzonderlijk hoog”) vertaald werden deze misschien niet als veroordelend ervaren maar is dat in Vlaanderen minder neutraal dan de labels ‘zeer laag’ en ‘zeer hoog’.
- **“Lage score”/ “Hoge score”** in plaats van door Hendriks (2020) voorgestelde “Benedengemiddeld”/ “Bovengemiddeld” omwille van een verwarring in de praktijk doordat deze zelfde termen (“Benedengemiddeld”/ “Bovengemiddeld”) binnen intelligentieonderzoek voor andere scores horend bij de Gausscurve wordt gebruikt (Resing, 2002). Bovendien is het model van Hendriks vertrokken uit een letterlijke vertaling van de American Academy of Clinical Neuropsychology. Hetgeen zorgt voor een goede aansluiting bij de internationale literatuur hetgeen in een wereld die zo globaliseert, de communicatie over landsgrenzen heen vergemakkelijkt. Echter, in Vlaanderen klinkt “benedengemiddeld” minder laag dan 'laaggemiddeld', terwijl het in de methode van Hendriks (2020) wel als lager bedoeld wordt. Daarom adviseren we in het Vlaams “lage score”/“hoge score” te hanteren en internationaal de Engelse termen te gebruiken.



Deze termen komen ook overeen met die zoals gehanteerd in de meeste Wechslerschalen in het Nederlands taalgebied; WAIS-IV-NL en WPPSI-IV-N, en de vorige versie van de WISC.

**In Nederland is door de Sectie Neuropsychologie van het Nederlands Instituut van Psychologen ondertussen dezelfde beslissing gevallen en zij zullen parallel met onze communicatie een monodisciplinaire richtlijn publiceren waarin ook zij adviseren om de beschrijving van Bouma et al te hanteren.**

We beseffen dat het niet makkelijk is om het systeem waar je nu mee werkt, los te laten. Wanneer er een professionele uitspraak wordt gedaan over iemand, moet deze kristalhelder zijn en mag er geen misverstand over kunnen bestaan, en is uniformiteit een noodzaak. Het voordeel van het hanteren van interdisciplinair uniforme labels is dat we een gemeenschappelijke taal creëren die communicatie tussen collega's en naar cliënten transparant maakt. Het VFD onderschrijft daarom het belang van een eenduidig systeem en adviseert om deze labels te gebruiken, ook als je je mogelijk niet helemaal in deze "consensus statement" kunt vinden. Bij testen met een normale verdeling is deze indeling theoretisch correct. Dit zou bijgevolg moeten terug te vinden zijn in de handleiding van een test met normaal verdeelde normscores. Indien de handleiding hiervan afwijkt (bijv. andere grenzen, terminologie), dien je kritisch na te gaan wat hiervoor de reden is. Om de helderheid in communicatie nog te vergroten, raden wij aan om de tabel waarin de betekenis van de labels wordt toegelicht, ook steeds mee op te nemen in het verslag.

Dit schrijven is de eerste stap richting uniformiteit in de communicatie binnen de diagnostiek, maar zeker nog niet de laatste. Zoals eerder vermeld is het **doortrekken van uniformiteit naar klinische interpretatie de volgende stap** waar de werkgroep nu nog mee aan de slag is, en waar er dus nog adviezen over zullen volgen. Voor het uitwerken van dergelijke klinische richtlijnen zullen eerst per discipline (bijv. logopedie, kinesitherapie, psychologie) de verschillende specifieke domeinen (bijv. reken en taal voor logopedie) in kaart gebracht worden. Vervolgens zal er per domein met experten groepen gekeken worden naar de belangen en bedenkingen voor de uitwerking van klinische richtlijnen zodat ook daar hopelijk geëvolueerd zal worden naar meer uniformiteit afgestemd op deze specifieke domeinen.

### Referentielijst

- Bouma, A., Mulder, J., Lindeboom, J. & Schmand, B. (2012). Handboek neuropsychologische diagnostiek (2e dr.). Amsterdam: Pearson Assessment and Information B.V.
- Thomas J. Guilmette, Jerry J. Sweet, Nancy Hebben, Deborah Koltai, E. Mark Mahone, Brenda J. Spiegler, Kirk Stucky, Michael Westerveld & Conference Participants (2020) American Academy of Clinical Neuropsychology consensus conference statement on uniform labeling of performance test scores, *The Clinical Neuropsychologist*, 34, 437-453, DOI: 10.1080/13854046.2020.1722244
- Hendriks, M.P.H., Mol, B.A.W., & Kessels, R.P.C. (2020). Uniformiteit in de kwalitatieve beschrijving van scores op prestatietaken. *Tijdschrift voor Neuropsychologie*, 15, 166-176.
- Resing, W.C.M. & Blok, J.B. (2002). De classificatie van intelligentiescores: Voorstel voor een eenduidig systeem. *De Psycholoog*, 37, 244-249.
- Schmand, B. & Bouma, A. (2020). Reactie op het voorstel van Hendriks e.a. over Uniformiteit in de kwalitatieve beschrijving van scores op prestatietaken. *Tijdschrift voor Neuropsychologie*, 15, 177-181.
- Van Zandvoort, M. (2020). Commentaar op de publicatie van Hendriks e.a. over Uniformiteit in de kwalitatieve beschrijving van scores op prestatietaken. *Tijdschrift voor Neuropsychologie*, 15, 183-185.