



Pearson

Pearson Benelux –R&D

Joëlle Dek – Sr. Product Developer  
Koen van Loenen – Manager R&D

November 2018

Paper Aanvullende Technische informatie

## Inleiding

In april 2018 is de WISC-V-NL ter beoordeling voorgelegd aan de COTAN (Commissie Testaangelegenheden Nederland). In het COTAN Beoordelingssysteem voor de kwaliteit van tests (Evers, Lucassen, Meijer & Sijtsma, 2010) worden aanbevelingen gedaan voor het verstrekken van psychometrische gegevens ter onderbouwing van de psychometrische kwaliteiten van een test. Gezien de grote hoeveelheid analyses en resultaten werd bij de samenstelling van de Technische handleiding van de WISC-V-NL omwille van de duidelijkheid en overzichtelijkheid de keuze gemaakt om niet alle tabellen en resultaten in de handleiding te publiceren. Ten behoeve van de COTAN-beoordeling is met betrekking tot een drietal psychometrische onderwerpen extra informatie aangeleverd, die niet in de handleiding is opgenomen. In deze paper worden deze gegevens ook voor WISC-V-NL gebruikers beschikbaar gemaakt.

In paragraaf 1 worden de item-restcorrelaties beschreven en weergegeven, in paragraaf 2 de test-hertestresultaten van twee leeftijdsgroepen en in paragraaf 3 informatie over verschillen tussen groepen. Per onderwerp worden aanvullende resultaten gepresenteerd en zal ook kort worden uiteengezet waarom er voor gekozen is om deze gegevens niet in de technische handleiding op te nemen.

### 1. Item-restcorrelaties

Vanuit testontwikkelingsperspectief ligt de nadruk op de kwaliteit van de subtests en indexen, omdat deze gebruikt worden in de beoordeling van de respondent. De subtestbetrouwbaarheid geeft een goede indicatie van de kwaliteit van de onderliggende items. Bovendien kan een item minder goed functioneren, maar de subtest als geheel een goede representatie geven van het te meten construct. Om deze reden is in de Technische handleiding alleen informatie over de subtestkwaliteit en niet over itemkwaliteit weergegeven. In het COTAN Beoordelingssysteem voor de kwaliteit van tests (Evers, Lucassen, Meijer & Sijtsma, 2010) wordt ter onderbouwing van de Betrouwbaarheid (criterium 5.2e) en Validiteit (criterium 6.2) wel gevraagd om gegevens over de psychometrische kwaliteit van items aan te leveren. Ten behoeve van de COTAN-beoordeling en vergelijkbaarheid met andere instrumenten is besloten de gegevens wel in dit aanvullende technische paper op te nemen.

In tabel 1 worden voor elke subtest, processcore en index per leeftijdsgroep de range, mediaan en gemiddelde Item-restcorrelaties weergegeven behaald door de Nederlandse steekproef. De item-restcorrelatie geeft aan in welke mate een item samenhangt met de overige items, oftewel in hoeverre een item hetzelfde meet als de overige items. Uit tabel 1 is af te lezen dat de range van item-restcorrelaties *Gemiddeld over alle leeftijdsgroepen* ligt tussen de 0.03 en 0.83. De mediaan ligt in alle afzonderlijke leeftijdsgroepen op alle subtests tussen de .25 en .64 en het gemiddelde tussen de .24 en .57. Incidenteel komt een negatieve Rir-waarde voor, dit gebeurt alleen bij (een van de) eerste items van een leeftijdsgroep. Deze negatieve waarden zijn niet consistent over leeftijden, zoals blijkt uit het ontbreken van negatieve waarden in de Rir-range van de overall resultaten in de kolom *Gemiddeld over alle leeftijdsgroepen*.

Tabel 1. Item-restcorrelaties per leeftijdsgroep, Nederlandse steekproef

	Gemiddeld over alle leeftijdsgroepen			6:0-6:11 jaar			7:0-7:11 jaar			8:0-8:11 jaar		
	Rir-range	Rir-mediaan	Rir-gemiddelde	Rir-range	Rir-mediaan	Rir-gemiddelde	Rir-range	Rir-mediaan	Rir-gemiddelde	Rir-range	Rir-mediaan	Rir-gemiddelde
BP	0.16 - 0.72	.43	.42	-0.03 - 0.59	.36	.35	0.00 - 0.60	.43	.40	0.00 - 0.62	.39	.34
OV	0.32 - 0.70	.56	.53	0.10 - 0.62	.38	.38	0.23 - 0.60	.43	.43	0.00 - 0.68	.35	.36
MR	0.14 - 0.57	.43	.40	0.06 - 0.58	.38	.39	0.15 - 0.63	.43	.42	0.07 - 0.52	.29	.28
WS	0.06 - 0.71	.48	.44	0.07 - 0.33	.26	.24	0.02 - 0.44	.26	.26	0.00 - 0.56	.30	.30
GW	0.06 - 0.82	.39	.43	0.07 - 0.70	.39	.39	0.11 - 0.74	.43	.45	-0.13 - 0.79	.28	.37
FS	0.10 - 0.65	.43	.42	0.10 - 0.60	.39	.35	0.13 - 0.66	.39	.40	0.06 - 0.60	.41	.36
PR	0.07 - 0.63	.47	.43	0.04 - 0.60	.34	.34	0.08 - 0.67	.40	.38	0.05 - 0.55	.36	.33
CLN	0.22 - 0.66	.47	.49	-	-	-	-	-	-	0.09 - 0.85	.61	.57
BG	0.36 - 0.71	.53	.54	0.14 - 0.42	.29	.29	0.14 - 0.55	.26	.29	0.04 - 0.50	.26	.27
RE	0.16 - 0.74	.52	.50	-0.05 - 0.59	.39	.39	0.09 - 0.63	.51	.46	0.16 - 0.61	.39	.41
BPd	0.08 - 0.83	.40	.44	-0.21 - 0.72	.41	.42	0.16 - 0.80	.48	.49	-0.20 - 0.72	.43	.42
CRV	0.03 - 0.60	.33	.31	0.16 - 0.44	.35	.33	0.20 - 0.53	.28	.32	0.13 - 0.49	.38	.35
CRA	0.10 - 0.61	.35	.36	0.12 - 0.59	.37	.38	0.16 - 0.57	.52	.45	-0.11 - 0.57	.43	.32
CRS	0.19 - 0.69	.50	.45	0.28 - 0.53	.40	.39	0.04 - 0.57	.37	.37	-0.11 - 0.65	.30	.34

Tabel 1. Item-restcorrelaties per leeftijdsgroep, Nederlandse steekproef (vervolg)

	9:0-9:11 jaar			10:0-10:11 jaar			11:0-11:11 jaar			12:0-12:11 jaar		
	Rir-range	Rir-mediaan	Rir-gemiddelde	Rir-range	Rir-mediaan	Rir-gemiddelde	Rir-range	Rir-mediaan	Rir-gemiddelde	Rir-range	Rir-mediaan	Rir-gemiddelde
BP	0.26 - 0.64	.43	.43	0.22 - 0.61	.48	.46	0.12 - 0.65	.40	.38	0.00 - 0.66	.40	.41
OV	0.01 - 0.59	.38	.37	-0.09 - 0.56	.37	.32	0.14 - 0.57	.40	.39	0.25 - 0.61	.41	.42
MR	-0.02 - 0.63	.39	.37	0.00 - 0.57	.37	.35	-0.21 - 0.53	.31	.28	0.04 - 0.60	.38	.38
WS	0.10 - 0.57	.26	.28	0.03 - 0.62	.37	.33	0.05 - 0.52	.33	.34	0.18 - 0.57	.36	.37
GW	0.20 - 0.79	.37	.45	0.03 - 0.76	.36	.40	-0.06 - 0.71	.44	.42	-0.02 - 0.78	.42	.45
FS	0.07 - 0.67	.49	.42	0.07 - 0.62	.38	.36	0.07 - 0.59	.36	.32	0.09 - 0.64	.35	.36
PR	0.17 - 0.59	.45	.43	0.15 - 0.69	.44	.42	0.15 - 0.69	.47	.42	0.08 - 0.62	.37	.34
CLN	0.12 - 0.77	.60	.51	0.02 - 0.54	.36	.31	0.28 - 0.66	.47	.49	0.14 - 0.73	.60	.53
BG	0.06 - 0.42	.26	.27	0.19 - 0.47	.32	.32	0.02 - 0.60	.32	.32	0.16 - 0.65	.36	.36
RE	0.20 - 0.66	.41	.42	0.20 - 0.59	.43	.40	0.15 - 0.64	.39	.38	0.07 - 0.63	.39	.41
BPd	0.33 - 0.81	.56	.55	0.27 - 0.80	.46	.51	0.10 - 0.75	.44	.46	-0.07 - 0.76	.46	.46
CRV	0.21 - 0.64	.46	.44	0.10 - 0.52	.33	.33	0.14 - 0.60	.31	.32	0.14 - 0.62	.40	.40
CRA	0.15 - 0.62	.40	.40	-0.16 - 0.58	.44	.37	0.27 - 0.55	.42	.40	-0.11 - 0.65	.43	.38
CRS	0.23 - 0.65	.40	.43	0.13 - 0.65	.38	.40	0.15 - 0.58	.43	.40	0.18 - 0.63	.51	.46

Tabel 1. Item-restcorrelaties per leeftijdsgroep, Nederlandse steekproef (vervolg)

	13:0-13:11 jaar			14:0-14:11 jaar			15:0-15:11 jaar			16:0-16:11 jaar		
	Rir-range	Rir-mediaan	Rir-gemiddelde	Rir-range	Rir-mediaan	Rir-gemiddelde	Rir-range	Rir-mediaan	Rir-gemiddelde	Rir-range	Rir-mediaan	Rir-gemiddelde
BP	0.00 - 0.74	.26	.34	0.03 - 0.76	.41	.39	0.21 - 0.73	.52	.48	-0.01 - 0.65	.32	.34
OV	0.25 - 0.56	.42	.41	0.14 - 0.60	.33	.37	0.15 - 0.56	.39	.37	0.15 - 0.61	.49	.44
MR	-0.06 - 0.58	.32	.31	-0.01 - 0.53	.37	.32	0.00 - 0.48	.35	.33	0.03 - 0.48	.25	.26
WS	0.12 - 0.61	.38	.38	0.11 - 0.55	.40	.39	0.15 - 0.62	.33	.38	0.06 - 0.63	.47	.45
GW	0.03 - 0.71	.39	.42	-0.06 - 0.75	.49	.42	0.19 - 0.79	.64	.56	-0.09 - 0.75	.43	.41
FS	0.06 - 0.60	.38	.37	-0.01 - 0.64	.39	.38	0.17 - 0.75	.48	.50	-0.1 - 0.61	.31	.31
PR	-0.11 - 0.61	.38	.34	0.12 - 0.64	.39	.37	0.08 - 0.60	.29	.32	0.08 - 0.59	.40	.36
CLN	0.19 - 0.58	.46	.43	0.07 - 0.50	.39	.35	-0.07 - 0.61	.47	.40	0.22 - 0.54	.42	.41
BG	0.06 - 0.59	.41	.39	0.17 - 0.58	.35	.37	0.16 - 0.67	.42	.42	0.08 - 0.67	.44	.41
RE	0.06 - 0.66	.45	.43	0.22 - 0.67	.47	.46	0.12 - 0.66	.45	.44	0.02 - 0.58	.35	.34
BPd	0.00 - 0.86	.61	.51	0.01 - 0.80	.45	.43	0.38 - 0.70	.58	.56	0.04 - 0.64	.46	.36
CRV	0.09 - 0.58	.28	.29	0.10 - 0.48	.34	.32	0.24 - 0.53	.41	.40	0.18 - 0.67	.47	.45
CRA	0.03 - 0.53	.40	.35	-0.01 - 0.62	.46	.40	-0.07 - 0.56	.33	.31	0.09 - 0.57	.40	.38
CRS	0.12 - 0.61	.46	.43	-0.03 - 0.62	.48	.41	0.26 - 0.65	.50	.47	0.14 - 0.58	.38	.35

## 2. Test-hertestresultaten van twee leeftijdsgroepen

In het COTAN Beoordelingssysteem voor de kwaliteit van tests (Evers, Lucassen, Meijer & Sijtsma, 2010) wordt aanbevolen om betrouwbaarheidscoëfficiënten per normgroep te berekenen (zie criterium 5.3b). Dit is echter voor onderzoeken buiten het normeringsonderzoek om, zoals test-hertest en interbeoordelaaronderzoek, praktisch onmogelijk omdat de omvang van de groepen dan te groot wordt, en een overbelasting van de proefpersonen ontstaat. In dit geval zouden voor het test-hertestonderzoek elf normgroepen x 50 proefpersonen (een minimale N van 50 is nodig per groep; Bartram & Lindley, 1999)= 550 proefpersonen twee maal getest moeten worden. Ook theoretisch gezien ontbreekt de noodzaak: er wordt in de literatuur geen trend beschreven over fluctuaties tussen leeftijdsgroepen, en tenzij er een consistente trend te zien is, zijn verschillen tussen subgroepen meestal een effect van de groeps grootte (steekproeffluctuaties). Om dezelfde reden zijn er geen hertestresultaten per normgroep gerapporteerd bij de WMS-IV-NL (COTAN, 2018a), RAKIT-2 (COTAN, 2018b) en WAIS-IV-NL (COTAN, 2018c) en kregen deze tests ondanks dit gegeven toch een beoordeling 'Goed'.

Tabel 5.5 uit de Technische handleiding waarin de test-hertest betrouwbaarheid gepresenteerd wordt, is ook in dit paper opgenomen als tabel 2. In tabel 3 en 4 zijn deze resultaten uitgesplitst naar twee aparte leeftijdsgroepen; 6-9 jaar en 10-16 jaar. Uit onderstaande resultaten in tabel 3 en 4 blijkt dat de varianties van de schaal in de twee aparte leeftijdsgroepen niet hoger liggen dan de varianties in de totale groep (tabel 2). De resultaten zijn zeer consistent met de resultaten over de totale groep; voor de kinderen van 6-9 jaar is de gecorrigeerde correlatie van het TIQ op de eerste en tweede meting .93 en de gecorrigeerde correlaties van de overige indexen tussen de .76 en .93. Op subtestniveau liggen de correlaties tussen de .64 en .87 (met uitzondering van Overeenkomsten en Gewichten, die respectievelijk .51 en .53 zijn). In de groep van 10 tot 16 jaar is de gecorrigeerde correlatie van het TIQ .96 en liggen de correlaties van de indexen tussen de .71 en .96. Op subtestniveau liggen de gecorrigeerde correlaties tussen de .61 en .83 (met uitzondering van Figuur Samenstellen en Matrix Redeneren, welke respectievelijk .54 en .47 zijn). Al met al zijn de resultaten over de gehele groep dus niet geflatteerd ten opzichte van deze twee aparte groepen. Omdat de resultaten over een grotere groep minder onderhevig zijn aan steekproeffluctuaties en er aan de resultaten van aparte groepen voor gebruikers geen conclusies verbonden kunnen worden en vanwege de overzichtelijkheid, is er voor gekozen in de handleiding alleen de resultaten van de totale test-hertest groep te presenteren.

Tabel 2. (tabel 5.5 uit de technische handleiding). Test-hertest betrouwbaarheid van de subtest-, proces- en indexscores (Nederland en Vlaanderen gecombineerd)

**Tabel 5.5** Test- hertest betrouwbaarheid van de subtest-, proces- en indexscores (Nederland en Vlaanderen gecombineerd)

Alle leeftijden Subtest-/Proces-/ Indexscore	Eerste afname		Tweede afname		$r_{12}^2$	Gecorrigeerde $r^2$	Standaard- verschil <sup>c</sup>
	Gem.	SD	Gem.	SD			
OV	10.2	2.3	11.2	2.4	.40	.63	.43
WS	10.8	3.1	11.1	2.6	.68	.68	.10
BG	10.8	2.9	11.1	3.0	.78	.79	.10
BP	10.1	2.3	11.1	2.3	.51	.70	.43
FS	10.6	2.5	11.3	2.7	.64	.75	.27
MR	10.4	2.7	11.4	2.6	.53	.63	.38
GW	10.6	3.0	11.4	3.2	.65	.65	.26
RE	11.1	2.7	10.9	2.8	.75	.82	-.07
CR	11.6	3.0	11.5	2.7	.79	.80	-.04
PR	10.8	2.9	10.9	2.5	.60	.62	.04
CLN	10.9	2.7	11.5	2.9	.68	.76	.21
SSC	10.6	2.9	12.5	3.2	.71	.73	.62
SZ	10.2	3.1	12.0	3.5	.73	.71	.54
FZ	10.0	2.8	11.5	2.8	.77	.79	.54
BPz	10.3	2.0	11.2	2.3	.50	.78	.42
BPd	9.9	2.4	11.0	2.4	.56	.73	.46
CRv	10.7	2.6	11.2	2.6	.56	.71	.19
CRA	11.1	2.6	11.4	2.6	.68	.76	.12
CRs	11.4	3.1	11.1	2.8	.74	.71	-.10
FZw	9.8	3.0	11.7	2.9	.70	.71	.64
FZg	9.9	2.6	11.2	2.8	.77	.83	.48
VBI	102.7	12.1	105.9	12.1	.62	.76	.26
VRI	101.8	11.6	106.3	11.6	.71	.83	.39
FRI	103.4	12.8	108.5	13.7	.68	.77	.38
Wgl	106.6	14.3	107.0	12.5	.84	.85	.03
Vsl	102.4	15.6	112.9	18.2	.78	.76	.62
TIQ	105.1	11.1	110.9	12.6	.91	.95	.49
KRI	104.9	14.6	106.4	15.1	.80	.81	.10
AWI	106.2	14.2	108.1	14.8	.81	.86	.13
NVI	104.9	12.0	111.0	13.3	.87	.92	.48
AVI	103.1	10.5	107.9	12.4	.83	.92	.42
CCI	105.6	15.9	112.0	16.3	.87	.85	.40

<sup>a</sup> Gemiddelde correlaties zijn berekend met behulp van een Fisher's z-transformatie.

<sup>b</sup> Correlaties werden gecorrigeerd voor de variabiliteit van de normeringssteekproef (Allen & Yen, 2002; Magnusson, 1967).

<sup>c</sup> Het standaardverschil is het verschil tussen de gemiddelden van de twee afnames, gedeeld door de wortel van de gecombineerde variantie, berekend met behulp van Cohen's (1996) Formule 10.4.

Tabel 3. Test-hertest betrouwbaarheid van de subtest-, proces- en indexscores in de leeftijdsgroep 6-9 jaar

Alle leeftijden Subtest- / Proces- / Indexscore	Eerste afname		Tweede afname		n	$r_{12}$	Gecorrigeerde $r$	Standaardverschil
	Gem.	SD	Gem.	SD				
OV	10.1	2.5	11.6	2.2	37	0.30	0.51	0.64
WS	10.4	2.6	10.6	2.3	37	0.65	0.74	0.08
BG	10.8	3.0	11.4	2.9	35	0.81	0.81	0.20
BP	9.8	2.5	10.8	2.0	37	0.52	0.67	0.44
FS	10.6	2.4	11.7	2.9	38	0.80	0.87	0.41
MR	10.8	2.5	11.6	2.4	38	0.64	0.75	0.33
GW	10.5	3.1	10.8	3.1	38	0.56	0.53	0.10
RE	11.5	2.3	11.4	2.9	38	0.72	0.84	-0.04
CR	11.8	2.9	11.7	2.3	38	0.74	0.76	-0.04
PR	10.7	3.1	10.7	3.0	38	0.68	0.66	0.00
CLN	10.7	2.3	11.3	2.7	19	0.67	0.81	0.24
SSC	10.3	3.2	12.2	3.5	33	0.72	0.68	0.57
SZ	10.1	2.9	12.0	3.2	37	0.62	0.64	0.62
FZ	9.9	3.1	11.1	2.8	35	0.76	0.74	0.41
BPz	10.2	1.9	10.9	2.1	35	0.59	0.84	0.35
BPd	9.5	2.5	10.8	2.3	31	0.45	0.62	0.54
CRv	11.2	2.1	11.8	2.3	37	0.45	0.73	0.27
CRa	11.0	2.4	11.0	2.6	38	0.59	0.74	0.00
CRs	11.7	3.4	11.2	2.5	38	0.72	0.64	-0.17
FZw	9.5	2.8	11.1	2.9	32	0.67	0.71	0.56
FZg	9.7	2.7	11.1	2.8	36	0.84	0.87	0.51
VBI	101.4	10.4	105.5	10.4	37	0.60	0.81	0.39
VRI	101.1	12.0	107.2	12.0	37	0.75	0.84	0.51
FRI	104.2	12.7	107.6	13.4	37	0.73	0.81	0.26
WgI	107.0	14.9	106.8	12.7	38	0.86	0.86	-0.01
VsI	101.6	15.9	111.7	18.3	34	0.79	0.76	0.59
TIQ	104.3	12.1	109.7	12.2	32	0.90	0.93	0.44
KRI	105.8	13.7	106.5	14.7	38	0.74	0.78	0.05
AWI	105.8	11.1	108.2	12.2	19	0.78	0.88	0.21
NVI	104.3	13.6	110.3	15.1	33	0.92	0.93	0.42
AVI	102.1	10.7	107.3	10.9	36	0.84	0.92	0.48
CCI	105.4	16.3	111.3	16.2	34	0.89	0.87	0.36



Tabel 4. Test-hertest betrouwbaarheid van de subtest-, proces- en indexscores in de leeftijdsgroep 10-16 jaar

Alle leeftijden Subtest- / Proces- / Indexscore	Eerste afname		Tweede afname		n	$r_{12}$	Gecorrigeerde $r$	Standaardverschil
	Gem.	SD	Gem.	SD				
OV	10.3	2.2	10.9	2.6	42	0.49	0.73	0.25
WS	11.2	3.4	11.5	2.9	38	0.70	0.61	0.09
BG	10.7	2.9	10.9	3.1	39	0.75	0.77	0.07
BP	10.3	2.2	11.2	2.5	41	0.50	0.73	0.38
FS	10.7	2.6	10.9	2.4	43	0.39	0.54	0.08
MR	10.0	2.8	11.2	2.9	39	0.39	0.47	0.42
GW	10.7	2.9	11.9	3.3	43	0.73	0.75	0.39
RE	10.8	2.9	10.3	2.7	43	0.78	0.79	-0.18
CR	11.3	3.0	11.4	3.0	43	0.83	0.83	0.03
PR	10.9	2.8	11.0	2.1	43	0.51	0.57	0.04
CLN	11.0	3.0	11.5	3.0	43	0.69	0.69	0.17
SSC	10.9	2.6	12.8	2.9	40	0.69	0.77	0.69
SZ	10.2	3.4	12.0	3.9	41	0.81	0.76	0.49
FZ	10.1	2.6	11.8	2.7	41	0.78	0.83	0.64
BPz	10.5	2.1	11.4	2.4	39	0.40	0.71	0.40
BPd	10.2	2.2	11.2	2.5	29	0.65	0.81	0.42
CRv	10.2	2.9	10.6	2.8	43	0.66	0.68	0.14
CRa	11.3	2.8	11.7	2.5	43	0.75	0.78	0.15
CRs	11.2	2.9	11.0	3.1	43	0.75	0.77	-0.07
FZw	10.0	3.1	12.1	2.8	41	0.72	0.70	0.71
FZg	10.1	2.5	11.3	2.8	42	0.67	0.77	0.45
VBI	103.9	13.5	106.2	13.6	39	0.64	0.71	0.17
VRI	102.5	11.4	105.4	11.3	41	0.67	0.81	0.26
FRI	102.8	13.0	109.3	14.2	40	0.63	0.72	0.48
WgI	106.3	14.0	107.1	12.5	43	0.81	0.83	0.06
VsI	103.1	15.4	114.0	18.2	40	0.77	0.76	0.65
TIQ	105.9	10.1	112.0	13.0	33	0.91	0.96	0.52
KRI	104.1	15.4	106.2	15.6	43	0.85	0.84	0.14
AWI	106.3	15.5	108.1	16.0	43	0.84	0.83	0.11
NVI	105.3	10.4	111.7	11.5	35	0.80	0.90	0.58
AVI	104.1	10.3	108.5	13.9	35	0.82	0.92	0.36
CCI	105.8	15.8	112.6	16.6	40	0.85	0.83	0.42

### 3. Verschillen tussen groepen

In het COTAN Beoordelingssysteem voor de kwaliteit van tests (Evers, Lucassen, Meijer & Sijsma, 2010) wordt ter onderbouwing van de validiteit gevraagd naar mogelijke verschillen tussen subgroepen (criterium 4.6). Deze mogelijke verschillen tussen subgroepen moeten onderbouwd zijn vanuit de literatuur om ondersteuning te kunnen leveren voor de validiteit van de test. De subgroepen die daarom in de handleiding besproken worden zijn klinische groepen met verschillende diagnoses (paragraaf 6.7) en verschillen tussen kinderen van verschillend opleidingsniveau en tussen kinderen met moeders van verschillend opleidingsniveau (paragraaf 6.8). Tussen andere te onderscheiden groepen met verschillende demografische kenmerken worden wel verschillen verwacht, maar niet altijd is de literatuur eenduidig over de richting van verschillen. Onder andere doordat er vaak sprake is van interactie-effecten, zoals dat verschillen tussen bijvoorbeeld mannen-vrouwen op bepaalde leeftijden groter zijn dan op andere leeftijden. Hierdoor zijn dergelijke resultaten moeilijk eenduidig te interpreteren en te gebruiken ter onderbouwing van het instrument. Het is belangrijk op te merken dat eventuele verschillen tussen subgroepen niet leiden tot het ontwikkelen van specifieke normen; de keuze voor normgroepen hangt samen met het doel van de test. Het doel van de WISC-V-NL is om de prestatie van een kind te vergelijken met zijn/ haar leeftijdsgenoten uit de totale populatie. Dit betekent dat er ongeacht eventuele verschillen tussen groepen geen aparte normen gecreëerd zijn, behalve leeftijdsnormen.

De demografische kenmerken waarop wel verschillen worden verwacht en waarover de literatuur ook duidelijk is omtrent de richting en verwachting zijn:

- **Leeftijd:** uit de literatuur is bekend dat intelligentie toeneemt tot een jaar of 25-35 en daarna weer afneemt (o.a. Wechsler, 2012). Hier is wel een verschil waar te nemen tussen fluid en crystallized taken; bij fluid subtests nemen de prestaties geleidelijk af, bij crystallized niet en nemen ze soms zelfs een beetje toe (Luteijn & Barelds, 2004). Bij de GIT-2 lag de piek gemiddeld rond de 40 jaar. Voor de leeftijdsgroep van de WISC-V-NL worden echter consistent toenemende scores verwacht. Omdat de verschillen tussen verschillende leeftijdsgroepen dusdanig groot en consistent oplopend zijn in het leeftijdsbereik van de WISC-V-NL zijn er voor deze variabele wel aparte normtabellen opgesteld.
- **Opleidingsniveau respondent:** de verschillen tussen kinderen van verschillende opleidingsniveaus staan vermeld in de Technische handleiding (tabel 6.13; Wechsler, 2018). Het is belangrijk dat gebruikers zich bewust zijn van de verschillen tussen respondenten van verschillende opleidingsniveau. Deze verschillen worden bovendien verwacht en geven daarom ondersteuning voor de validiteit van de test.
- **Sociaal-Economische Status:** dit is geoperationaliseerd als opleidingsniveau moeder en de resultaten hiervan staan vermeld in de Technische handleiding (tabel 6.14; Wechsler, 2018). Het is belangrijk dat gebruikers zich bewust zijn van de verschillen tussen respondenten met moeders van verschillende opleidingsniveaus. Deze verschillen worden bovendien verwacht en geven daarom ondersteuning voor de validiteit van de test.

Om toch enige informatie te verschaffen over verschillen tussen subgroepen met verschillende demografische kenmerken zijn een aantal extra MANOVA-analyses uitgevoerd. Er is hierbij gekeken naar groepen die worden onderscheiden op basis sekse, leeftijd, wel of geen migratie-achtergrond, opleidingsniveau en Sociaal-Economische achtergrond (SES). Hieronder wordt per achtergrondcriterium eerder onderzoek en bevindingen op de WISC-V-NL gepresenteerd.

- Sekse: uit literatuur en eerder onderzoek (Luteijn & Barelds, 2004; Wechsler, 2012; Wechsler, 2014) is bekend dat scores van mannen en vrouwen regelmatig significant van elkaar verschillen. De richting van dit verschil is echter verschillend, onder meer afhankelijk van cognitief domein, leeftijd en opleidingsniveau. Dit blijkt ook bij de WISC-V-NL het geval. Bij toetsing met behulp van MANOVA's worden significante verschillen gevonden tussen beide seksen op FRI, Wgl, Vsl en TIQ met kleine effectgroottes en een verklaarde variantie  $\eta^2$  tussen 0.01 en 0.06. Op FRI scoren jongens hoger, op de overige indexen waar een significant verschil gevonden wordt, scoren meisjes hoger.
- Migratie-achtergrond: op het gebied van het wel of niet hebben van een migratie-achtergrond worden er geen verschillen tussen subgroepen verwacht. Wel is het mogelijk dat thuistaal, omgeving en cultuur van invloed zijn op de testresultaten van kinderen. Uit MANOVA-analyses op de WISC-V-NL-data waarin de variabelen sekse, land en migratie-achtergrond verwerkt zijn, blijkt dat er op alle indexen significante verschillen bestaan tussen kinderen met en kinderen zonder een migratie-achtergrond. Deze verschillen hebben in alle gevallen een zeer kleine verklaarde variantie van  $\eta^2$  tussen 0.01-0.04. Zie tabel 5.

Index	Vergelijking	Statistic	F-					Multivariate $\eta^2$
			Waarde	waarde	NumDF	DenDF	ProbF	
VBI	sekse	Wilks' Lambda	0.99	3.62	3	1423	0.01	0.01
VBI	land	Wilks' Lambda	1.00	1.26	3	1423	0.29	0.00
VBI	achtergrond	Wilks' Lambda	0.98	5.86	3	1033	0.00	0.02
VRI	sekse	Wilks' Lambda	1.00	1.08	2	1430	0.34	0.00
VRI	land	Wilks' Lambda	1.00	2.66	2	1430	0.07	0.00
VRI	achtergrond	Wilks' Lambda	0.96	20.88	2	1035	0.00	0.04
FRI	sekse	Wilks' Lambda	0.96	20.71	3	1425	0.00	0.04
FRI	land	Wilks' Lambda	1.00	0.76	3	1425	0.51	0.00
FRI	achtergrond	Wilks' Lambda	0.98	8.20	3	1033	0.00	0.02
Wgl	sekse	Wilks' Lambda	0.98	6.07	3	1127	0.00	0.02
Wgl	land	Wilks' Lambda	0.99	5.48	3	1127	0.00	0.01
Wgl	achtergrond	Wilks' Lambda	0.99	3.26	3	815	0.02	0.01
Vsl	sekse	Wilks' Lambda	0.94	29.66	3	1381	0.00	0.06
Vsl	land	Wilks' Lambda	1.00	1.45	3	1381	0.23	0.00
Vsl	achtergrond	Wilks' Lambda	0.96	13.67	3	1025	0.00	0.04
TIQ	sekse	Wilks' Lambda	0.95	13.52	5	1389	0.00	0.05
TIQ	land	Wilks' Lambda	0.97	7.81	5	1389	0.00	0.03
TIQ	achtergrond	Wilks' Lambda	0.96	9.23	5	1028	0.00	0.04

Bovenstaande bevindingen geven aan dat de WISC-V-NL wel differentieert tussen groepen, maar de verschillen zijn niet eenduidig te interpreteren, wat in lijn is met eerder onderzoek.

Naast het feit dat onderzoek naar subgroepen gebruikt wordt ter onderbouwing van de validiteit moeten gebruikers bekend zijn met mogelijke verschillen tussen subgroepen als mogelijke verklaring van resultaten van de respondent. Uiteraard heeft de gebruiker ook een eigen verantwoordelijkheid om bekend te zijn met de literatuur omtrent intelligentie en onderzochte verschillen tussen subgroepen (zie ook de verwijzing naar de eigen verantwoordelijkheid van de gebruiker in de Technische handleiding; Wechsler, 2018). Op basis van bovenstaande groepsverschillen kunnen echter geen individuele conclusies getrokken worden voor de diagnostische praktijk.

## Referenties

- Bartram, D. & Lindley, P.A. (1999). *Psychologisch testen module 3. Betrouwbaarheid en de standaardmeetfout*. Swets & Zeitlinger: Lisse.
- COTAN (2018a). Testbeoordeling WMS-IV-NL. Geraadpleegd op 29-3-2018 van <https://www.cotandocumentatie.nl/beoordelingen/b/14786/wechsler-memory-scale/>
- COTAN (2018b). Testbeoordeling RAKIT-2. Geraadpleegd op 29-3-2018 van <https://www.cotandocumentatie.nl/beoordelingen/b/14723/revisie-amsterdamse-kinder-intelligentietest--2e-editie/>
- COTAN (2018c). Testbeoordeling WAIS-IV-NL. Geraadpleegd op 29-3-2018 van <https://www.cotandocumentatie.nl/beoordelingen/b/14695/wechsler-adult-intelligence-scale-fourth-edition--nederlandstalige-bewerking/>
- Evers, A., Lucassen, W., Meijer, R., & Sijtsma, K. (2010). *COTAN beoordelingssysteem voor de kwaliteit van tests*. NIP: Amsterdam.
- Luteijn, F. & Barelds, D. (2004). *GIT-2 Handleiding*. Amsterdam: Pearson Benelux B.V.
- Wechsler, D. (2012). *Wechsler Adult Intelligence Scale, vierde editie; Nederlandstalige bewerking*. Amsterdam: Pearson Benelux B.V.
- Wechsler, D. (2014). *Wechsler Memory Scale, vierde editie; Nederlandstalige bewerking*. Amsterdam: Pearson Benelux B.V.
- Wechsler, D. (2018). *Wechsler Intelligence Scale for Children, vijfde editie; Nederlandstalige bewerking*. Amsterdam: Pearson Benelux B.V.