CoVaT-CHC Basisversie

Cognitieve VaardigheidsTest volgens het CHC-model

V. Decaluwé

M. Tierens

W. Magez

Onderzoeksrapport

Test-hertestbetrouwbaarheid

COÖRDINATIE PSYCHODIAGNOSTIEK Antwerpen

PDC CENTRUM VOOR PSYCHODIAGNOSTIEK @THOMAS MORE

COVAT – CHC BASISVERSIE ONDERZOEKSRAPPORT: TEST-HERTESTBETROUWBAARHEID

V. Decaluwé, M. Tierens, & W. Magez (2017)

Methodologische introductie

Stabiliteit van de CoVaT-CHC Basisversie

In een eerder studie hebben we de betrouwbaarheid op het niveau van de interne consistentie onderzocht beschreven in het onderzoeksrapport Betrouwbaarheid en betrouwbaarheidsintervallen (Tierens & Magez, 2016a; zie hoofdstuk 5.1 op het CoVaT – CHC platform). In huidige studie toetsen we via de test-hertestmethode een ander belangrijk aspect van de testbetrouwbaarheid, namelijk de teststabiliteit doorheen de tijd. Het onderzoek benadert deze vorm van betrouwbaarheid op het niveau van de subtests, de brede cognitieve vaardigheden (BCV) en het IQ (GCV).

Afgeleide uitslagen

In de meeste intelligentietests wordt de test-hertestbetrouwbaarheid van de subtests, indexen en het totaal IQ (TIQ) berekend aan de hand van de *afgeleide uitslagen*. Deze benadering werd ook gevolgd bij het schatten van de test-hertestbetrouwbaarheid van de CoVaT-CHC (Magez et al., 2015).

Voor zowel het schatten van de betrouwbaarheid van de subtests, de indexen voor de brede cognitieve vaardigheden (BCV indexen) als de globale cognitieve vaardigheid (GCV of TIQ) van de CoVaT – CHC Basisversie werd de *ruwe uitslag* omgezet in een genormaliseerde standaardscore op basis van de algemene normgroep van de CoVaT-CHC Basisversie. Voor een gedetailleerde beschrijving van deze normgroep verwijzen we naar het onderzoeksrapport Normering op het CoVaT – CHC platform (Tierens & Magez, 2016b; zie hoofdstuk 3.1).

Op basis van deze gegevens werden de ruwe uitslagen van de subtests getransformeerd naar een andere gemeenschappelijk schaal, met name een afgeleide uitslag (AU). Bij de subtests werd de test-hertest betrouwbaarheid ($\mathbf{r_{12}}$) nagegaan op de afgeleide uitslagen (AU) met gemiddelde 10 en standaarddeviatie 3 (zoals bij de Wechsler - tests). Deze subtest AU waren voorheen nog niet

beschikbaar. Een beschrijving van de afgeleide uitslagen per subtest volgt nog in een latere publicatie en zullen worden besproken in een apart hoofdstuk bij "Normering" (hoofdstuk 3.6).

De BCV indexen en IQ/GCV hebben een gemiddelde van 100 met standaarddeviatie van 15. Al deze scores (subtest AU en indexen) zijn genormaliseerde standaardscores bepaald op de algemene CoVaT-CHC Basisversie normgroep. Voor een beschrijving van de normgroep en toelichting bij de gehanteerde statistische procedures bij de normscores verwijzen we naar de hoofdstukken 3.1 en 3.2 (Tierens & Magez, 2016b; Magez, Van Parijs, & Tierens, 2016)

Methode

Steekproef

De test-hertest betrouwbaarheid van de CoVaT-CHC Basisversie werd nagegaan in een afzonderlijke steekproef van 106 Vlaamse kinderen die twee keer werd getest. De steekproef bevatte kinderen uit drie verschillende leerjaren: drie klassen uit het 4de leerjaar, twee klassen uit het 6de leerjaar en drie klassen uit het 2de middelbaar (leeftijdsgroepen: 9j6 maanden - 11 jaar, 10 – 12 jaar en 12 – 13j11 maanden). De kinderen waren afkomstig uit twee verschillende Antwerpse scholen.

Voor het bepalen van de test-hertestbetrouwbaarheid werd de test tweemaal door dezelfde proefleider afgenomen (met uitzondering van de afnames in het 2de middelbaar). Het gaat om een groepsgewijze afname die binnen de schoolcontext plaatsvond. Bij elke klas werd de CoVaT-CHC tweemaal afgenomen met een tussentijd van ten minste 1 maand. Het interval tussen de afnames varieerde van 35 dagen tot 46 dagen, met een gemiddelde tijd tussen beide afnames van 40 dagen.

De kenmerken van deze steekproef staan in Tabel 1, zowel voor de volledige steekproef als voor de drie afzonderlijke leerjaren. De gehele steekproef bestond ongeveer voor de helft uit jongens (52%) en meisjes (48%). De gemiddelde leeftijd van de volledige groep op het tijdstip van de eerste afname was 11.79 jaar (*SD*= 1.24). Bij de verschillende leeftijdsgroepen apart, is de gemiddelde leeftijd voor de groep uit het 4de leerjaar 10.07 jaar (*SD* = 0.54), voor het 6de leerjaar 11.00 jaar (*SD* = 0.60) en voor het 2de middelbaar 12.98 jaar (*SD* = 0.14).

Tabel 1: Kenmerken van de test-herteststeekproef

Leerjaar	Aantal	Leeftijd in jaren		Geslacht (%)		Thuistaal (%)	
	N	***M***	***(SD)***	**Jongens**	**Meisjes**	**Nederlands**	**Niet-Nederlands**
4de leerjaar	16	10.07	(0.54)	56%	44%	88%	12%
6de leerjaar	40	11.00	(0.60)	45%	55%	83%	17%
2de middelbaar	50	12.98	(0.14)	56%	44%	52%	48%
Totaal	**106**	**11.79**	**(1.24)**	**52%**	**48%**	**69%**	**31%**

Alle kinderen zijn Nederlandstalig. Een derde van deze steekproef spreekt echter thuis een andere taal (bijv. Frans, Arabisch,…). Vooral bij de leerlingen van het 2de middelbaar merken we op dat bijna de helft van de leerlingen thuis een andere taal spreekt dan het Nederlands.

Meer gedetailleerde beschrijving per leeftijdsgroep

De groep van het 4de leerjaar is afkomstig uit drie klassen. In totaal namen 46 kinderen deel aan het onderzoek. Echter, een groot deel van de kinderen was op het moment van de eerste afname jonger dan 9 jaar en 6 maanden (*N* = 30). Aangezien de CoVaT-CHC Basisversie bedoeld is voor kinderen vanaf 9 jaar en 6 maanden, werden de resultaten van deze kinderen niet opgenomen in huidige studie. Hierdoor komen we op een totaal van 16 kinderen die voldeden aan het leeftijdscriterium. Deze groep bestaat uit 56% jongens (*N* = 9) en 44% meisjes (*N* = 7), met een gemiddelde leeftijd van 10 jaar (*SD* = 0.54, spreiding = 9.5 -11.37 jaar).

De groep van het 6de leerjaar is afkomstig uit twee klassen. In totaal namen 43 kinderen deel aan het onderzoek. Echter, drie respondenten namen slechts deel aan één testafname waardoor hun resultaten niet werden opgenomen in deze studie. De uiteindelijk groep bestaat uit 40 kinderen waarvan 45% jongens (*N* = 18) en 55% meisjes (*N* = 22), met een gemiddelde leeftijd van 11 jaar (*SD* = 0.60, spreiding = 10 -12 jaar).

De groep van het 2de middelbaar (2de leerjaar, 1ste graad) is afkomstig uit drie klassen van de A-stroom. In totaal namen 62 jongeren deel aan het onderzoek. Echter, 10 scholieren waren op het moment van de eerste afname ouder dan 13 jaar en 11 maanden. Aangezien de CoVaT-CHC Basisversie bedoeld is voor kinderen en jongeren tot 13 jaar en 11 maanden, werden de resultaten van deze jongeren niet opgenomen in huidige studie. Hierdoor komen we op een totaal van 50 jongeren die voldeden aan het leeftijdscriterium. Deze groep bestaat uit 56% jongens (*N* = 28) en 44% meisjes (*N* = 22), met een gemiddelde leeftijd van 13 jaar (*SD* = 0.14, spreiding = 12 - 13 jaar).

Statistische procedure

Voor het schatten van de test-hertestbetrouwbaarheid werden Pearson correlaties berekend tussen de afgeleide uitslagen (subtests) en indexen (BCV) en IQ/GCV van tijdstip 1 en 2. Daarnaast wordt de grootte van de verschilscores besproken.

Correlaties

In eerste instantie werd de teststabiliteit geschat aan de hand van de Pearson correlaties tussen de afgeleide uitslagen/indexen bij afname 1 en de afgeleide uitslagen/indexen bij afname 2 ($\mathbf{r_{12}}$). Voor de volledige groep (N = 106) werd voor elke subtest en Brede Cognitieve vaardigheid (BCV- module) van de CoVaT-CHC Basisversie een Pearson correlatie berekend tussen de eerste en tweede afname (zie Tabel 2 tot en met Tabel 4). Tegelijk werd ook het significantieniveau bepaald van de berekende correlaties. In deze tabellen zijn eveneens de gemiddelde scores en de standaarddeviaties voor de beide afnames vermeld. Voor de berekening van de scores (afgeleide uitslagen resp. indexen) is uitgegaan van de leeftijd van het kind op het moment van de testafnames. De test-hertest correlaties ($\mathbf{r_{12}}$) zijn berekend op de cijfers zoals ze werkelijk vastgesteld zijn en we extrapoleerden zo weinig mogelijk. Een correctie voor de variantie (of spreiding) binnen de steekproef t.o.v. de algemene populatie hebben we niet toegepast.

Een correctie voor attenuatie is echter wel toegepast (Guilford, 1956). Bij deze correctie wordt rekening gehouden met de betrouwbaarheid of onbetrouwbaarheid van de test. Deze methode is bestemd om zicht te krijgen op de 'ware intercorrelatie' als we twee tests met elkaar wensen te correleren. Aangezien de intercorrelatie tussen twee tests in deze situatie kan geïnterpreteerd worden als een vorm van test-hertestcorrrelatie, kan de gecorrigeerde test-hertestbetrouwbaarheid via deze correctie geschat worden.

De formule is: $r_{xw} = \frac{r_{xy}}{\sqrt{r_{xx'} r_{yy'}}}$

waarbij r_{xw}= gecorrigeerde intercorrelatie,

r_{xy}= intercorrelatie tussen test 1 en test 2,

$r_{xx'}$= betrouwbaarbaarheid van test 1 en $r_{yy'}$ = betrouwbaarheid van test 2

In huidige studie, waar op twee tijdstippen dezelfde test werd afgenomen, wordt aangenomen dat de interne consistentie van de test (zie vorig hoofdstuk 5.1; Tierens & Magez, 2016a) voor de beide testmomenten dezelfde is (i.e. $r_{xx'} = r_{yy'} = \lambda_2$, dus $\sqrt{r_{xx'} r_{yy'}} = \lambda_2$).

Vervolgens kan de formule dan eenvoudig herleid worden tot:

$$r_{12c} = \frac{r_{12}}{\lambda_2}$$

waarbij r_{12C} = gecorrigeerde test-hertest betrouwbaarheid,

r_{12} = vastgestelde correlatie tussen afname 1 en afname 2,

λ_2 = interne consistentie van de test

Het hanteren van de gecorrigeerde correlatie sluit onmiddellijk aan bij de stelregel dat een score steeds gehanteerd wordt mét zijn betrouwbaarheidsinterval. Als bij de tweede afname de score met zijn betrouwbaarheidsinterval binnen het betrouwbaarheidsinterval van de score bij de eerste afname valt - ook al zijn ze verschillend (zie ongecorrigeerde correlatie) - zijn deze scores *niet echt* verschillend (getoetst via de correctie voor attenuatie op de oorspronkelijke correlatie).

Om de sterkte van het verband te interpreteren worden volgende interpretatieniveaus gehanteerd: “klein ” (.10 < r < .30), “matig” (.30 < r < .50) en “zeer sterk” (r > .50) (Vanhoomissen & Valkeneers, 2014).

Verschilscores en effect size

Bij afname 1 en bij afname 2 werden telkens de gemiddelden (*M)* en standaarddeviaties (*SD*) berekend op de scores (AU/Indexen) en daaruit hun verschilscore op de twee afnamemomenten. Bij de subtests gaat het dus om AU met gemiddelde 10 en *SD* van 3. Bij de BCV indexen en IQ/GCV ligt het gemiddelde op 100 met *SD*=15. De verschilscores werden berekend via gepaarde t-toetsen (t-toets voor afhankelijke steekproeven). Op basis van deze verschilscores werd eveneens de effectgrootte bepaald.

De formule is:

$$Cohen's\ d = \frac{(M_1 - M_{2)}}{SD_{diff}}$$

waarbij

M_1 = gemiddelde op meetmoment 1

M_2 = gemiddelde op meetmoment 2

SD_{diff}= standaarddeviatie van de verschilscore

Om de effect grootte te interpreteren worden volgende interpretatieniveaus gehanteerd: “klein effect” (.10 - .50), “matig effect” (.50 - .80) en “zeer sterk effect” (> .80) (Cohen, 1988; Johnson & Eagly, 2000)

Resultaten

Correlaties

Subtests

In Tabel 2 wordt een overzicht gegeven van de betrouwbaarheidscoëfficiënten op basis van de afgeleide uitslagen op de verschillende subtests.

Voor de volledige steekproef variëren de gecorrigeerde test-hertestcorrelaties van .64 tot .95. Alle gecorrigeerde correlaties zijn zeer sterk. De hoogste gecorrigeerde correlaties worden gevonden voor de subtest Tegenstellingen .95 en Figuurreeksen .90. De laagste gecorrigeerde correlatie vinden we voor de subtest Puntreeksen .64

De lage test-hertestcorrelatie van Puntreeksen heeft mogelijk te maken met de aard van de subtest zelf. Puntreeksen blijkt een zeer gemakkelijke subtest te zijn (zie hoofdstuk 6.2; Magez & Tierens, 2016). Bij leerlingen in de voorliggende steekproef uit het zesde leerjaar was de gemiddelde ruwe score bij de eerste afname al 24.4 op 30. Er is weinig spreiding naar *boven* toe (plafondeffect) en de

moeilijkste items (29-30) blijken minder betrouwbaar, wat samen de intercorrelatie tussen beide afname drukt.

*Tabel 2: Gemiddelden en standaardafwijkingen voor de test-hertest scores, correlaties (**r_{12}** = ongecorrigeerd en r_{12C} = gecorrigeerd) tussen beide scores (N = 106)*

Subtests	**Eerste afname**		**Tweede afname**			
	M	***SD***	***M***	***SD***	**r_{12}**	**r_{12C}**
Puntreeksen	10.22	2.78	11.85	2.91	.56*	.64*
Figuurreeksen	10.35	2.99	11.20	3.05	.86*	.90*
Schiftingen	10.83	2.77	11.54	2.71	.62*	.82*
Tegenstellingen	11.03	2.70	11.42	2.39	.77*	.95*
Geheugen A	11.06	2.72	13.47	2.56	.62*	.70*
Geheugen B	11.40	2.44	12.70	2.08	.67*	.77*
Gedraaide figuren	10.34	3.18	12.10	3.39	.69*	.79*
Dozen plooien	10.47	2.69	11.81	3.03	.72*	.82*
Geheime code	10.29	2.45	12.07	2.78	.68*	.81*

Noot. * = $p < .01$; r_{12} = ongecorrigeerde correlatie; r_{12C} = gecorrigeerde correlatie

Indexen en GCV

Tabel 3 heeft dezelfde opbouw als Tabel 2 maar geeft een overzicht van de betrouwbaarheidscoëfficiënten voor de BCV's en IQ/GCV.

De gecorrigeerde test-hertestcorrelatie voor de globale cognitieve vaardigheid (GCV of IQ) is .94 voor de volledige onderzoeksgroep. De gecorrigeerde test-hertestcorrelaties van de BCV indexscores zijn hoog en variëren van .75 tot .91. De hoogste test-hertestcorrelaties worden gevonden voor Gc (.91) en Gf (.90). De laagste test-hertestcorrelatie is gevonden voor Gsm (.75).

*Tabel 3: Gemiddelden en standaardafwijkingen voor de test-hertest scores en correlaties (**r_{12}** = ongecorrigeerd en en r_{12C} = gecorrigeerd) tussen beide indexscores en IQ-scores (N = 106)*

Indexen	**Eerste afname**		**Tweede afname**			
	M	***SD***	***M***	***SD***	**r_{12}**	**r_{12C}**
Gf	101.45	14.76	106.72	15.37	.86*	.90*
Gc	104.25	13.11	107.33	12.27	.79*	.91*
Gsm	105.93	12.49	117.36	11.80	.70*	.75*
Gv	100.81	13.81	110.63	15.05	.75*	.82*
Gs	102.70	11.64	110.99	13.66	.70*	.83*
IQ /GCV	**102.02**	**13.03**	**110.63**	**14.43**	**.91***	**.94***

Noot. * = p < .01; r_{12} = ongecorrigeerde correlatie; r_{12C} = gecorrigeerde correlatie; Gf = Vloeiende Intelligentie; Gc = Gekristalliseerde Intelligentie; Gsm = Kortetermijngeheugen; Gv = Visuele Informatieverwerking; Gs = Verwerkingssnelheid; IQ/GCV = Globale Cognitieve Vaardigheid gebaseerd op 5 brede cognitieve vaardigheden.

Verschilscores en effect size

Subtests

Het is niet onverwacht dat de gemiddelde scores in de tweede afname systematisch hoger liggen dan bij de eerste afname wat een leereffect of een geheugen effect kan impliceren (zie Tabel 4). Voor alle subtests is er een significant verschil tussen de eerste en tweede afname. De effectgroottes van de verschilscores (AU) zijn in het algemeen matig. Voor de Gc subtests Schiftingen en Tegenstellingen is er een klein effect. Voor de Gf subtests (Puntreeksen en Figuurreeksen), Geheugen B en Dozen plooien is er een matig effect. Het grootste leer- en geheugeneffect komt voor bij Gedraaide figuren, Geheime code en Geheugen A.

De subtests Schiftingen en vooral Tegenstellingen blijken maar vrij licht te stijgen. De subtest Geheugen A stijgt sterk wat niet zo verwonderlijk is gezien de testopdracht en de korte tijd tussen de twee testafnames. Opvallend is de stijging bij Geheime code en Gedraaide figuren, wat duidelijk een leereffect impliceert bij de hertesting.

Tabel 4: Verschilscores tussen de eerste en tweede afname en effect grootte (N = 106)

Subtesten	Verschilscores (AU)		
	M $(M_2 - M_1)$	***SD***	**Effect size**
Puntreeksen	+1.63	2.66	.61
Figuurreeksen	+0.85	1.60	.53
Schiftingen	+0.71	2.37	.30
Tegenstellingen	+0.40	1.77	.22
Geheugen A	+2.41	2.31	1.05
Geheugen B	+1.30	1.88	.69
Gedraaide figuren	+2.10	2.60	.81
Dozen plooien	+1.34	2.18	.61
Geheime code	+1.76	2.11	.83

Noot. M_2 = gemiddelde op meetmoment 2; M_1 = gemiddelde op meetmoment 1; effect size = Cohen's d

Indexen en GCV

Uiteraard treedt hier ook de stijging op bij afname 2. Dezelfde vaststellingen als bij de subtests komen hier naar voren. De gemiddelde indexscores in de tweede afname zijn significant hoger dan bij de eerste afname. Belangrijk is de vaststelling dat er gemiddeld 8.62 IQ punten winst is bij de hertesting. Er is een klein effect voor Gc, een matige effect voor de Gf en een sterk effect voor Gs, Gsm en Gv.

Relevant blijkt het verschil in 'winst ' tussen enerzijds Gc en anderzijds Gv en Gs. In de literatuur rond de Wechslerschalen (zie oa. WISC-III handleiding, blz. 54) blijkt ook dat het toenmalige PIQ en de POF bij hertesting sterker stegen dan het toenmalige VIQ en de VBF. De CoVaT-CHC BCV Gf sluit in dit opzicht eerder aan bij de BCV Gc dan bij de BCV Gv en Gs en onderscheidt zich zo ook van het WISC III P-IQ of POF (Schittekatte, 2000; Wechsler, 2005).

Tabel 5: Verschilscores tussen de eerste en tweede afname en effect grootte (N = 106)

Indexen	**Verschilscores (Indexen-IQ/GCV)**		
	M $(M_2 - M_1)$	***SD***	**Effect size**
Gf	+5.26	8.03	.66
Gc	+3.08	8.26	.37
Gsm	+11.42	9.42	.91
Gv	+9.82	10.27	1.21
Gs	+8.29	9.87	.84
IQ /GCV	+8.62	5.95	1.45

Noot. M_2 = gemiddelde op meetmoment 2; M_1 = gemiddelde op meetmoment 1; effect size = Cohen's d

Conclusie

Voor de globale cognitieve vaardigheid vinden we een zeer sterke test-hertestcorrelatie wat ook mag worden verwacht van een test die cognitieve vaardigheden meet. De test-hertestbetrouwbaarheid van de CoVaT – CHC Basisversie van het gemeten totaal IQ of GCV is .94. Op basis van de regels volgens de COTAN bij tests voor belangrijke beslissingen op individueel niveau, wijst de betrouwbaarheidscoëfficiënten van het IQ van de CoVaT-CHC Basisversie op een goede betrouwbaarheid (Evers, Lucassen, Meijer, & Sijtsma, 2010). Volgens de EFPA richtlijnen wijst de betrouwbaarheidscoëfficiënt van het CoVaT-CHC IQ op een zeer goede betrouwbaarheid (Evers et al., 2013). De test voldoet zo ruimschoots aan de eisen die men aan een goede test voor cognitieve vaardigheden mag stellen. Met dit niveau is de test geschikt voor uitspraken in functie van belangrijke beslissingen op individueel niveau.

De test-hertestbetrouwbaarheidcoëfficiënten van de BCV Indexen variëren van .75 tot .91. Op basis van de regels volgens de COTAN bij tests voor belangrijke beslissingen op individueel niveau, wijzen de betrouwbaarheidscoëfficiënten van de BCV Indexen van de CoVaT-CHC Basisversie op een goede (Gf, Gc, Gv en Gs) betrouwbaarheid met uitzondering van Gsm wat een "voldoende" krijgt (Evers et al., 2010). Dit doet echter niets af van de kwaliteit van Gsm. Zo is het juist een betrouwbaar gegeven dat de leerwinst bij deze BCV de hoogste is van de vijf BCV's. In dit kader zou de betrouwbaarheid van Gsm waarschijnlijk nog correcter geschat kunnen worden via de paralleltestmethode. Volgens de EFPA richtlijnen wijzen de betrouwbaarheidscoëfficiënten van de gemeten BCV Indexen op een goede

(Gsm) tot zeer goede (Gf, Gc, Gv,en Gs) betrouwbaarheid (Evers et al., 2013). De test - hertestbetrouwbaarheid op het niveau van de BCV laat toe deze indexen - met de gepaste voorzichtigheid - ondersteunend te hanteren bij beslissingen op individueel niveau .

Op subtestniveau liggen de betrouwbaarheidscoëfficiënten tussen .64 tot .95. Met uitzondering van Puntreeksen zijn de test-hertestcorrelaties .70 of hoger. Volgens de COTAN richtlijnen (Evers et al., 2010) en CAP vzw (Rauws, Bos, Magez, Geerinck, & Van Elsacker, 2016) zijn de betrouwbaarheidscoëfficiënten 'Voldoende' tot 'Goed' (uitgezonderd voor Puntreeksen) voor minder belangrijke beslissingen. Op zich zijn dit dus behoorlijk resultaten. Dit neemt echter niet weg dat voor belangrijke beslissingen op individueel niveau een betrouwbaarheid van méér dan. 80 een referentie is (CAP vzw e.a.).

Het advies is dan ook om enkel op niveau van de indexen of op basis van de globale cognitieve vaardigheid (IQ) belangrijke uitspraken te doen, iets wat ondersteund wordt door de theorie van het CHC – model (Flanagan & Harrison, 2012). De test-hertestbetrouwbaarheden van de subtests laten enkel toe deze scores eventueel 'mogelijk richtinggevend' te hanteren, maar dit dient met de grootste zorgvuldigheid te gebeuren. We raden het gebruik ervan niet aan, met uitzondering echter van de scores voor de subtest Figuurreeksen en eventueel Tegenstellingen, al blijft een zekere voorzichtigheid gepast bij de interpretatie.

Praktische implicaties

Tussen de eerste en tweede afname (interval van ongeveer 40 dagen) steeg het CoVaT-CHC IQ gemiddeld met ruim 8.5 IQ-punten. Dit betekent dat als men de CoVaT-CHC Basisversie met een korte tussentijd nogmaals afneemt, er een toename in de totaalscore zal plaatsvinden van gemiddeld 8.5 IQ punten. Het is dan ook af te raden om in het kader van een 'klassiek IQ' onderzoek kinderen en jongeren in een relatief korte tussentijd voor een tweede keer met de CoVaT-CHC Basisversie te onderzoeken.

Dit onderzoek sluit aan bij de literatuurstudie van Schittekatte (2000). Schittekatte toont ook een IQ stijging aan bij de hertesting. Deze stijging is logischerwijze des te groter naarmate de tijd tussen de eerste afname en de tweede korter is. Concluderend komt hij tot het besluit:" Het minimuminterval bij het hertesten van de WISC-III ligt ergens tussen negen maanden en twee jaar. Indien mogelijk moet men, om alle test-hertesteffecten te vermijden, tussen de twee afnames twee jaar laten verlopen.

Wacht men minimaal één jaar, dan zullen de invloeden van herhaalde meting beperkt zijn" (Schittekatte, 2000, pg 17).

We sluiten ons hier globaal bij aan wat de CoVaT-CHC Basisversie betreft en raden aan om een tijdsinterval tussen een test en een hertesting aan te houden van minimum 1 jaar of te hertesten met een andere intelligentietest dat ook gebaseerd op het CHC-model (Flanagan & Harrison, 2012).

De gestelde regels lijken ons zeker te gelden voor de BCV Gsm, Gv en Gs. Voor Gc lijkt het dat, zo nodig, het tijdsinterval korter mag/kan zijn. In onze studie bedroeg het effect op Gc slechts + 3 Indexpunten met een maand (ongeveer 40 dagen) tussentijd. Dit lijkt ons ook te gelden voor de subtests Figuurreeksen en Tegenstellingen. Voor dezelfde intervalperiode bedroeg het effect voor de subtests Figuurreeksen en Tegenstellingen respectievelijk slechts + 1 en + 0.5 AU punten.

Globaal kan gesteld worden dat de vastgestelde test-hertestbetrouwbaarheid toelaat de BCV indexen te hanteren - steeds met de gepaste voorzichtigheid - bij beslissingen op individueel niveau. Dit geldt zeker voor het IQ/GCV met zijn test- hertestingsbetrouwbaarheid van .94 wat wijst op een zeer sterk - excellent - betrouwbaarheidsniveau.

Referenties

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Earlbaum Associates.

Evers, A., Hagemeister, C., Høstmælingen, A., Lindley, P., Muñiz, J., & Sjöberg, A. (2013). EFPA Review model for the description and evaluation of psychological and educational tests: Test review form and notes for reviewers, version 4.2.6. Geraadpleegd op 20 januari 2017 via http://www.efpa.eu/professional-development/assessment

Evers, A., Lucassen, W., Meijer, R., & Sijtsma, K. (2010). COTAN Beoordelingssysteem voor de kwaliteit van tests. Amsterdam: NIP. Geraadpleegd op 20 januari 2017 via https://www.psynip.nl/wp-content/uploads/2016/07/COTAN-Beoordelingssysteem-2010.pdf

Flanagan, D. P. & Harrison, P.L. (Eds.) (2012). Contemporary Intellectual Assessment Third Edition: Theories, Tests, and Issues. New York: The Guilford press.

Guilford, J. P. (1956). *Fundamental statistics in psychology and education*. New York: McGraw – Hill Book Company.

Johnson, B. T. & Eagly, A. H. (2000). Quantitative synthesis of social psychological research. *CHIP Documents, 13*, 496 - 529.

Magez, W., Tierens, M., Van Huynegem, J, Van Parijs, K., Decaluwé, V., & Bos, A. (2015). *CoVaT-CHC Basisversie: Cognitieve vaardigheidstest volgens het CHC-model.* Antwerpen: Psychodiagnostisch Centrum en CAP vzw.

Magez, W. & Tierens, M. (2016) *Onderzoeksrapport moeilijkheidsgraden van items, subtests en BCV met duiding voor de praktijk.* Geraadpleegd op 1 februari 2017 via http://www.thomasmore.be/psychodiagnostisch-centrum/covat-chc-basisversie-0

Magez, M., Van Parijs, K., & Tierens, M. (2016). *Onderzoeksrapport methodologie bij de gehanteerde normschalen.* Geraadpleegd op 1 februari 2017 via http://www.thomasmore.be/psychodiagnostisch-centrum/covat-chc-basisversie-0

Rauws, G., Bos A., Magez, W., Geerinck, K., & Van Elsacker, M. (2016). *CAP-vademecum van diagnostische instrumenten en methoden voor CLB.* Brasschaat: CAP. Geraadpleegd op 20 januari 2017 via http://www.cap-vademecum.be

Schittekatte, M. (2000). Het hertesten van intelligentie bij kinderen: een literatuurstudie. *Significant*, *1-21*. Geraadpleegd op 31 januari 2017 via http://www.sig-net.be/uploads/artikels_signaal/significant_hertesting_schittekatte_2000_nr0.pdf.

Tierens, M. & Magez, W. (2016a). *Onderzoeksrapport betrouwbaarheid en betrouwbaarheidsintervallen.* Geraadpleegd op 22 november 2016 via http://www.thomasmore.be/psychodiagnostisch-centrum/covat-chc-basisversie-0

Tierens, M. & Magez, W. (2016b). *Onderzoeksrapport normering.* Geraadpleegd op 7 september 2016 via http://www.thomasmore.be/psychodiagnostisch-centrum/covat-chc-basisversie-0

Vanhoomissen, T. & Valkeneers, G. (2014). *Inductieve statistiek voor de gedragswetenschappen: Toegepaste hypothesetoetsing met SPSS.* Leuven: Acco.

Wechsler, D. (2005). *Wechsler Intelligence Scale for Children, Third edition, Nederlandstalige bewerking (WISC-III ^{NL}): Technische handleiding.* Amsterdam: Pearson Assessment en Information B.V. (bewerking door: Kort, W., Schittekatte, M., Bosmans, M., Compaan, E.L., Dekker, P.H., Vermeir, G., & Verhaeghe, P.)