

VALIDERING VAN DE CHILDREN'S ICAR BIJ LEERLINGEN VAN HET ZESDE LEERJAAR

Aantal woorden: 11.170

Eline Rombaut

Studentennummer: 02011693

Promotor: Prof. dr. Nicolas Dirix

Co-promotor: dr. Stijn Schelfhout

Begeleider: Merel Dutry

Masterproef voorgelegd voor het behalen van de graad master in de klinische psychologie

Academiejaar: 2023 – 2024

Disclaimer

In deze masterproef werd eenzelfde cognitieve test gecrossvalideerd als in de masterproef van Sari Van Haute (01809054), maar bij een jongere populatie en met een ander cross-validerend instrument. Hierdoor is enige overlap mogelijk in de methode, analyse en discussie. De tekst van deze masterproef, het uitgevoerde kwantitatieve onderzoek en de uitgevoerde analyses zijn echter individueel en kwamen in overleg met de promotor tot stand.

Dankwoord

Vier jaar geleden rondde ik mijn studies in de toegepaste psychologie af en ging ik aan de slag binnen de forensische sector. Ondanks de voldoening die ik hieruit haalde, had ik zin in meer en besloot ik voor een master klinische psychologie te gaan. Deze beslissing resulteerde in een avontuur waarin ik niet alleen mijn vaardigheden als toekomstig klinisch psycholoog verder ontwikkelde, maar ook mijzelf beter heb leren kennen. Ik wil eerst en vooral mijn ouders bedanken om mij de kans te bieden mijn dromen na te streven en mijzelf op deze manier verder te ontwikkelen. Zonder hun vertrouwen in mijn kunnen, zou ik niet staan waar ik nu sta. Bovendien wil ik ook Jarich, Julie en Margaux bedanken voor hun eindeloze geduld en onvoorwaardelijke steun. Het vertrouwen dat zij steeds opnieuw hebben getoond in mijzelf, en de bemoedigende woorden die zij mij hebben geboden wanneer ik het even niet meer zag zitten, hebben mij in staat gesteld om deze opleiding en bijhorende masterproef met trots af te ronden. Mijn leidinggevende en collega's van mijn huidige job verdienen ook een eerbetuiging; zonder hun begrip en flexibiliteit had ik dit niet gekund. Verder wens ik mijn mentor, Merel, te bedanken. Dankzij haar waardevolle feedback en uitstekende begeleiding slaagde ik erin om deze masterproef tot een goed einde te brengen. Ook mijn promotor en co-promotor, prof. dr. Nicolas Dirix en dr. Stijn Schelfhout, verdienen een dankwoord om mij de mogelijkheid te bieden dit onderzoek uit te voeren. Tot slot betuig ik mijn dank aan de deelnemende scholen en hun leerlingen; zonder hun bijdrage had dit onderzoek niet mogelijk geweest.

Met trots presenteer ik hierbij mijn masterproef: het symbolische sluitstuk van mijn studies en het begin van mijn professionele carrière als klinisch psycholoog. Het is een eer om deze mijlpaal te bereiken en ik hoop dat de inspanningen die ik heb geleverd voor dit onderzoek, een waardevolle bijdrage kunnen betekenen voor de academische wereld.

Abstract

Om tegemoet te komen aan de uitdagingen waar onderzoekers op stuiten bij het in kaart brengen van cognitieve vaardigheden op grootschalig niveau, werd de ICAR ontwikkeld. Dit openbaar toegankelijke meetinstrument werd reeds door verschillende onderzoekers wereldwijd gevalideerd, echter ontbreekt het tot op heden aan een gevalideerde versie voor kinderen. Dit onderzoek spitst zich toe op de validering van de Ch-ICAR, een specifieke versie van de ICAR, ontwikkeld voor het in kaart brengen van cognitieve vaardigheden bij kinderen tussen 11 en 14 jaar. Uit de resultaten van dit onderzoek blijkt een goede betrouwbaarheid van de Ch-ICAR in het algemeen, echter vergen de subtesten Verbaal Redeneren en Matrix Redeneren verdere aandacht. Qua interne structuur vertoont de Ch-ICAR positieve onderlinge correlaties, wat de aanwezigheid van een overkoepelende g-factor lijkt te bevestigen. De subtesten Verbaal Redeneren, Cijferreeksen en Figuur Analogieën blijken voornamelijk Gf te meten waarbij Verbaal Redeneren ook Gc meet en Figuur Analogieën tevens Gv meet. De subtest Matrix Redeneren blijkt voornamelijk Gc te meten. Verder onderzoek dient zich te focussen op replicaties met grotere steekproeven en dient verder in kaart te brengen welke brede cognitieve vaardigheden worden gemeten door de verschillende subtesten. Samengevat kan de Ch-ICAR een waardig alternatief bieden in het onderzoek naar cognitieve vaardigheden bij kinderen.

Inhoudstafel

Inleiding.....	3
Wat is intelligentie?.....	4
<i>Spearman's g-factor</i>	4
<i>Biologische basis g-factor</i>	5
<i>Stabiliteit intelligentie</i>	5
<i>Geslachtsverschillen in intelligentie</i>	6
<i>Academische prestaties en de g-factor</i>	8
<i>CHC-model</i>	9
Meetinstrumenten intelligentie.....	13
<i>De Wechsler schalen</i>	13
<i>CoVaT-CHC Basisversie</i>	14
<i>ICAR</i>	15
Huidig onderzoek	18
Methode.....	20
Participanten	20
Materialen.....	20
Opzet en procedure.....	23
Resultaten.....	25
Betrouwbaarheid	25
Verdeling Ch-ICAR totaalscores	25
Interne structuur	26
Validiteit.....	27
Geslachtsverschillen.....	28
Discussie	29
Inleiding tot discussie.....	29
Bespreking resultaten en vergelijking literatuur	30
Beperkingen en toekomstig onderzoek	33
Conclusie	34
Referenties.....	36

Inleiding

Intelligentie is een veelzijdig begrip dat de cognitieve vaardigheden van een persoon weerspiegelt, en waar sinds de 19e eeuw onderzoek naar wordt gedaan (Boake, 2002). In 1905 werd de eerste intelligentietest voor kinderen ontwikkeld, namelijk de Binet-Simon-schaal (Doll, 1917; Terman, 1916). Deze schaal wordt doorgaans beschouwd als de basis voor de moderne intelligentietest zoals we ze nu kennen (Boake, 2002; White, 2000). Het ontwikkelen van betrouwbare meetinstrumenten voor intelligentie is van cruciaal belang om een dieper inzicht te verkrijgen in de cognitieve vermogens van een persoon, waardoor individuele verschillen beter kunnen worden begrepen en benut.

Deze masterproef zal zich toespitsen op de validering van een specifieke versie van het meetinstrument ICAR, namelijk de Children's International Cognitive Ability Resource (Ch-ICAR; Dutry et al., 2023). De Ch-ICAR betreft een samenstelling van ICAR-items, bedoeld voor het meten van cognitieve vaardigheden bij kinderen tussen 11 en 14 jaar. De oorspronkelijke ICAR is een openbaar toegankelijk meetinstrument dat reeds door verschillende onderzoekers werd gevalideerd. Uit deze onderzoeken blijkt dat de ICAR een valide instrument is om cognitieve vaardigheden in kaart te brengen voor onderzoeksdoeleinden (Condon & Revelle, 2014; Dworak et al., 2021).

Het doel van deze masterproef is om de validiteit en betrouwbaarheid van de Ch-ICAR in kaart te brengen door middel van een *cross battery analysis*. De analyse gebeurt door de Ch-ICAR te vergelijken met een reeds bestaand en gevalideerd meetinstrument voor kinderen, de Cognitieve Vaardigheidstest gebaseerd op het Cattell-Horn-Carroll model van intelligentie (CoVaT-CHC; Magez et al., 2015). Een *cross battery analysis* tussen beide meetinstrumenten maakt het mogelijk om de Ch-ICAR grondig te evalueren en te bepalen of het instrument daadwerkelijk meet wat het beoogt te meten.

Alvorens dieper wordt ingegaan op de ontwikkeling van de Ch-ICAR en het huidige onderzoek, worden enkele belangrijke begrippen nader toegelicht. Allereerst volgt een beschrijving van het begrip intelligentie en zijn biologische basis, en wordt besproken hoe intelligentie evolueert doorheen de levensloop van een persoon. Verder wordt gekeken of intelligentie verschilt bij mannen en vrouwen, en wordt de link gelegd tussen de cognitieve vaardigheden van een individu en zijn academische prestaties. Vervolgens wordt de theorie rond het eerdergenoemde CHC-model uitgebreid toegelicht, alsook een overzicht geboden van de

meest relevante meetinstrumenten die de cognitieve functies van een individu in kaart brengen. Zo zullen de Wechsler schalen, de CoVaT en de ICAR aan bod komen. Tot slot volgt een toelichting inzake de meerwaarde van deze masterproef voor het onderzoeksveld en worden de verwachtingen van dit onderzoek gepresenteerd.

Wat is intelligentie?

Het begrip intelligentie, ondanks dat het wereldwijd wordt gebruikt, valt moeilijk te definiëren. Volgens Cattell (1987) stelden verschillende disciplines doorheen de geschiedenis drie adequate definities. Zo omschrijven deskundigen uit het onderwijs intelligentie als ‘de capaciteit om te leren’ (Cattell, 1987). Binnen de schoolcontext ligt de focus voornamelijk op de vaardigheid om zaken te memoriseren en toe te passen. Kinderen die over deze vaardigheid beschikken, worden doorgaans als intelligent beschouwd. Verder omschrijven anderen intelligentie als ‘de capaciteit om abstract te denken’. Deze benadering vloeit eerder voort uit de filosofische en wiskundige vakgebieden, waarbij abstract redeneren centraal staat (Cattell, 1987). Tot slot stelt Cattell (1987) dat, vanuit een evolutionair oogpunt, intelligentie verwijst naar ‘de capaciteit om zich aan te passen aan nieuwe situaties’. Intelligentie kent dus verschillende omschrijvingen, afhankelijk van welk doel voorop wordt gesteld.

Spearman's g-factor

Wanneer in dit onderzoek het begrip intelligentie aan bod komt, wordt verwezen naar de algemene intellectuele vaardigheid zoals omschreven door Spearman (Spearman, 1904). Spearman trachtte correlaties te ontdekken tussen de prestaties van personen op verschillende cognitieve taken (Brody, 1999). Naar aanleiding van Spearman's onderzoek bekroonden deskundigen hem tot pionier op het vlak van factoranalyses; een techniek die voor een ware revolutie zorgde binnen verschillende wetenschappelijke disciplines (Jensen, 2014). Spearman ontdekte door middel van factoranalyses dat er inderdaad een gemeenschappelijke factor over de prestaties op cognitieve taken heen bestaat, hetgeen volgens hem de algemene cognitieve capaciteit van een persoon reflecteert. Deze gemeenschappelijke factor noemde Spearman de ‘g-factor’ (Spearman, 1904). Spearman bevestigde hiermee de hypothese rond intelligentie waar onderzoekers zich al enige tijd over bogen. Diverse onderzoekers stelden immers dat wanneer iemand hoog zou scoren op een cognitieve taak die beroep doet op één aspect van onze cognitieve vermogens, diezelfde persoon ook hoog zou scoren op een taak dat een ander aspect van onze

cognitieve vaardigheden meet (Gottfredson, 1998). De g-factor bevestigde dus het bestaan van deze algemene intellectuele vaardigheid.

Biologische basis g-factor

Later onderzoek doet vermoeden dat de g-factor een biologische basis kent. Zo haalt Gottfredson (1998) aan dat verschillende aspecten van het brein invloed uitoefenen op de algemene cognitieve vaardigheden van een persoon. Uit onderzoek blijkt dat, na correctie voor zowel geslacht als fysieke grootte, de grootte van het brein positief correleert met intellectuele capaciteiten (Deary et al., 2022; Gottfredson, 1998; Pol et al., 2006; Posthuma et al., 2002). Verder zou de snelheid waarmee de zenuwen informatie doorsturen ook positief correleren met de intellectuele vermogens van een persoon. Daarnaast rapporteren onderzoekers correlaties tussen de kwaliteit van de hersengolven van de participanten en hun intelligentiequotiënt (IQ). Bovendien merken zij op dat personen die hoger scoren wat betreft IQ, significant minder energie nodig hebben tijdens moeilijke taken waarbij zij beroep moeten doen op hun probleemoplossend vermogen (Gottfredson, 1998).

Vanuit deze waarnemingen wordt de hypothese gesteld dat verschillen in de g-factor mogelijk voortvloeien uit de verschillen in snelheid en efficiëntie waarmee het brein informatie verwerkt op neurale niveau (Deary et al., 2022; Gottfredson, 1998). Hierbij maken de onderzoekers de opmerking dat indien deze hypothese waar zou zijn, dit zou betekenen dat omgevingsinvloeden van belang zijn in die zin dat modificatie van het brein zou kunnen leiden tot een betere cognitieve vaardigheid. De hypothese dat een hogere intelligentie gecorreleerd zou zijn met een efficiëntere werking van de neurale structuren, wordt bevestigd door Li en collega's (2009) in hun onderzoek. Zij vermoeden dat de neurale netwerkgeleiding een korter pad volgt bij intelligentere personen en zo bijdraagt aan een efficiëntere informatieverwerking. Uit de bevindingen van het onderzoek van Deary en collega's (2022) blijkt inderdaad dat een verband bestaat tussen algemene intelligentie en efficiëntere neurale activiteit in de prefrontale cortex.

Stabiliteit intelligentie

Intelligentie betreft een construct dat zich kenmerkt als relatief stabiel doorheen de levensjaren (Caniyez & Watkins, 1998). Uit een onderzoek van Salthouse (2010) blijkt echter dat enkele facetten van intelligentie afnemen wanneer de leeftijd toeneemt. Deze afname in cognitieve capaciteiten begint reeds vanaf een leeftijd van 20 jaar (Salthouse, 2010). Het betreft cognitieve vaardigheden zoals verwerkingssnelheid, redeneren, ruimtelijk inzicht en enkele

aspecten van het geheugen (Salthouse, 2010). Een ander longitudinaal onderzoek levert gelijkaardige resultaten op waaruit blijkt dat voornamelijk cognitieve taken die beroep doen op verwerkingsnelheid en probleemoplossend vermogen, het meest gevoelig zouden zijn aan een toenemende leeftijd (Schwartzman et al., 1987).

Wanneer onderzoekers de stabiliteit van intelligentie in kaart trachten te brengen, stuiten zij vaak op enkele moeilijkheden. Een eerste probleem doet zich voor in het kiezen van een meetinstrument om het IQ te meten. Tijdens longitudinale onderzoeken verandert het aanbod in meetinstrumenten zodanig dat het bijna onmogelijk blijkt om resultaten te vergelijken op basis van hetzelfde instrument (Deary, 2014). Indien het toch zou lukken om hetzelfde meetinstrument te gebruiken, kan het zijn dat de normen van dergelijk meetinstrument veranderd zijn of niet meer geschikt blijken voor het leeftijdsbereik waarin de participanten zich op dat moment bevinden (Deary, 2014). Omwille van het Flynn-effect, het fenomeen waarbij volgende generaties een betere prestatie op cognitieve testen vertonen, dienen normen regelmatig te worden bijgesteld (Flynn, 1987; Sundet et al., 2004). Het in gebruik nemen van verschillende meetinstrumenten en het berekenen van de correlaties hiertussen, zou een oplossing kunnen bieden voor bovengenoemde problemen (Deary, 2014). Indien verschillende meetinstrumenten sterk gecorreleerde scores opleveren, kan dit indiceren dat enige mate van stabiliteit in intelligentie aanwezig is, ongeacht de veranderingen in meetinstrumenten.

Geslachtsverschillen in intelligentie

Wat betreft de vraag of mannen en vrouwen verschillen vertonen qua cognitieve vaardigheden lopen de antwoorden sterk uiteen. De afgelopen decennia onderzochten verschillende onderzoekers deze vraag, hetgeen conflicterende onderzoeksresultaten opleverde. Pezzuti en Orsini (2016) trachten een antwoord te bieden op deze vraag door de resultaten tussen jongens en meisjes van 6 tot 16 jaar te vergelijken, dewelke zijn verzameld door middel van de gestandaardiseerde intelligentietest Wechsler Intelligence Scale for Children-IV (WISC-IV). Zowel de WISC-IV als de andere Wechsler schalen worden verder in deze masterproef besproken. De IQ-score, zoals berekend door de WISC-IV, wordt onderverdeeld in vier indices: Verbaal Begrip, Perceptueel Redeneren, Werkgeheugen en Verwerkingsnelheid. Pezzuti en Orsini (2016) concluderen dat, met de geobserveerde variabelen, geen geslacht gerelateerde verschillen kunnen worden weerhouden wat betreft Verbaal Begrip, Perceptueel Redeneren, Werkgeheugen en totaal IQ-score. Zij nemen slechts significante geslacht gerelateerde verschillen waar in de gemiddelde scores op latente variabelen waarbij zij stellen dat jongens

hoger zouden scoren op Verbaal Begrip in vergelijking met meisjes, terwijl meisjes hoger zouden scoren op Verwerkingssnelheid (Pezzuti & Orsini, 2016).

Een ander onderzoek, gevoerd door Daseking, Petermann en Waldmann (2017), tracht mogelijke geslachtsverschillen te onderzoeken aan de hand van de Wechsler Adult Intelligence Scale-IV (WAIS-IV). De beoogde leeftijdsgroep voor WAIS-IV loopt van 16 tot 89 jaar en bestaat uit dezelfde opdeling qua indices als de WISC-IV, namelijk Verbaal Begrip, Perceptueel Redeneren, Werkgeheugen en Verwerkingssnelheid. Uit dit onderzoek blijkt dat vrouwen op alle indices lager scoren dan hun mannelijke tegenhangers, met uitzondering van de index Verwerkingssnelheid (Daseking et al., 2017). De onderzoekers nuanceren deze resultaten met de opmerking dat het om een klein effect gaat. Daarnaast stellen zij de vraag of de schoolse achtergrond van een persoon invloed uitoefent op de testresultaten. Dit bleek een significant groter effect op te leveren dan het effect van geslacht op de onderzoeksresultaten (Daseking et al., 2017). De onderzoekers maken hierbij de nuance dat deze resultaten mogelijk onderhevig kunnen zijn aan bepaalde bias, zoals het selecteren van bepaalde items, en dat het niet is aangetoond dat de geobserveerde verschillen voortvloeien uit een werkelijk effect van de onderzochte variabelen op de onderzoeksresultaten (Daseking et al., 2017). Uit een andere studie, gevoerd door Weiss en collega's (2002), blijkt dat vrouwen hoger zouden scoren wat betreft verbale vaardigheden, terwijl mannen het beter zouden doen qua visuo-spatiale taken. Ook zij maken de nuance dat de effecten eerder klein zijn (Weiss et al., 2002).

De bekomen onderzoeksresultaten met betrekking tot mogelijke geslachtsverschillen conflicteren vaak wat betreft de domeinen waar mannen en vrouwen hoger op zouden scoren ten opzichte van elkaar. Dit doet de vraag rijzen in hoeverre we hier een uitspraak kunnen over doen en of de cognitieve prestaties van personen wel degelijk geslachtsverschillen zouden vertonen. Een review van Hyde (2014) biedt een overzicht van verschillende meta-analyses van de afgelopen decennia die dergelijke geslachtsverschillen in kaart trachten te brengen. Hierbij haalt zij een review aan van Maccoby en Jacklin (1974) waarin de onderzoekers beweren dat geslachtsverschillen aanwezig zijn op het vlak van mathematische, spatiale en verbale cognitieve taken. Maccoby en Jacklin (1974) stellen dat mannen hoger zouden scoren wat betreft mathematische en spatiale taken, terwijl vrouwen hoger zouden scoren qua verbale taken. Echter toont een andere meta-analyse van Lindberg, Hyde, Petersen en Linn (2010) dat geen statistisch significant verschil kan worden waargenomen wat betreft de prestaties van mannen en vrouwen op mathematische taken. Een onderzoek van Ardila, Rosselli, Matute en Inozemtseva (2011) vergelijkt de prestaties van jongens en meisjes tussen 5 en 16 jaar op verschillende cognitieve

taken die peilen naar verbale en spatiale vaardigheden. Ook hun resultaten bevestigen de afwezigheid van een significant geslachtsverschil wat betreft prestaties op taken die de cognitieve vaardigheden meten (Ardila et al., 2011).

Hyde (2014) komt in haar review tot de conclusie dat veel bewijs voorhanden is dat de geslachtsgelijke hypothese steunt, waarbij zij stelt dat beide geslachten over het algemeen gelijke prestaties vertonen. Op basis van de meta-analyses die zij aanhaalt, stelt zij dat eerder kleine ($d = 0.20$) of insignificante ($d \leq 0.10$) geslachtsverschillen worden waargenomen op het gebied van rekenkundige prestaties en verbale vaardigheden, terwijl gemiddelde ($d = 0.50$) of grote ($d = 0.80$) verschillen worden waargenomen op 3D mentale rotatie. In haar review vermeldt Hyde (2014) tevens andere gebieden waarop dergelijke geslachtsverschillen worden geconstateerd; zoals temperament, beloningsgevoeligheid, en andere kenmerken van een persoon. Hierop stelt Cole (2009) dat geslachtseffecten nooit afzonderlijk mogen worden bekeken en dienen te worden begrepen in zijn context. Daaruit kunnen we concluderen dat ieder persoon een complex samenspel is van kenmerken en dat mogelijke waargenomen verschillen in cognitieve vaardigheden niet louter vanuit een geslachtsverschil te verklaren zijn (Cole, 2009; Hyde, 2014; Lindberg et al., 2010).

Academische prestaties en de g-factor

Rond het verband tussen de g-factor en academische prestaties van een individu bestaat reeds veel onderzoek. Een onderzoek van Spinath en collega's (2006) tracht de belangrijkste predictor voor academische prestaties te achterhalen. Uit hun onderzoek blijkt de g-factor de belangrijkste predictor voor de schoolse prestaties van een individu (Spinath et al., 2006). Een longitudinaal onderzoek van Deary en collega's (2007) levert soortgelijke resultaten op. In hun onderzoek analyseren zij het verband tussen cognitieve vaardigheden op de leeftijd van 11 jaar en academische prestaties op de leeftijd van 16 jaar. Uit de onderzoeksresultaten blijkt dat de correlatie tussen deze twee factoren .81 bedraagt (Deary et al., 2007). Deze bevindingen liggen in lijn met de resultaten uit andere onderzoeken die dit verband onderzoeken (Lu et al., 2011; Petrill et al., 2000; Roth et al., 2015). Deary en collega's (2007) maken bovendien de vergelijking tussen meisjes en jongens om zo te achterhalen of sprake kan zijn van mogelijke geslachtsverschillen. Uit hun bevindingen blijkt dat meisjes significant beter scoren op alle vakken, met uitzondering van fysica (Deary et al., 2007). Verder testen zij of dit effect te wijten valt aan de waargenomen betere prestaties van meisjes op het vlak van verbale vaardigheden, hetgeen niet het geval blijkt te zijn (Deary et al., 2007). Fergusson en Horwood (1997) wierpen

reeds de hypothese op of meisjes beter presteren op school omwille van het verschil in gedrag in de klas tussen jongens en meisjes. Ondanks dat de g-factor de belangrijkste predictor zou zijn voor academische prestaties, kan deze niet alle variantie verklaren (Deary et al., 2007). Deary en collega's (2007) vermoeden dat de rest van de variantie te wijten valt aan verschillende andere factoren zoals de mate van abstinentie, persoonlijkheidstrekken, opvoeding, leerkrachtkenmerken en motivatie van de student. Ondanks dat academische prestaties onderhevig zijn aan de invloed van verschillende factoren, lijken de huidige beschikbare onderzoeken te bevestigen wat Spearman in zijn tijd al vermoedde: dat de g-factor zeer sterk correleert met academische prestaties van een persoon (Deary et al., 2007; Lu et al., 2011; Petrill et al., 2000; Roth et al., 2015; Spinath et al., 2006).

CHC-model

Dankzij de ontdekking van de eerder besproken g-factor van Spearman (1904) en het factor analytische werk van Thurstone (1931), ontwikkelde Raymond Cattell in de jaren 40 het dichotome Gf-Gc model (Cattell, 1941; Flanagan & Dixon, 2014; Tierens & Van Huynegem, 2017). Cattell ging ervan uit dat de cognitieve vaardigheid van een persoon op te delen valt in twee soorten vaardigheden; vloeiende intelligentie (Gf) en gekristalliseerde intelligentie (Gc). In de jaren 60 voegde John Horn hier vier bijkomende vaardigheden aan toe: visuele informatieverwerking (Gv), kortetermijngeheugen (Gsm), langetermijngeheugen (Glr) en verwerkingssnelheid (Gs) (Horn, 1965). Later voegde hij nog auditieve informatieverwerking (Ga) en reactiesnelheid (Gt) toe aan zijn theoretisch model (Flanagan & Dixon, 2014). Vanuit Horns bijdrage aan Cattells Gf-Gc theorie, werd het Cattell-Horn Gf-Gc model geboren (Flanagan & Dixon, 2014).

In de jaren 90 publiceerde John Carroll een survey van factor analytische studies met betrekking tot de cognitieve vaardigheden van de mens, waarbij hij zijn *three-stratum theory* introduceerde (Carroll, 1993). Carroll deelde de cognitieve vaardigheden van de mens op in drie strata. De brede vaardigheden plaatste hij onder stratum II en omvatten vaardigheden zoals Gf en Gc, de twee originele factoren van het Gf-Gc model zoals opgesteld door Cattell (Carroll, 1993; Cattell, 1941). Volgens Carroll worden brede vaardigheden gedefinieerd als “de fundamentele kenmerken van een individu die een breed scala aan gedragingen binnen een specifiek domein kunnen beheersen of beïnvloeden” (Carroll, 1993). Deze brede vaardigheden zouden de nauwe vaardigheden omvatten, dewelke Carroll onder stratum I plaatste. Nauwe vaardigheden representeren volgens Carroll “geavanceerde specialisatie van vaardigheden, vaak

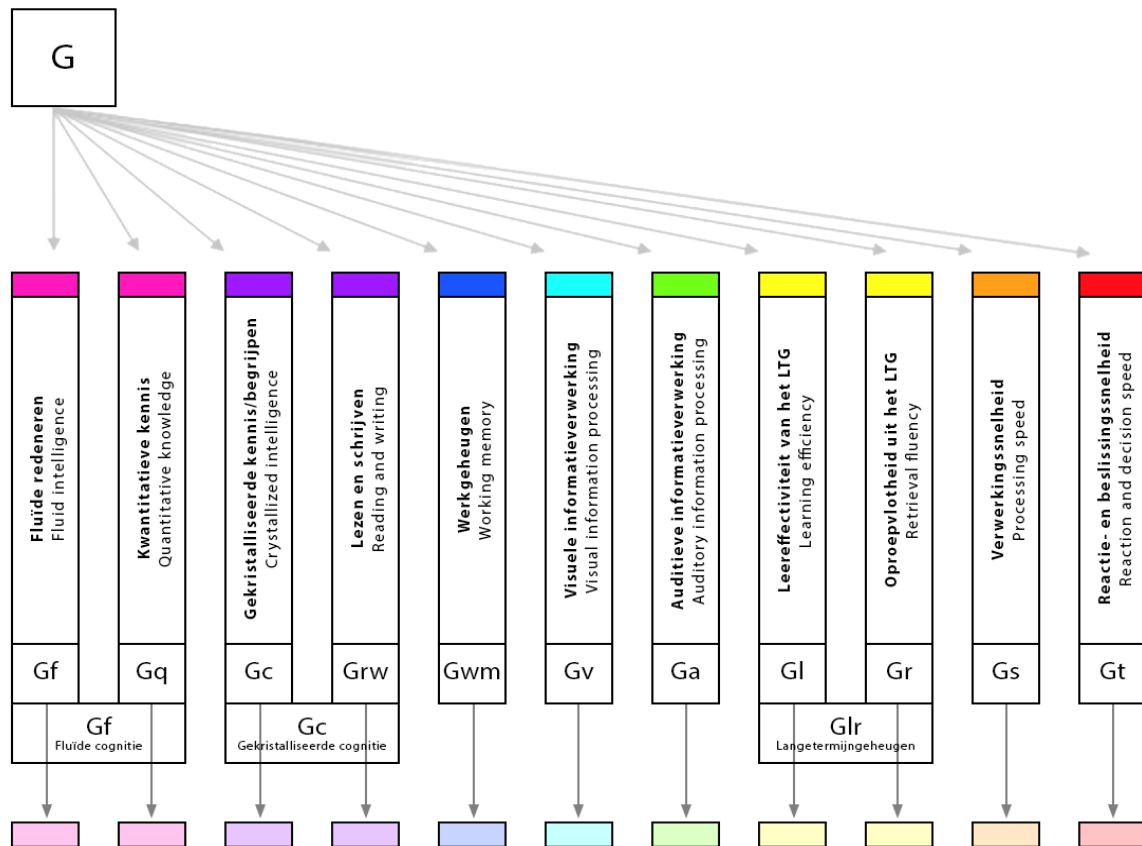
op zeer specifieke wijze, die de resultaten weerspiegelen van leerprocessen en ervaring” (Carroll, 1993). Overkoepelend aan deze twee strata plaatste Carroll de g-factor van Spearman op stratum III, dewelke de algemene factor van intelligentie representeert en betrokken is bij complexe hogere-orde cognitieve processen (Carroll, 1993; Flanagan & Dixon, 2014).

Het Cattell-Horn Gf-Gc model en Carrolls *three-stratum theory* bevatten zowel belangrijke gelijkenissen als structurele verschillen (Flanagan & Dixon, 2014). Ondanks deze verschillen zag Kevin McGrew het potentieel in de combinatie van deze twee modellen. McGrew stelde een geïntegreerd Cattell-Horn en Carroll model voor, teneinde een classificatiesysteem te ontwikkelen dat als basis kon dienen voor toekomstig onderzoek naar cognitieve vaardigheden (Flanagan & Dixon, 2014; McGrew, 1997). Daarop stelden McGrew en Flanagan (1998) een gereviseerde versie van het geïntegreerde model voor, dewelke later werd verfijnd door Flanagan, McGrew en Ortiz (2000). Het geïntegreerde model werd zowel door John Horn als John Carroll geaccepteerd waardoor het nu bekend staat als het Cattell-Horn-Carroll-model (CHC-model; Flanagan & Dixon, 2014).

Het uiteindelijke CHC-model, afgebeeld in Figuur 1, kent drie niveaus of strata (Magez et al., 2015; Schneider & McGrew, 2012). De g-factor van Spearman (1904) behoort tot stratum III, dit niveau betreft de algemene intelligentie (Carroll, 1993). Stratum II omvat de brede cognitieve vaardigheden (BCV). Deze BCV's omvatten vloeiende intelligentie (Gf), kwantitatieve kennis (Gq), gekristalliseerde intelligentie (Gc), lezen en schrijven (Grw), kortetermijngeheugen (Gsm), visuele informatieverwerking (Gv), auditieve informatieverwerking (Ga), langetermijngeheugen (Glr), verwerkingssnelheid (Gs) en reactiesnelheid (Gt) (Carroll, 1993; Flanagan & Harrison, 2012; Horn, 1965; Magez et al., 2015). Samen bepalen deze vaardigheden hoe een persoon zal presteren op verschillende domeinen (Carroll, 1993). De BCV's worden in stratum I verder opgedeeld in nauwere vaardigheden (Carroll, 1993). Deze nauwere vaardigheden omvatten specifieke vaardigheden die nodig zijn voor de BCV's (Carroll, 1993; Magez et al., 2015). Het CHC-model wordt ondertussen breed erkend als een uitgebreide en empirisch ondersteunde theorie voor de structuur van cognitieve vaardigheden (Flanagan & Dixon, 2014; Flanagan & Harrison, 2012). Het wordt frequent toegepast om intelligentietesten te structureren en de resultaten te duiden, om die reden zal het CHC-model ook in deze masterproef als basis dienen.

Figuur 1

CHC-model



Magez, W. (2021). Bouwstenen en structuur voor een cognitief persoonlijkheidsprofiel: een CHC – benadering. In: Magez, W., Rauws, G. & Bos, A. (Red.). *Diagnostiek van cognitieve vaardigheden en aansluitend handelen binnen het CHC-denkkader*. Brasschaat: CAP vzw

Deze masterproef zal zich toespitsen op drie BCV's: Gf, Gc en Gv. Zodoende worden deze vaardigheden kort besproken. Gf betreft het vermogen tot redeneren om problemen op te lossen (McGrew, 2009). Meer specifiek gaat het om problemen waar de persoon nooit eerder in contact mee kwam en waarvoor hij dus geen beroep kan doen op automatische processen om het probleem op te lossen (Magez et al., 2015). De nauwe vaardigheden die hiermee samenhangen omvatten redeneren, inductie en kwantitatief redeneren (Flanagan & Harrison, 2012; Magez et al., 2015; McGrew, 2009; Schneider & McGrew, 2012). Samengevat hebben deze nauwe vaardigheden betrekking op het vermogen om vanuit gekende regels een nieuw probleem op te lossen, onderliggende eigenschappen te achterhalen, en te redeneren met wiskundige concepten (Magez et al., 2015). Verder bestaat Gc uit de verworven kennis van een persoon en hoe deze persoon zijn kennis toepast in het dagelijkse leven (Flanagan & Harrison, 2012; Magez et al.,

2015; McGrew, 2009; Schneider & McGrew, 2012). De bijhorende nauwe vaardigheden omvatten taalontwikkeling, lexicale kennis, luisteren, algemene kennis, culturele kennis, wetenschappelijke kennis, geografische kennis, communicatieve vaardigheden, grammaticale vaardigheden, kennis van vreemde talen en geschiktheid voor vreemde talen (Magez et al., 2015; McGrew, 2009). Tot slot staat Gv voor het vermogen om visuele stimuli te verwerken en bewerken. Nauwe vaardigheden die hierbij horen zijn visueel geheugen, visuele voorstelling, visuele closure van snelheid en flexibiliteit, versnelde rotaties van ruimtelijke rotaties, scannen, perceptuele integratie van visuele patronen, schatting van lengte, weerstand tegen visuele illusies, alterneren tussen visuele prikkels en verbeelding (Magez et al., 2015; McGrew, 2009).

Het belang van een classificatiesysteem zoals het CHC-model werd reeds erkend door McGrew (Flanagan & Dixon, 2014; McGrew, 1997). Een goed classificatiesysteem brengt ons namelijk in de mogelijkheid om complexe fenomenen te ordenen, bovendien biedt het volgens Schneider en McGrew (2012) “een overzicht van eerdere bevindingen waardoor het hiaten in onze kennis blootlegt die verder onderzoek vereisen”. Daarnaast zorgt het voor een universeel referentiekader voor onderzoekers (Schneider & McGrew, 2012). Dergelijk referentiekader levert criteria waarop onderzoekers zich kunnen baseren om de objectiviteit zo goed als mogelijk te waarborgen. Dit zorgt voor meer uniformiteit van de onderzoeksresultaten, wat het eenvoudiger maakt vergelijkingen te maken tussen verschillende onderzoeken (Magez et al., 2015). Schneider en McGrew (2012) stellen dat een valide theorie resistent dient te zijn tegen kritische onderzoeksinspanningen die het proberen weerleggen, wat het CHC-model na bijna een eeuw onderzoek uitvoerig heeft bewezen. Dankzij de talloze onderzoeken die zijn gevoerd naar het CHC-model kunnen we op dit moment stellen dat het model zorgt voor een evidence-based kader waar onderzoekers van verschillende disciplines beroep op kunnen doen (Magez et al., 2015). Daarnaast kan het de basis bieden voor de ontwikkeling van nieuwe meetinstrumenten (Keith & Reynolds, 2010). Het CHC-model toont tevens op pragmatisch niveau zijn meerwaarde. De categorisatie van cognitieve vaardigheden stelt onderzoekers in staat om een gedetailleerd cognitief profiel te ontwikkelen in plaats van louter te rapporteren over een IQ-score. Dit biedt meer mogelijkheden voor onderzoek naar onder andere leerproblemen of neuropsychologische aandoeningen (Magez et al., 2015). Door middel van een dergelijk cognitief profiel kunnen onderzoekers uitspraken doen over de cognitieve sterktes en zwaktes van individuen. Dit maakt het mogelijk om adequaat in te spelen op de cognitieve uitdagingen die personen ervaren, waardoor een geïndividualiseerde begeleiding kan worden aangeboden.

Tot slot omvat het CHC-model enkele BCV's die op niet-talige wijze kunnen getoetst worden, wat de mogelijkheid biedt om meer cultuurfair te testen (Magez et al., 2015).

Meetinstrumenten intelligentie

Sinds het midden van de 19^e eeuw streven deskundigen naar het in kaart brengen van cognitieve vaardigheden bij individuen (Boake, 2002). Zo introduceerde de Amerikaan James McKeen Cattell in 1890 het begrip 'mentale test'. In 1905 ontwikkelden Alfred Binet en Theodore Simon de eerste bruikbare intelligentietest, die bekend staat als de Binet-Simon-schaal (Boake, 2002; Doll, 1917). Deze schaal heeft een basis geleverd voor intelligentietesten zoals we ze nu kennen. Tot op heden zijn er in de Wechsler schalen items terug te vinden die hun oorsprong kennen in de Binet-Simon-schaal (Boake, 2002). In het volgende deel van deze masterproef worden beknopt enkele relevante meetinstrumenten besproken, namelijk de Wechsler schalen, de CoVaT-CHC Basisversie en de ICAR.

De Wechsler schalen

De meest gebruikte meetinstrumenten voor cognitieve vaardigheden zijn de Wechsler schalen (Magez et al., 2015). De Wechsler schalen bestaan momenteel uit vier verschillende intelligentietesten. De Wechsler Preschool and Primary Scale of Intelligence (WPPSI) kent een leeftijdsbereik van 2 tot 6 jaar (Watkins & Beaujean, 2014), de Wechsler Intelligence Scale for Children (WISC) is bedoeld voor kinderen van 6 tot 16 jaar (Wilson et al., 2023), de Wechsler Abbreviated Scale of Intelligence (WASI) kan gebruikt worden bij personen van 6 tot 89 jaar (Ryan et al., 2003) en tot slot kent de Wechsler Adult Intelligence Scale (WAIS) een leeftijdsbereik van 16 tot 89 jaar (Benson et al., 2010)

De brede toepasbaarheid van de Wechsler schalen heeft zonder twijfel sterk bijgedragen aan hun populariteit. De Wechsler schalen brengen echter ook enkele nadelen met zich mee (Bisconer & Suttie, 1998; Flanagan & Kaufman, 2004). Ondanks het wijdverspreide gebruik van deze intelligentietesten blijkt dat de Wechsler schalen over een onvoldoende betrouwbaarheid beschikken wanneer onderzoekers lage IQ-scores willen meten. Bovendien zijn de Wechsler schalen ontwikkeld door Westerse deskundigen, wat voor een sterke culturele bias zorgt (Bisconer & Suttie, 1998; Flanagan & Kaufman, 2004). Dit zorgt ervoor dat wanneer diagnostici de Wechsler schalen gebruiken om de intelligentie te meten van personen die niet tot de Westerse cultuur behoren, dit mogelijk een vertekend beeld van hun cognitieve vermogens oplevert. Verder vereisen de Wechsler schalen deskundigheid van de afnemer (Bisconer & Suttie, 1998;

Flanagan & Kaufman, 2004). De testen zijn volledig gestandaardiseerd zodat de afname steeds op dezelfde wijze verloopt. Op die manier tracht men de invloeden van externe factoren te beperken. Bijgevolg dient de deskundige die de testen afneemt deze zeer goed in te oefenen. Wanneer onderzoekers de Wechsler schalen wensen te gebruiken voor wetenschappelijke doeleinden dringen bijkomende problemen zich op. De schalen zijn namelijk zeer tijdrovend om af te nemen, wat de verzameling van grote data samples bemoeilijkt (Nelissen et al., *In preparation*). Tot slot kozen de ontwikkelaars van de Wechsler schalen ervoor om de brongegevens waarop de schalen ontwikkeld zijn geheim te houden. Dit maakt dat deze data beschermd worden door auteursrechten (Dworak et al., 2021). Dit zorgt ervoor dat het gebruik van de schalen een erg hoge financiële kost met zich meebrengt (Nelissen et al., *In preparation*). Bovengenoemde zaken bemoeilijken het gebruik voor wetenschappelijk onderzoek aanzienlijk.

CoVaT-CHC Basisversie

Het CHC-model stelde Vlaamse onderzoekers in staat om een CHC-cognitieve vaardigheidstest te ontwikkelen: de CoVaT-CHC Basisversie (Magez et al., 2015). Deze test meet vijf BCV's: Gf, Gc, Gv, Gsm, en Gs. De beslissing om zich tot deze vijf BCV's te beperken kent verschillende redenen. Allereerst dient een intelligentietest, teneinde als kwaliteitsvol en betrouwbaar te worden beschouwd, minstens vier BCV's binnen het CHC-model te meten waaronder zeker Gf en Gc (Tierens & Van Huynegem, 2017). Ten tweede zijn de andere BCV's die de CoVaT omvat, zijnde Gv, Gsm en Gs, mee opgenomen omwille van hun schoolcurriculumoverstijgende karakter. BCV's zoals Gq en Grw zijn eerder schoolcurriculumgebonden en bovendien bestaan hiervoor specifieke meetinstrumenten die aanvullend aan de CoVaT kunnen worden afgenomen (Tierens & Van Huynegem, 2017). Daarnaast stellen Tierens en Van Huynegem (2017) dat de resterende BCV's Ga, Glr en Gt voorlopig niet worden opgenomen omwille van het tijdrovende aspect om deze vaardigheden te operationaliseren. Om elke BCV te meten maakt de CoVaT telkens gebruik van twee subtesten, met uitzondering van Gs die slechts door één subtest wordt gemeten (Magez et al., 2015; Tierens & Van Huynegem, 2017). De totale gewogen som van de BCV's gemeten door de CoVaT kan uitgedrukt worden in de Globale Cognitieve Vaardigheid, kortweg GCV genoemd (Magez, 2016). Deze masterproef zal trachten Gf, Gc, en Gv te meten en bijgevolg gebruik maken van de subtesten Puntreeksen en Figuurreeksen (Gf), Schiftingen en Tegenstellingen (Gc) en tot slot Gedraaide Figuren en Dozen Plooien (Gv) (Magez & Tierens, 2015)

De CoVaT is geschikt voor kinderen vanaf 9 jaar en 6 maanden tot en met 13 jaar en 11 maanden. Betreffende normering biedt deze test informatie op leeftijds- en klasniveau, voor zowel gewoon als buitengewoon onderwijs. Dankzij de beschikbare klasnormen kunnen onderzoekers uitspraken doen over de prestaties van kinderen tegenover hun medeleerlingen. Klasnormen geven echter nooit een zuiver beeld omwille van de grote variatie aan leeftijden binnen één klas (Tierens & Van Huynegem, 2017). De leeftijdsnormen komen tegemoet aan deze beperking en bieden de mogelijkheid om alsnog uitspraken te doen over de algemene cognitieve capaciteit van een kind en deze te vergelijken met leeftijdsgenoten (Tierens & Van Huynegem, 2017).

Het gebruik van de CoVaT brengt enkele voordelen met zich mee. Eerst en vooral kan deze intelligentietest zowel individueel als in groep worden afgenomen (Tierens & Van Huynegem, 2017). Dit maakt dat een groot aantal kinderen in één keer een afname kan volbrengen, wat de testafname minder tijdrovend maakt voor onderzoekers. De afnameduur van de totale test bedraagt doorgaans 2,5 à 3 uur. Daarnaast concluderen Tierens en Van Huynegem (2017) dat de CoVaT over de mogelijkheid beschikt om deze cultuurfair af te nemen. Geen enkel diagnostisch instrument zal volledig cultuurvrij zijn, echter zorgt de constructie van de test ervoor dat deze zo cultuurfair mogelijk kan worden afgenomen. Dit kan dankzij de beschikbaarheid van de niet-talige subtesten, de mogelijkheid om de items zelf te kiezen, en de ondersteunende diapresentatie die visueel illustreert hoe items kunnen worden opgelost (Tierens & Van Huynegem, 2017). In tegenstelling tot de Wechsler schalen vereist de afname van de CoVaT-CHC Basisversie geen uitgebreide deskundigheid. Onderzoekers kunnen de intelligentietest eenvoudig afnemen zonder hiervoor een training te moeten volgen. Bovendien maakt de CoVaT geen gebruik van instap-, afbreek- of omkeerregels (Tierens & Van Huynegem, 2017). Tot slot betreft de CoVaT een Vlaamse test waarvan de kostprijs voor aankoop eenmalig 495 euro bedraagt, waarna de gebruiker al het benodigde materiaal permanent tot zijn beschikking heeft (PDC Thomas More Antwerpen & CAPvzw, 2018).

ICAR

De International Cognitive Ability Resource, kortweg ICAR, tracht met zijn *open access* karakter een oplossing te bieden voor de problemen waar onderzoekers op stuiten bij het gebruik van cognitieve meetinstrumenten voor wetenschappelijke doeleinden. Het belang van de beschikbaarheid van openbaar toegankelijke instrumenten voor onderzoeksdoeleinden werd reeds door verschillende onderzoekers aangehaald (Condon & Revelle, 2014; Dworak et al.,

2021; Revelle et al., 2020). Zo blijkt dat een groot aantal instrumenten dat momenteel op de markt is in het kader van onderzoek naar cognitieve vaardigheden, wordt beschermd door auteursrechten (Dworak et al., 2021). Bijgevolg kunnen onderzoekers deze instrumenten niet vrij gebruiken. Wanneer instrumenten auteursrechtelijk beschermd zijn, brengt dit vaak een hoge kostprijs met zich mee voor andere onderzoekers die het instrument willen gebruiken (Dworak et al., 2021). Bovendien wordt bij het gebruik van dergelijke instrumenten vaak een kost per afname gerekend, dit betekent dat de kostprijs oploopt naarmate meer afnames plaatsvinden (Dworak et al., 2021). Om deze reden kunnen onderzoekers opteren om kleinere steekproeven te gebruiken teneinde de kostprijs te drukken (Revelle et al., 2020). Het gebruik van kleinere steekproeven heeft als nadeel dat de generaliseerbaarheid van de bevindingen beperkter is en dus weinig adequate uitspraken kunnen worden gedaan over de gehele populatie (Revelle et al., 2020). Bovendien blijkt een groot aantal van de beschikbare openbaar toegankelijke instrumenten onvoldoende gevalideerd (Dworak et al., 2021). Om deze problemen aan te pakken, ontwikkelden onderzoekers de ICAR (Condon & Revelle, 2014).

Het doel van de ICAR is een openbaar toegankelijk meetinstrument ter beschikking stellen waar onderzoekers uit alle onderzoeksdomeinen gebruik van kunnen maken zonder eerdergenoemde restricties (Condon & Revelle, 2014; Dworak et al., 2021). De ontwikkelaars van de ICAR wensen een collaboratief platform te bieden waar onderzoekers samen kunnen werken aan de ontwikkeling, verfijning en validering van het instrument (Revelle et al., 2020). Verder vereist het afnemen van de ICAR geen klinische achtergrond van de onderzoeker, wat maakt dat deze door verschillende soorten deskundigen kan worden gebruikt. Dworak en collega's (2021) hopen dat de ontwikkeling van de ICAR kan bijdragen aan onderzoeksdomeinen zoals onderwijs, economie, psychologie, biomedisch onderzoek en onderzoek naar cognitief functioneren. Onderzoeksdomeinen die voorheen het gebruik van traditionele meetinstrumenten voor cognitieve vaardigheden niet konden veroorloven, kunnen hier nu ook op laagdrempelige wijze gebruik van maken (Dworak et al., 2021). Dit wordt volgens Dworak en collega's (2021) reeds waargenomen in de praktijk, zo stellen zij vast dat een groot deel van de huidige ICAR-gebruikers bestaat uit graduaatstudenten. De openbare toegankelijkheid van de ICAR zorgt er namelijk voor dat onderzoekers over een goedkoop alternatief beschikken om cognitieve vaardigheden in kaart te brengen. Bovendien, doordat dit instrument online kan worden afgenomen, beperken afnames zich niet tot experimentele settings (Condon & Revelle, 2014). Dit betekent ook dat onderzoekers grote groepen tegelijk kunnen testen, wat zorgt voor een grote hoeveelheid aan bruikbare data (Dworak et al., 2021). Verder

kunnen deskundigen dankzij de openbare toegankelijkheid van de ICAR de inhoud van de test beter controleren en wordt zodanig de kwaliteit ervan gewaarborgd. Dit stelt onderzoekers in staat om de onderlinge verbanden tussen de subtesten beter in kaart te brengen (Condon & Revelle, 2012). Doordat de ICAR de mogelijkheid biedt om zelf te kiezen welke items worden opgenomen in een onderzoek, kan bijgevolg de moeilijkheidsgraad van de testafname bepaald worden (Dworak et al., 2021). Op die manier kunnen deskundigen gericht de domeinen onderzoeken die zij van belang achten en hun testinhoud afstemmen op hun doelpubliek. Bovengenoemde zaken maken van de ICAR een flexibel meetinstrument dat breed kan worden ingezet.

De ICAR betreft een tool voor onderzoeksdoeleinden waarmee onderzoekers een beeld kunnen scheppen van de algemene cognitieve vaardigheden van een persoon (Condon & Revelle, 2014; Dworak et al., 2021). De eerste versie van de ICAR kreeg de naam ICAR60, verwijzend naar de 60 items waaruit de oorspronkelijke versie bestond (Condon & Revelle, 2014). De ICAR60 omvat vier subtesten: Letter- en Cijferreeksen, Verbaal Redeneren, Matrix Redeneren en Driedimensionale Rotatie (Condon & Revelle, 2014). Een uitgebreid validatie-onderzoek van Condon en Revelle (2014) bracht onder andere de betrouwbaarheid van de ICAR60 in kaart. Uit hun resultaten blijkt dat zowel de test in zijn geheel als de subtest Driedimensionale Rotatie een goede betrouwbaarheid vertonen. Alsook blijken de subtesten Letter- en Cijferreeksen en Verbaal Redeneren een bevredigende mate van betrouwbaarheid te vertonen, terwijl de subtest Matrix Redeneren slechts een marginaal voldoende betrouwbaarheid laat zien (Condon & Revelle, 2014). Verder deden Nelissen en collega's (*In preparation*) onderzoek naar de Vlaamse versie van de ICAR. Hieruit blijkt eveneens een bevredigende mate van betrouwbaarheid wat betreft de integrale test, alsook de subtesten Driedimensionale Rotatie, Letter- en Cijferreeksen en Matrix Redeneren. De subtest Verbaal Redeneren daarentegen toont een eerder twijfelachtige betrouwbaarheid in de Vlaamse versie (Nelissen et al., *In preparation*). De onderzoekers verklaren deze bevindingen door te stellen dat enkele items van Verbaal Redeneren eerder basisvaardigheden op het gebied van wiskunde en algemene kennis dan de vaardigheid van verbaal redeneren lijken te meten.

Een ander onderzoek, gevoerd door Young en Keith (2020), onderzocht de convergente validiteit van de ICAR16. De ICAR16, ook bekend als de ICAR Sample Test, is een verkorte subtest van de oorspronkelijke ICAR60 (Condon & Revelle, 2014). De ICAR16 bestaat uit telkens vier items van de subtesten van de oorspronkelijke ICAR60 (Condon & Revelle, 2014). Young en Keith (2020) onderzochten de convergente validiteit van de ICAR16 door deze te

vergelijken met een ‘gouden standaard’ testbatterij in het onderzoeksveld naar cognitieve vaardigheden: de WAIS-IV. Zij trachtten te achterhalen welke BCV’s gemeten worden door de verschillende subtesten van de ICAR16. Uit hun resultaten blijkt dat de subtest Letter- en Cijferreeksen vooral Gf meet, terwijl de subtesten Verbaal Redeneren, Matrix Redeneren en Driedimensionale Rotatie voornamelijk Gv lijken te meten. Gezien deze verrassende resultaten maken de onderzoekers hierbij de nuance dat replicaties van het onderzoek met grotere steekproeven nodig zijn om deze bevindingen te staven (Young & Keith, 2020).

Op dit moment draait het ICAR-project op volle toeren en is het instrument nog volop in ontwikkeling waarbij het wereldwijd door onderzoekers wordt getest en gevalideerd (Condon & Revelle, 2012). Momenteel omvat de ICAR 19 types subtesten, met een pool van meer dan 1000 items, die verschillende cognitieve vaardigheden meten zoals ruimtelijke vaardigheden, verbale vaardigheden, mathematische vaardigheden en perceptuele vaardigheden (Dworak et al., 2021).

Huidig onderzoek

Deze masterproef zal zich toespitsen op de validering van de Childrens’ ICAR (Ch-ICAR); een nieuwe samenstelling van ICAR-items bedoeld om cognitieve vaardigheden bij kinderen tussen 11 en 14 jaar in kaart te brengen (Dutry et al., 2023). De ICAR60 werd reeds gevalideerd door onderzoekers bij verschillende leeftijdsgroepen (Condon & Revelle, 2014) en in verschillende talen (Atmaca & Baloglu, 2023; Kirkegaard & Nordbjerg, 2015; Nelissen et al., *In preparation*). Een versie voor kinderen jonger dan 14 jaar ontbreekt echter tot op heden (Kirkegaard & Bjerrekær, 2016; Revelle et al., 2020). Eerder werd aangehaald welke uitdagingen en restricties onderzoekers ondervinden bij het gebruik van bestaande instrumenten; zo vereisen deze vaak een zekere deskundigheid van de afnemer, zijn ze tijdrovend en kostenintensief van aard en worden ze auteursrechtelijk beschermd (Bisconer & Suttie, 1998; Dworak et al., 2021; Flanagan & Kaufman, 2004). In het licht van deze uitdagingen wordt duidelijk dat een behoefte bestaat aan een meetinstrument voor cognitieve vaardigheden bij kinderen dat specifiek is ontworpen om te voldoen aan de vereisten van wetenschappelijk onderzoek. De Ch-ICAR kan deze behoefte vervullen dankzij zijn kostenefficiënte en tijdbesparende karakter. Bijkomend voordeel betreft dat het instrument ook toegankelijk is voor onderzoekers met diverse achtergronden, zonder de noodzaak van gespecialiseerde bekwaamheden. De Ch-ICAR zou onderzoekers in staat stellen om op efficiënte wijze de cognitieve vaardigheden van kinderen te onderzoeken en begrijpen, terwijl tegelijkertijd de validiteit en betrouwbaarheid van de metingen wordt gewaarborgd.

De Ch-ICAR werd ontwikkeld door Dutry en collega's (2023) en bestaat uit 31 items verdeeld over vier subtesten: Verbaal Redeneren, Matrix Redeneren, Figuur Analogieën, en Cijferreeksen. Om de kwaliteit van de Ch-ICAR na te gaan zullen volgende factoren in kaart worden gebracht: 1) de betrouwbaarheid van de test en subtesten, 2) de verdeling van de Ch-ICAR scores, 3) de aanwezigheid van de g-factor, 4) de constructvaliditeit van de Ch-ICAR door de resultaten te vergelijken met de resultaten van de CoVaT, en 5) mogelijke geslachtsverschillen die van invloed kunnen zijn op de Ch-ICAR totaalscores. Wat betreft de betrouwbaarheid wordt een McDonalds omega coëfficiënt van minimaal .70 verwacht voor zowel de totale test als voor de subtesten afzonderlijk (Nunnally, 1978). De Ch-ICAR scores worden verwacht een normaalverdeling te volgen, zoals de verdeling van intelligentie zich presenteert in de populatie (Kaplan & Saccuzzo, 2005). De Ch-ICAR kan als kwaliteitsvol worden beschouwd indien een onderliggende g-factor aanwezig is, bijgevolg zou een positieve correlatie tussen alle subtesten waarneembaar moeten zijn (Cattell, 1987; Gottfredson, 1998; Spearman, 1904). Daarnaast gaan we de aanwezigheid van de g-factor ook formeel toetsen aan de hand van een confirmatorische factoranalyse waarbij we de vier Ch-ICAR subtesten laten laden op een hogere-orde latente factor: de g-factor (Carroll, 1993; McGrew, 2009). De constructvaliditeit wordt nagegaan door de resultaten op de Ch-ICAR (sub)testen te vergelijken met deze op de CoVaT (sub)testen. Algemeen wordt een sterk positieve correlatie verwacht tussen de Ch-ICAR totaalscore en de CoVaT totaalscore. Daarnaast verwachten we, ons baserend op de CHC-theorie en expertbeoordelingen, dat de Ch-ICAR subtest Verbaal Redeneren voornamelijk Gf en Gc meet en dus sterker zal samenhangen met de BCV's Gf en Gc zoals gemeten door de CoVaT subtesten Puntreeksen, Figuurreeksen, Schiftingen en Tegenstellingen (Magez et al., 2015; Tierens & Van Huynegem, 2017). Ondanks de naam Verbaal Redeneren doet vermoeden, meet deze subtest eerder wiskundige vaardigheden en algemene kennis (Nelissen et al., *In preparation*). Verder verwachten we dat de Ch-ICAR subtest Cijferreeksen vooral Gf meet en dus sterker zal samenhangen met de BCV Gf zoals gemeten door de CoVaT subtesten Puntreeksen en Figuurreeksen (Magez et al., 2015; Young & Keith, 2020). Met betrekking tot de subtesten Figuur Analogieën en Matrix Redeneren verwachten we dat deze vooral Gf en Gv meten en dus sterker zullen samenhangen met de BCV's Gf en Gv zoals gemeten door de CoVaT subtesten Puntreeksen, Figuurreeksen, Gedraaide Figuren en Dozen Plooien (Magez et al., 2015; Young & Keith, 2020). Tot slot verwachten we dat we geen geslachtsverschillen zullen zien in de totaalscore van de Ch-ICAR (Cole, 2009; Hyde, 2014; Lindberg et al., 2010).

Methode

Participanten

De steekproef omvatte 96 leerlingen, afkomstig uit vier verschillende klassen van het 6^e leerjaar. De deelnemende klassen waren gevestigd in Oost- en West-Vlaanderen. De steekproef bestond uit een evenredige verdeling van 48 meisjes (50%) en 48 jongens (50%). De gemiddelde leeftijd was 11.05 jaar ($SD = .27$). De thuistaal van de overgrote meerderheid (96,88%) van de leerlingen was Nederlands, met uitzondering van drie leerlingen die een andere thuistaal hadden (voor een leerling was dat Engels en voor de overige twee leerlingen Frans). Drie leerlingen gaven aan naast het Nederlands thuis ook Spaans, Albanees of Engels te spreken.

Materialen

Children's ICAR. De Ch-ICAR wordt online afgenomen en bestaat uit vier subtesten met een totaal van 31 items; zijnde Verbaal Redeneren, Cijferreeksen, Figuur Analogieën, en Matrix Redeneren. Bij aanvang van elke subtest worden geschreven instructies gegeven en worden de subtesten Figuur Analogieën en Matrix Redeneren voorzien van een voorbeelditem ter illustratie. In Tabel 1 wordt een overzicht geboden van de Ch-ICAR subtesten met telkens een voorbeeld van een op te lossen item. De subtest Verbaal Redeneren bestaat uit vragen die beroep doen op de woordenschat, algemene kennis en wiskundige vaardigheden van de participant. Verbaal Redeneren bestaat uit acht items met telkens acht antwoordopties, waaronder ook de antwoordopties '*Geen van deze*' en '*Ik weet het niet*'. Bij de subtest Cijferreeksen zijn er eveneens acht items op te lossen. Een item bestaat uit een reeks cijfers die aangevuld dient te worden, waarbij de reeksen kunnen variëren van vier tot zes getallen. Bij deze subtest worden geen antwoordopties aangeboden, in plaats daarvan kan de deelnemer zijn antwoord zelf typen in een leeg veld. Verder bestaat de subtest Figuur Analogieën uit zeven items. Een item bestaat uit drie gekleurde figuren en een vraagteken. Het doel is om de figuur aan te duiden dat op de plaats van het vraagteken dient te komen, zodat de verhouding tussen de eerste twee figuren dezelfde is als tussen de laatste twee figuren. De relatie tussen twee figuren kan betrekking hebben op kleur, de positie van het cirkeltje en de positie van het vierkantje. Deelnemers krijgen de keuze uit zes verschillende antwoordopties waarbij een figuur staat afgebeeld, alsook de opties '*Geen van deze*' en '*Ik weet het niet*'. Tot slot bestaat de subtest Matrix Redeneren uit acht items. Elk item stelt een matrix voor met figuren waarin één figuur ontbreekt. Deelnemers dienen uit zes verschillende figuren de juiste figuur te kiezen die de

matrix aanvult. Naast de zes figuren krijgen zij tevens de optie om 'Geen van deze' of 'Ik weet het niet' aan te duiden.

Tabel 1

Ch-ICAR Subtesten en Voorbeelditems

Subtest	Voorbeelditem							
Verbaal Redeneren	Indien het overmorgen twee dagen voor donderdag is, welke dag is het dan vandaag? (1) vrijdag, (2) maandag, (3) woensdag, (4) zaterdag, (5) dinsdag, (6) zondag, (7) geen van deze, (8) ik weet het niet							
Cijferreeksen	<table border="1" style="margin-left: auto; margin-right: auto;"> <tr> <td style="padding: 5px;">34</td> <td style="padding: 5px;">150</td> <td style="padding: 5px;">35</td> <td style="padding: 5px;">160</td> <td style="padding: 5px;">36</td> <td style="padding: 5px;">170</td> <td style="padding: 5px;">?</td> </tr> </table>	34	150	35	160	36	170	?
34	150	35	160	36	170	?		
Figuur Analogieën								
Matrix Redeneren								

CoVaT-CHC Basisversie. In dit onderzoek wordt gebruik gemaakt van de pen-en-papier versie van de CoVaT, dewelke wordt ondersteund door een diapresentatie. Omwille van tijdsbeperkingen wordt voor deze masterproef een selectie van CoVaT subtesten gebruikt. De selectie bestaat uit CoVaT subtesten die de BCV's meten die de Ch-ICAR wordt verwacht te meten, namelijk Gf, Gc, en Gv. Tabel 2 geeft een overzicht van de gebruikte CoVaT subtesten, alsook welke BCV's zij meten, het aantal items, de tijdslimiet en de mate van betrouwbaarheid per BCV. De subtest Puntreeksen kent een totaal van vijftien items met telkens twee te geven antwoorden (Magez & Tierens, 2015). Om een item op te lossen dienen punten te worden getekend die de voorafgaande reeks aanvullen. Elk juist antwoord levert 1 punt op. Verder bestaat de subtest Figuurreeksen uit 25 items (Magez & Tierens, 2015). Een item bestaat uit een reeks figuren die dient worden aangevuld volgens een bepaald patroon. Per item moeten drie figuren worden aangevuld, elk juist getekend antwoord levert 1 punt op. De subtesten Schiftingen en Tegenstellingen bestaan beide uit 35 items (Magez & Tierens, 2015). Bij de subtest Schiftingen is het de bedoeling om uit een reeks woorden het woord te kiezen dat niet in het rijtje past. Bij Tegenstellingen wordt een woord gegeven en dient de deelnemer uit een reeks woorden het woord te kiezen dat het tegengestelde is van het woord uit de opgave. Elk juist antwoord levert 1 punt op. Verder bestaat de subtest Gedraaide Figuren uit 20 items (Magez, 2015). Het is de bedoeling om de figuren, die geroteerd kunnen zijn, aan te duiden die hetzelfde zijn als de voorbeeldfiguur. Elk item omvat vijf antwoordalternatieven waarvan meerdere antwoorden juist kunnen zijn. Elke juist aangeduide figuur levert 1 punt op terwijl een fout antwoord een minpunt oplevert. Tot slot omvat de subtest Dozen Plooien 26 items (Magez & Tierens, 2015). Bij deze subtest wordt telkens een doos getoond met daarnaast vijf antwoordalternatieven; de antwoordalternatieven zijn figuren van een open geplooid doos. De deelnemer dient de open geplooid doos aan te duiden die na plooien overeenkomt met de voorbeelddoos. Elk juist antwoord levert 2 punten op.

Tabel 2*Overzicht BCV's gemeten door CoVaT Subtesten*

BCV	Subtest	Aantal items	Tijdslimiet (in minuten)	Betrouwbaarheid per BCV (Lambda 2)
Gf	Puntreeksen	15	5	.96
	Figuurreeksen	25	20	
Gc	Schiftingen	35	7	.87
	Tegenstellingen	35	7	
Gv	Gedraaide Figuren	20	5	.91
	Dozen Plooien	26	15	

Opzet en procedure

Deze masterproef kadert binnen lopend doctoraatsonderzoek van Merel Dutry. Het onderzoek werd uitgevoerd volgens het algemeen ethisch protocol van de Faculteit Psychologie en Pedagogische Wetenschappen van de Universiteit Gent (2022) en volgens de ethische protocollen van de American Psychological Association (2017). Het dossier (referentie 2021/59) getiteld 'Studie-oriëntering in het secundair onderwijs' werd op 2 september 2021 goedgekeurd door de Ethische Commissie van de Faculteit Psychologie en Pedagogische Wetenschappen van de Universiteit Gent.

Om participanten voor dit onderzoek te verwerven werden 50 scholen van het regulier basisonderwijs in Oost- en West-Vlaanderen per mail gecontacteerd met de vraag of de klassen van het 6^e leerjaar wenselijk deel te nemen aan het onderzoek. Van deze 50 scholen reageerden drie scholen dat zij bereid waren om deel te nemen, wat resulteerde in een totaal van vier klassen van het 6^e leerjaar die zich beschikbaar stelden. Alvorens het onderzoek van start kon gaan, werd een *informed consent* meegegeven met alle leerlingen. In dit toestemmingsformulier werd het doel van het onderzoek verduidelijkt en werd toestemming gevraagd aan de ouders of voogd om relevante gegevens van hun zoon of dochter te verwerken en te gebruiken voor onderzoeksdoeleinden. Zonder actieve toestemming van de ouders of voogd konden leerlingen

niet deelnemen aan het onderzoek. Over alle klassen heen waren er zes leerlingen zonder toestemming, zij werden niet verder betrokken in deze masterproef. Van de 99 leerlingen die wel mochten deelnemen waren drie leerlingen ziek op de dag van de testafname, wat resulteerde in een totale steekproef van 96 leerlingen. Alle verzamelde data werden gepseudonimiseerd in het kader van de privacywetgeving.

Bij elke klas werden twee testafnames uitgevoerd: een met de Ch-ICAR en een met de CoVaT. Tussen twee testafnames bij dezelfde klas zat een onderbreking van minimaal een week en maximaal twee weken. Om volgorde-effecten te minimaliseren werd afgewisseld welke test als eerste werd afgenomen. De testen gingen door in groepsverband waarbij de leerlingen op hun gebruikelijke plaats in de klas zaten. Om een eerlijk verloop te waarborgen, werden mappen tussen de leerlingen geplaatst zodoende zij niet konden spieken. Tijdens de testafnames was de onderzoeker in de klas aanwezig, alsook in de meeste gevallen de leerkracht. De Ch-ICAR afname verliep online waarbij leerlingen gebruik maakten van een iPad of computer die voorzien was door de school. Zij kregen een weblink die zij dienden in te voeren in hun internetbrowser, deze link voerde hen naar de test. De Ch-ICAR werd geprogrammeerd in Qualtrics (<https://www.qualtrics.com>). Er werd een volledig lesuur (50 minuten) voorzien voor de afname van de Ch-ICAR, echter was er geen tijdslimiet bij het afleggen van de test. Een volledig lesuur bleek ruim voldoende voor alle deelnemers. Na een algemene mondelinge toelichting door de onderzoeker konden de leerlingen op eigen tempo de Ch-ICAR invullen. Teruggaan in de test om antwoorden aan te passen was geen mogelijkheid, vragen openlaten was eveneens niet mogelijk. Indien deelnemers het antwoord niet wisten hadden ze de optie om 'ik weet het niet' aan te duiden. Voor de CoVaT afname werden drie lesuren voorzien. De afname verliep via pen en papier waarbij de subtesten werden uitgelegd aan de hand van een diapresentatie die werd geprojecteerd in de klas. Voorafgaand aan elke subtest dienden de leerlingen klassikaal de voorbeelditems op te lossen, waarna zij individueel enkele oefenitems konden oplossen. Leerlingen kregen bij het uitleggen van de voorbeelditems de mogelijkheid om vragen te stellen indien iets niet duidelijk was. Bij het oplossen van de testitems werd echter geen hulp meer geboden. Eenmaal elke leerling begreep wat het doel van de subtest was en geen vragen meer had, ging de subtest van start. De subtesten kenden een tijdslimiet zoals vastgelegd in de handleiding. Eenmaal de tijdslimiet bereikt was, dienden leerlingen hun balpen neer te leggen en hun bladzijde om te draaien.

Om de data-analyse van deze masterproef uit te voeren, werd gebruik gemaakt van twee statistische programma's: RStudio (RStudio Team, 2020) en SPSS (IBM Corp., 2022). De

beslissing om beide programma's te gebruiken kwam voort uit hun complementariteit in het uitvoeren van statistische analyses.

Resultaten

Betrouwbaarheid

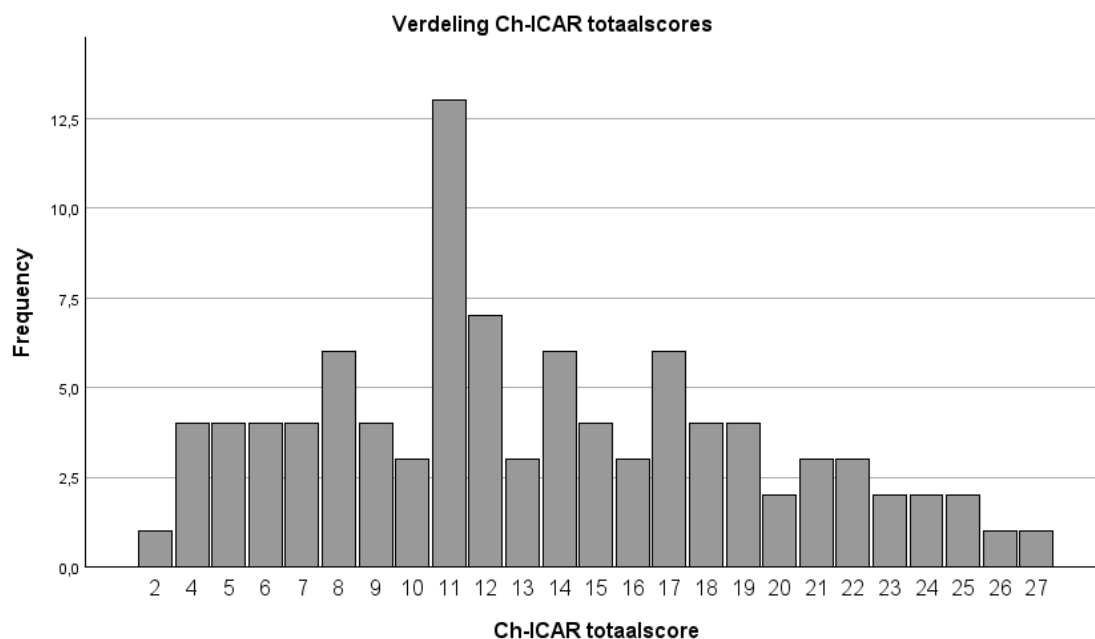
Om een antwoord te bieden op de eerste onderzoeksvraag werd de betrouwbaarheid van de Ch-ICAR en zijn subtesten nagegaan. De betrouwbaarheid werd in kaart gebracht door middel van de McDonalds omega coëfficiënt (McDonald, 1999), dewelke werd berekend in R met behulp van het R-package MBESS (Kelly, 2022). De totale Ch-ICAR test toont een goede mate van betrouwbaarheid ($\omega = .85$, 95% BI [.80, .89]). Verder vertonen twee subtesten eveneens een goede betrouwbaarheid, namelijk Cijferreeksen ($\omega = .79$, 95% BI [.71, .84]) en Figuur Analogieën ($\omega = .77$, 95% BI [.70, .82]). De subtest Verbaal Redeneren vertoont eerder een twijfelachtige betrouwbaarheid ($\omega = .60$, 95% BI [.22, .71]) en de subtest Matrix Redeneren ($\omega = .54$, 95% BI [.20, .68]) blijkt onvoldoende betrouwbaar.

Verdeling Ch-ICAR totaalscores

Teneinde de verdeling van de totaalscore na te gaan en een antwoord te bieden op de tweede onderzoeksvraag, werd een histogram van de geobserveerde Ch-ICAR totaalscores in SPSS gegenereerd (IBM Corp., 2022). Uit het histogram, afgebeeld in Figuur 2, blijkt een afwijking van de symmetrische normaalverdeling. De curve toont enerzijds geen graduele stijging en daling, anderzijds blijkt het zwaartepunt zich aan de linkerkant te bevinden.

Figuur 2

Verdeling Ch-ICAR Totaalscores



Interne structuur

Om inzicht te verwerven in de interne structuur van de Ch-ICAR werden Pearson correlaties berekend in SPSS (IBM Corp., 2022). Een overzicht van deze correlaties is terug te vinden in Tabel 3. De onderlinge correlaties tussen de subtesten van de Ch-ICAR vertonen een range van .35 tot .82 en blijken allen significant positief. Verder tonen de subtesten sterke positieve correlaties tot de Ch-ICAR totaalscores met correlaties variërend van .71 tot .82. Om de interne structuur verder te evalueren werd nagegaan of de Ch-ICAR de structuur van het CHC-model volgt. Een eerste model werd opgesteld waarbij de Ch-ICAR items de indicatoren van de latente cognitieve vaardigheden representeren (Verbaal Redeneren, Cijferreeksen, Figuur Analogieën en Matrix Redeneren) en waarbij g fungeert als hogere orde factor. Daarnaast werd een tweede model opgesteld; een één-factor model waarbij alle items op g laden en de subtesten achterwege werden gelaten. Voor beide modellen werd een confirmatorische factoranalyse (CFA) uitgevoerd met behulp van het R-package lavaan en de schattingsmethode Maximum Likelihood (ML; Rosseel, 2012). Om de fitmaten van de CFA te interpreteren werd gebruik gemaakt van de vuistregels uit het boek van Kline (2015). Uit de CFA van het hogere orde model blijkt dat het model een vrij goede fit vertoont ($\chi^2(430) = 471.24, p = .08; RMSEA = .03, 90\%$

BI [.00, .05]; CFI = .95; TLI = .94; SRMR = .15). Het hogere orde model blijkt de data significant beter te verklaren dan het één-factor model, wat consistent is met de CHC-theorie. Uit de CFA van het één-factor model ($\chi^2(434) = 521.22, p = .003$; RMSEA = .05, 90% BI [.03, .06]; CFI = .89; TLI = .88; SRMR = .17) blijkt dat het model een minder goede fit vertoont. De twee modellen werden tot slot formeel getoetst aan de hand van een modelvergelijkingstoets, ANOVA, in R. Het hogere orde model blijkt een significant betere fit te vertonen dan het één-factor model ($\Delta\chi^2(4) = 35.71, p < .001$).

Validiteit

Wat betreft constructvaliditeit van de Ch-ICAR werden de correlaties tussen de Ch-ICAR subtesten en de BCV's, zoals gemeten door de CoVaT subtesten, in kaart gebracht. Een overzicht van deze correlaties is terug te vinden in Tabel 4. Om de totaalscore van de CoVaT te vergelijken met deze van de Ch-ICAR werd de gewogen totaalscore berekend ($2 \cdot Gf + 2 \cdot Gc + Gv$). Alle correlaties blijken significant te zijn. Tussen de totaalscore van de Ch-ICAR en de gewogen totaalscore van de CoVaT wordt een matige positieve correlatie geconstateerd. Verder wordt tussen de Ch-ICAR subtest Verbaal Redeneren en de BCV's Gf en Gc matige positieve correlaties gemeten, waarbij Verbaal Redeneren het sterkst correleert met Gf. Tussen Cijferreeksen en de BCV's Gf en Gc worden eveneens matige positieve correlaties waargenomen, ook deze subtest correleert het sterkst met Gf. Daarnaast worden tussen de subtest Figuur Analogieën en de BCV's Gf, Gc en Gv gelijkaardige, zwakke correlaties waargenomen. Tot slot vertoont Matrix Redeneren gelijkaardige, matige correlaties met de BCV's Gf, Gc en Gv.

Tabel 3

Correlaties Ch-ICAR

(Sub)test	1	2	3	4	5
1. Verbaal Redeneren	1				
2. Cijferreeksen	.64**	1			
3. Figuur Analogieën	.36**	.43**	1		
4. Matrix Redeneren	.35**	.38**	.45**	1	
5. Ch-ICAR totaal	.77**	.82**	.74**	.71**	1

Noot. ** $p < .01$

Tabel 4

Correlaties tussen de Ch-ICAR (Sub)testen, de CoVaT Totaalscore en Drie BCV's Gemeten door de CoVaT

	Gf	Gc	Gv	CoVaT totaal
Ch-ICAR VR	.53**	.45**	.28**	.53**
Ch-ICAR CR	.50**	.44**	.34**	.53**
Ch-ICAR FA	.29**	.23*	.28**	.33**
Ch-ICAR MR	.37**	.38**	.31**	.43**
Ch-ICAR totaal	.56**	.49**	.40**	.60**

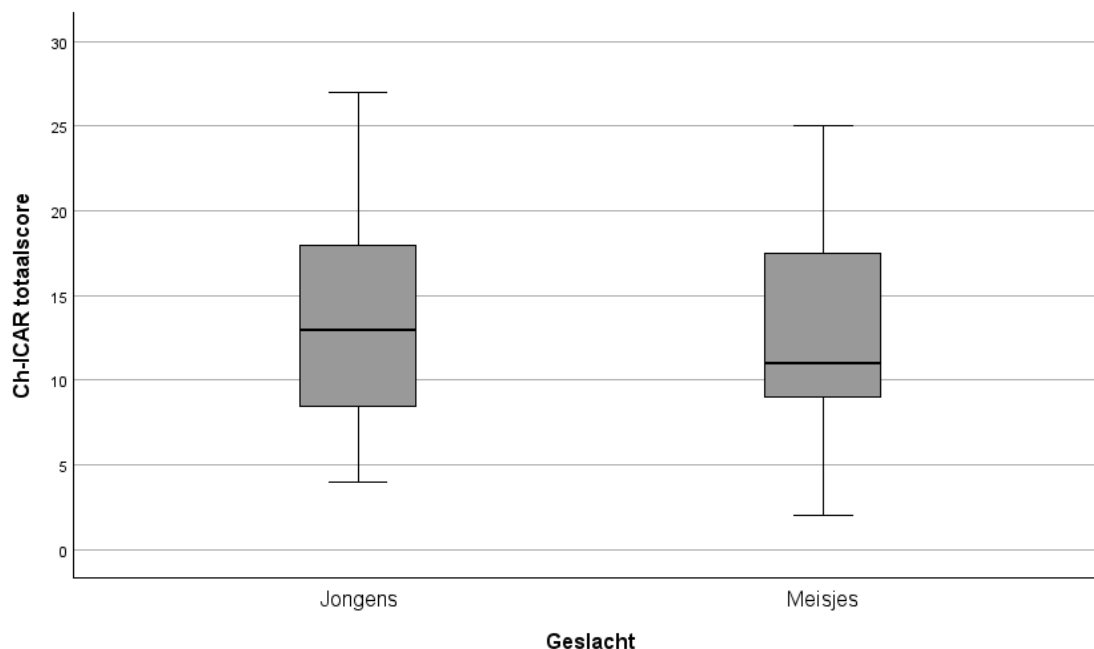
Noot. VR = Verbaal Redeneren, CR = Cijferreeksen, FA = Figuur Analogieën, MR = Matrix Redeneren, ** $p < .01$

Geslachtsverschillen

Tot slot werd, teneinde een antwoord te bieden op de vijfde onderzoeksvraag, onderzocht of geslachtsverschillen aanwezig zijn wat betreft de Ch-ICAR totaalscore. Een visuele voorstelling van de verdeling van de Ch-ICAR totaalscore volgens geslacht is terug te vinden in Figuur 3. Gezien de data leek af te wijken van de normaalverdeling werd gekozen om de Mann-Whitney toets uit te voeren. De toets werd uitgevoerd in SPSS waarbij geslacht fungeerde als *grouping variable* en de Ch-ICAR totaalscore als afhankelijke variabele. Wat betreft de Ch-ICAR totaalscore konden geen significante verschillen worden weerhouden tussen jongens ($M = 13.79$, $SD = 6.10$) en meisjes ($M = 12.83$, $SD = 5.79$) op basis van de Mann-Whitney toets ($U = 1042.5$, $z = -.80$, $p = .42$, Cohen's $d = .16$).

Figuur 3

Verdeling Ch-ICAR Totaalscore volgens Geslacht



Discussie

Inleiding tot discussie

De Ch-ICAR betreft een specifieke versie van het meetinstrument ICAR dat ontwikkeld werd voor het in kaart brengen van cognitieve vaardigheden bij kinderen tussen 11 en 14 jaar (Dutry et al., 2023). In het huidige landschap van onderzoek naar cognitieve vaardigheden bij kinderen bestaan reeds meerdere meetinstrumenten, echter vertonen deze instrumenten beperkingen wanneer ze worden ingezet voor wetenschappelijk onderzoek. Zo brengen zij vaak aanzienlijke kosten met zich mee, zijn ze tijdsintensief om af te nemen en te verwerken, en vereisen zij dat onderzoekers specifieke deskundigheid ontwikkelen om de testafnames correct te laten verlopen (Bisconer & Suttie, 1998; Flanagan & Kaufman, 2004; Nelissen et al., *In preparation*). Deze nadelige aspecten vormen een barrière voor grootschalige onderzoeksprojecten (Revelle et al., 2020). Deze masterproef heeft als doel deze noden op te vullen door bij te dragen aan de validering van de Ch-ICAR zodoende dit instrument in de toekomst kan worden ingezet voor onderzoeksdoeleinden bij kinderen. Om de kwaliteit van de Ch-ICAR te onderzoeken, werd de betrouwbaarheid van de (sub)testen nagegaan, de verdeling

van de totaalscores in kaart gebracht, de interne structuur geëvalueerd, de constructvaliditeit onderzocht en werd onderzocht of geslachtsverschillen aanwezig zijn inzake de totaalscores.

Bespreking resultaten en vergelijking literatuur

Wat betreft de betrouwbaarheid van de Ch-ICAR in zijn geheel constateren we een goede mate van betrouwbaarheid, hetgeen in lijn ligt met andere onderzoeken (Dutry et al., 2023; Young & Keith, 2020). Op subtestniveau vertonen de subtesten Cijferreeksen en Figuur Analogieën een goede betrouwbaarheid terwijl Verbaal Redeneren eerder een twijfelachtige mate van betrouwbaarheid vertoont en Matrix Redeneren onvoldoende betrouwbaar blijkt. In het onderzoek van Nelissen en collega's (*In preparation*) bleek de subtest Verbaal Redeneren eveneens een twijfelachtige betrouwbaarheid te vertonen. Volgens de onderzoekers zou dit mogelijk verklaard kunnen worden doordat enkele items van deze subtest eerder vaardigheden met betrekking tot wiskundige en algemene kennis meten, in plaats van vaardigheden die gerelateerd zijn aan verbaal redeneren. Om een betere betrouwbaarheid te bekomen voor deze subtest zouden de slechts ladende items kunnen vervangen worden door nieuwe items toe te voegen die meer beroep doen op verbaal redeneren dan op wiskundige en algemene kennis. De lage betrouwbaarheid van de subtest Matrix Redeneren ligt in lijn met bevindingen van Condon en Revelle (2014). De lage betrouwbaarheid van Verbaal Redeneren en Matrix Redeneren suggereert dat deze subtesten niet eenduidig één latent construct meten; het is aangewezen verder onderzoek te voeren naar de items en te achterhalen op welke latente factor(en) zij laden. Bovenstaande bevindingen inzake de betrouwbaarheid van de Ch-ICAR impliceren dat, over het geheel genomen, het instrument geschikt is voor gebruik in zijn totaliteit. Echter dient bij het overwegen van het gebruik van afzonderlijke subtesten de nodige voorzichtigheid in acht te worden genomen om de betrouwbaarheid van de test te bewaken.

Verder werd bij de analyse van de verdeling van de Ch-ICAR totaalscore een afwijking van de normaalverdeling geobserveerd waarbij het zwaartepunt van de curve zich aan de linkerkant bevindt. Deze afwijking indiceert dat de verdeling van de steekproef niet symmetrisch is rond het midden; dergelijke asymmetrie impliceert dat de moeilijkheidsgraad van de test te hoog lag. Mogelijk werd deze test als te moeilijk ervaren door de steekproef. Het zou kunnen dat de steekproef lager scoort op het gebied van academische prestaties en bijgevolg lager op de Ch-ICAR testvragen. Uit eerder onderzoek is al gebleken dat cognitief vermogen sterk correleert met de academische prestaties van een individu (Deary et al., 2007; Lu et al., 2011; Petrill et al., 2000; Roth et al., 2015; Spinath et al., 2006). Een mogelijke verklaring voor deze afwijking kan

de beperkte steekproefomvang zijn, waarbij de steekproef niet voldoende representatief is voor de populatie. Verder onderzoek met grotere steekproeven blijkt aangewezen, alsook dienen de achterliggende factoren van deze afwijking van de normaalverdeling onderzocht te worden.

Teneinde indicaties te vinden voor de aanwezigheid van de g-factor werden de onderlinge correlaties van de Ch-ICAR (sub)testen in kaart gebracht. Zoals verwacht constateren we onderling tussen alle Ch-ICAR subtesten en de totale testscore positief significante correlaties. Dit wijst op een samenhang van de cognitieve vaardigheden gemeten door de verschillende Ch-ICAR subtesten, waarbij individuen die goed presteren op één subtest de neiging hebben om goed te presteren op andere subtesten. Dit fenomeen ondersteunt het idee van een overkoepelende cognitieve factor, de g-factor (Carroll, 1993; Gottfredson, 1998; McGrew, 2009; Spearman, 1904). Bovendien dragen deze bevindingen bij aan de betrouwbaarheid van de Ch-ICAR gezien ze impliceren dat een hoge mate van consistentie bestaat tussen de verschillende cognitieve domeinen die worden gemeten. Deze positieve samenhang tussen de subtesten suggereert dat de Ch-ICAR mogelijk effectief kan worden ingezet als instrument voor het in kaart brengen van cognitieve vaardigheden door middel van één test.

De uitgevoerde CFA bood inzicht in de structuur van de Ch-ICAR. De geobserveerde indicatoren, vertegenwoordigd door de individuele testitems, bleken te laden op de latente factoren van de eerste orde die op hun beurt allen laden op een hogere-orde latente factor. Deze resultaten ondersteunen het idee dat een overkoepelende cognitieve factor bestaat die inherent is aan alle geteste vaardigheden. Bovendien lijken deze bevindingen de aanwezigheid van een hiërarchische structuur te bevestigen, analoog aan de structuur van het CHC-model, waarbij een overkoepelende factor de gemeenschappelijke variabiliteit tussen subtesten kan verklaren. De uitgevoerde modelvergelijkingstoets ondersteunt deze stelling; het hogere orde model blijkt significant een betere fit te vertonen dan het één-factormodel.

Om de constructvaliditeit van de Ch-ICAR na te gaan, werden de Ch-ICAR (sub)testcores in verband gebracht met de scores op de BCV's gemeten door de CoVaT subtesten. De Ch-ICAR totaalscore en CoVaT totaalscore tonen naar verwachting een significante positieve correlatie, wat aantoont dat deelnemers gelijkaardig scoren op beide testen. Deze bevinding ligt in lijn met ander onderzoek waarbij de Ch-ICAR werd vergeleken met de WAIS-IV (Young & Keith, 2020). Hieruit kunnen we afleiden dat individuen die op de ene test goed scoren ook op de andere test goed scoren. Zoals verwacht vertonen alle subtesten van de Ch-ICAR en de BCV's gemeten door de CoVaT subtesten onderling positieve significante

correlaties, hetgeen opnieuw indiceert dat een overkoepelende factor aanwezig is (Carroll, 1993; Gottfredson, 1998; McGrew, 2009; Spearman, 1904). Bovendien impliceren deze resultaten, dankzij de vergelijking met een ‘gouden standaard’ batterij zoals de CoVaT, dat de Ch-ICAR effectief meet wat hij beoogt te meten.

De Ch-ICAR subtest Verbaal Redeneren toont naar verwachting de sterkste correlaties met de BCV's Gf en Gc, dit in tegenstelling tot een onderzoek van Young en Keith (2020) waarbij Verbaal Redeneren voornamelijk leek te laden op Gv. Deze discrepantie kan mogelijk verklaard worden door de verschillende meetinstrumenten die in de studies zijn gebruikt. In tegenstelling tot Young en Keith (2020), die de WAIS-IV hanteerden voor hun vergelijking, werd in dit onderzoek de CoVaT als referentiepunt genomen. Desondanks deze discrepantie mogelijk van invloed kan zijn bleken ook Young en Keith (2020) verrast door hun resultaten. Zij stelden dat Gf en Gv weliswaar sterk samenhangend van aard zijn, hoewel Gv meer beroep doet op visuele stimuli. Bijgevolg vinden zij de samenhang tussen Verbaal Redeneren en Gv zeer verrassend gezien het visuele aspect compleet afwezig is bij Verbaal Redeneren (Young & Keith, 2020). Zij wierpen de hypothese op dat de WAIS-indicatoren die Gv meten mogelijk meerdere vaardigheden meten dan louter Gv (Young & Keith, 2020). Andere verklaringen zouden zijn dat dergelijke resultaten het gevolg zijn van een meetfout of een te beperkte steekproef (Young & Keith, 2020). De onderzoekers benadrukken het belang van replicaties van hun onderzoek met grotere steekproeven. Verder werd, naar verwachting, vastgesteld dat de subtest Cijferreeksen de sterkste correlatie vertoont met de BCV Gf. Opmerkelijk genoeg is Cijferreeksen echter ook sterk gecorreleerd met Gc. Ondanks de verwachting dat Cijferreeksen primair een beroep zou doen op wiskundige kennis, impliceert deze observatie dat andere elementen ook een rol spelen in het oplossen van deze subtest. Gezien Gc wordt gedefinieerd als de verworven kennis van een persoon en hoe deze persoon zijn kennis toepast in het dagelijkse leven, maken participanten wellicht ook gebruik van hun algemene kennis om de items van Cijferreeksen op te lossen (Flanagan & Harrison, 2012; Magez et al., 2015; McGrew, 2009; Schneider & McGrew, 2012). Deze bevinding suggereert dat de cognitieve processen bij het oplossen van deze subtest breder zijn dan de processen die louter beroep doen op pure wiskundige kennis. Daarnaast blijkt de Ch-ICAR subtest Figuur Analogieën gelijkmatig te correleren over de drie BCV's heen, hoewel de correlaties eerder zwak zijn, waarbij de sterkste correlaties worden waargenomen met Gf en Gv. Deze bevindingen zijn conform de vooropgestelde hypothese. De gelijkaardige correlatie met Gc zou enerzijds te wijten kunnen zijn aan het feit dat deze subtest ook beroep doet op algemene kennis, anderzijds zou dit opnieuw

te verklaren vallen omwille van de overkoepelende g-factor die een onderlinge positieve correlatie tussen alle cognitieve vaardigheden suggereert (Flanagan & Harrison, 2012; Magez et al., 2015; McGrew, 2009; Schneider & McGrew, 2012; Spearman, 1904). Het feit dat de correlaties tussen Figuur Analogieën en de BCV's eerder zwak zijn, kan duiden op een beperkte validiteit van deze subtest. De Ch-ICAR subtest Matrix Redeneren werd verondersteld het sterkst te correleren met Gf en Gv, echter toont deze subtest de sterkste correlatie met Gc. Deze opmerkelijke bevinding kan mogelijk te wijten zijn aan de lage betrouwbaarheid van deze subtest. Het is van belang om verder onderzoek te voeren naar deze subtest en zijn items, zodoende te achterhalen welk latent construct hierbij wordt gemeten.

Naar verwachting toonde de analyse van de totaalscores door middel van de Mann-Whitney toets geen significante geslachtsverschillen. Dit ligt in lijn met de meest recente literatuur waarbij wordt gesuggereerd dat de geslachtsverschillen die mogelijk worden waargenomen in onderzoek naar cognitieve capaciteiten meestal kleine effecten betreffen of eerder te wijten zijn aan andere factoren zoals persoonskenmerken of schoolse achtergrond (Daseking et al., 2017; Hyde, 2014). De afwezigheid van significante geslachtsverschillen met betrekking tot de totaalscores van de Ch-ICAR suggereert dat de test gelijkwaardige, betrouwbare metingen neemt voor zowel jongens als meisjes. Uit de resultaten van huidig onderzoek kunnen we concluderen dat op theoretisch niveau deze resultaten bijdragen aan het begrip van het theoretisch kader rond de Ch-ICAR, gezien de literatuur hieromtrent tot op heden eerder beperkt is. Op praktisch niveau suggereren de resultaten van dit onderzoek dat de Ch-ICAR een kwaliteitsvol en kostenefficiënt alternatief kan bieden voor grootschalig onderzoek naar cognitieve vaardigheden bij kinderen.

Beperkingen en toekomstig onderzoek

Desondanks dit onderzoek waardevolle inzichten biedt, kent het ook enkele beperkingen. Een eerste beperking van dit onderzoek betreft de grootte van de steekproef. De steekproef bestond uit 96 leerlingen, wat eerder als een beperkt aantal kan worden beschouwd in het kader van wetenschappelijk onderzoek. Tevens bleek uit de analyse van de verdeling van de totaalscore van de Ch-ICAR dat deze afwijkt van de normaalverdeling. Het dient verder te worden onderzocht of deze afwijking een gevolg is van de beperkte steekproefomvang, of in verband staat met andere factoren. Het zou bijvoorbeeld kunnen dat de items als te moeilijk werden ervaren, wat resulteert in lagere scores. Verder bestond de steekproef uit leerlingen van maximum 12 jaar met een gemiddelde leeftijd van 11.05 jaar. Dit is een belangrijke beperking

gezien de Ch-ICAR bedoeld is voor kinderen van 11 tot 14 jaar, en dus vanuit dit onderzoek geen data voorhanden is van leerlingen tussen 12 en 14 jaar. Replicaties van dit onderzoek met steekproeven binnen de leeftijdscategorie van 12 tot 14 jaar zijn sterk aangewezen. Bovendien waren alle deelnemers afkomstig uit de regio's Oost- en West-Vlaanderen, wat een geografische beperking oplevert. Een meer diverse steekproef, afkomstig uit alle regio's van Vlaanderen, zou een completer perspectief kunnen bieden. Daarnaast werd in dit onderzoek noch de sociaaleconomische status, noch de academische prestaties van de deelnemers bevestigd. In het kader van toekomstig onderzoek is het aangeraden om een grotere en meer diverse steekproef te overwegen zodoende een representatievere mix te bekomen om de generaliseerbaarheid van de resultaten te verbeteren.

Wat betreft de subtesten worden eveneens beperkingen opgemerkt. De subtest Figuur Analogieën vertoonde eerder zwakke correlaties met de BCV's, hetgeen kan duiden op een beperkte validiteit. Toekomstig onderzoek dient de validiteit van deze subtest verder na te gaan. Daarnaast bleek de betrouwbaarheid van de subtesten Verbaal Redeneren en Matrix Redeneren verder onderzoek te vereisen. Hoewel de verklaring voor de lage betrouwbaarheid van deze twee subtesten geen deel uitmaakt van huidig onderzoek, zou vervolgonderzoek zich kunnen focussen op een itemanalyse van deze subtesten en de slechtst ladende items vervangen door nieuwe items. Bovendien zou een verbeterde betrouwbaarheid van deze subtesten ook de algemene kwaliteit van de Ch-ICAR verbeteren, hetgeen in de toekomst eventueel zou kunnen toelaten om de subtesten afzonderlijk te gebruiken in plaats van het instrument in zijn geheel.

Conclusie

Samengevat kunnen we stellen dat dit onderzoek belangrijke inzichten biedt in de validiteit van de Ch-ICAR. De resultaten suggereren dat de Ch-ICAR een waardevol alternatief meetinstrument kan zijn voor onderzoek naar cognitieve vaardigheden bij kinderen tussen 11 en 14 jaar. Over het algemeen toont de Ch-ICAR een goede betrouwbaarheid, duiden de correlaties binnen de interne structuur op een positieve samenhang en wordt een significant positieve correlatie waargenomen tussen de totaalscore op de Ch-ICAR en de CoVaT. Deze resultaten suggereren de aanwezigheid van een overkoepelende g-factor over de verschillende cognitieve domeinen. Bovendien werd geen geslachtsverschil geconstateerd op totaalniveau van de Ch-ICAR, wat indiceert dat de test geen genderbias vertoont in zijn metingen. Op subtestniveau werden enkele beperkingen inzake de betrouwbaarheid waargenomen, meer specifiek bij de subtesten Verbaal Redeneren en Matrix Redeneren. Een diepgaandere analyse van de

subtestitems lijkt aangewezen om meer inzicht te krijgen in de onderliggende constructen die worden gemeten. Verder onderzoek op subtestniveau zou de kwaliteit ten goede kunnen komen zodoende de subtesten ook afzonderlijk zouden kunnen worden afgenomen. Concluderend kunnen we stellen dat, in het kader van de eerder besproken behoeften binnen het onderzoeksveld, de Ch-ICAR zich presenteert als een potentieel waardevol instrument. Het heeft de capaciteit om met beperkte financiële middelen en op efficiënte wijze te worden ingezet om de cognitieve vaardigheden van grote groepen kinderen tegelijk te meten.

Referenties

- Ardila, A., Rosselli, M., Matute, E., & Inozemtseva, O. (2011). Gender differences in cognitive development. *Developmental Psychology*, *47*(4), 984–990.
<https://doi.org/10.1037/a0023819>
- Atmaca, F., & Baloğlu, M. (2023). Are we as intelligent as we think? Adaptation study of ICAR-16 into Turkish. *Personality and Individual Differences*, *213*.
<https://doi.org/10.1016/j.paid.2023.112284>
- Benson, N., Hulac, D. M., & Kranzler, J. H. (2010). Independent examination of the Wechsler Adult Intelligence Scale-Fourth Edition (WAIS-IV): What does the WAIS-IV measure? *Psychological Assessment*, *22*(1), 121–130. <https://doi.org/10.1037/a0017767>
- Bisconer, S. W., & Suttie, J. N. (1998). *Mental retardation*. Pergamon Press.
- Boake, C. (2002). From the Binet-Simon to the Wechsler-Bellevue: Tracing the history of intelligence testing. *Journal of Clinical and Experimental Neuropsychology*, *24*(3), 383–405.
- Brody, N. (1999). *What is intelligence?*
- Caniyez, G. L., & Watkins, M. W. (1998). Long-term stability of the Wechsler Intelligence Scale for Children-Third Edition. *Psychological Assessment*. <http://www.ucl.eiu.edu/~cfglc>
- Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. Cambridge University Press.
- Cattell, R. B. (1941). Some theoretical issues in adult intelligence testing. *Psychological Bulletin*, *38*(9), 592–616.
- Cattell, R. B. (1987). *Intelligence: Its structure, growth and action*. Elsevier Science Publishers B. V.
- Cole, E. R. (2009). Intersectionality and research in psychology. *American Psychologist*, *64*(3), 170–180. <https://doi.org/10.1037/a0014564>

- Condon, D. M., & Revelle, W. (2014). The International Cognitive Ability Resource: Development and initial validation of a public-domain measure. *Intelligence*, 43(1), 52–64. <https://doi.org/10.1016/j.intell.2014.01.004>
- Condon, D., & Revelle, W. (2012). *The International Cognitive Ability Resource: Development and initial validation of a public-domain measure*.
- Daseking, M., Petermann, F., & Waldmann, H. C. (2017). Sex differences in cognitive abilities: Analyses for the German WAIS-IV. *Personality and Individual Differences*, 114, 145–150. <https://doi.org/10.1016/j.paid.2017.04.003>
- Deary, I. J. (2014). The stability of intelligence from childhood to old age. *Current Directions in Psychological Science*, 23(4), 239–245. <https://doi.org/10.1177/0963721414536905>
- Deary, I. J., Cox, S. R., & Hill, W. D. (2022). Genetic variation, brain, and intelligence differences. *Molecular Psychiatry*, 27(1), 335–353. <https://doi.org/10.1038/s41380-021-01027-y>
- Deary, I. J., Strand, S., Smith, P., & Fernandes, C. (2007). Intelligence and educational achievement. *Intelligence*, 35(1), 13–21. <https://doi.org/10.1016/j.intell.2006.02.001>
- Doll, E. A. (1917). A brief Binet-Simon scale. *The Psychological Clinic*, 11(7), 197–211.
- Dutry, M., Vereeck, A., Schelfhout, S., Duyck, W., Janse, M., De Herdt, K., Schittekatte, M., Dirix, N., & Derous, E. (2023). Validation of the International Cognitive Ability Resource for children. *44th Annual Conference of the International School Psychology Association (ISPA 2023), Abstracts. Presented at the 44th Annual Conference of the International School Psychology Association (ISPA 2023)*.
- Dworak, E. M., Revelle, W., Doebler, P., & Condon, D. M. (2021). Using the International Cognitive Ability Resource as an open source tool to explore individual differences in cognitive ability. *Personality and Individual Differences*, 169. <https://doi.org/10.1016/j.paid.2020.109906>
- Fergusson, D. M., & Horwood, J. (1997). Gender differences in educational achievement in a New Zealand birth cohort. *New Zealand Journal of Educational Studies*.

- Flanagan, D. P., & Dixon, S. G. (2014). The Cattell-Horn-Carroll theory of cognitive abilities. *Encyclopedia of Special Education*. Wiley.
<https://doi.org/10.1002/9781118660584.ese0431>
- Flanagan, D. P., & Harrison, P. L. (2012). *Contemporary intellectual assessment: Theories, tests, and issues* (3rd ed.). Guilford Press.
- Flanagan, D. P., & Kaufman, A. S. (2004). *Essentials of WISC-V assessment*. John Wiley & Sons Inc.
- Flanagan, D. P., McGrew, K. S., & Ortiz, S. O. (2000). *The Wechsler Intelligence Scales and Gf-Gc theory: A contemporary approach to interpretation*. Allyn & Bacon.
- Flynn, J. R. (1987). Massive IQ gains in 14 nations: What IQ tests really measure. *Psychological Bulletin*.
- Gottfredson, L. S. (1998). The general intelligence factor. *Scientific American Presents*.
- Horn, J. L. (1965). *Fluid and crystallized intelligence: A factor analytic study of the structure among primary mental abilities*. University of Illinois.
- Hyde, J. S. (2014). Gender similarities and differences. *Annual Review of Psychology*, 65, 373–398. <https://doi.org/10.1146/annurev-psych-010213-115057>
- IBM Corp. (2022). IBM SPSS Statistics for Macintosh, Version 29.0. Armonk, NY: IBM Corp
- Jensen, A. R. (2014). Charles E. Spearman: The Discoverer of g. *Portraits of pioneers in psychology*.
- Kaplan, R. M., & Saccuzzo, D. P. (2005). *Psychological testing: Principles, applications, and issues* (6th ed.). Thomson Wadsworth.
- Keith, T. Z., & Reynolds, M. R. (2010). Cattell-Horn-Carroll abilities and cognitive tests: What we've learned from 20 years of research. *Psychology in the Schools*, 47(7), 635–650.
- Kelly, K. (2022) The MBESS R Package (4.9.2). R Studio.

- Kirkegaard, E. O. W., & Bjerrekær, J. D. (2016). ICAR5: design and validation of a 5-item public domain cognitive ability test. *Open Differential Psychology*.
<https://doi.org/10.26775/odp.2016.02.12>
- Kirkegaard, E. O. W., & Nordbjerg, O. (2015). Validating a Danish translation of the International Cognitive Ability Resource sample test and Cognitive Reflection Test in a student sample. *Open Differential Psychology*. <https://doi.org/10.26775/odp.2015.07.31>
- Kline, R. B. (2015). *Principles and Practice of Structural Equation Modeling* (4th ed.). New York Guilford Press.
- Li, Y., Liu, Y., Li, J., Qin, W., Li, K., Yu, C., & Jiang, T. (2009). Brain anatomical network and intelligence. *PLoS Computational Biology*, 5(5).
<https://doi.org/10.1371/journal.pcbi.1000395>
- Lindberg, S. M., Hyde, J. S., Petersen, J. L., & Linn, M. C. (2010). New Trends in gender and mathematics performance: A meta-analysis. *Psychological Bulletin*, 136(6), 1123–1135.
<https://doi.org/10.1037/a0021276>
- Lu, L., Weber, H. S., Spinath, F. M., & Shi, J. (2011). Predicting school achievement from cognitive and non-cognitive variables in a Chinese sample of elementary school children. *Intelligence*, 39(2–3), 130–140. <https://doi.org/10.1016/j.intell.2011.02.002>
- Maccoby, E. E., & Jacklin, C. N. (1974). *The psychology of sex differences*. Stanford Univ. Press.
- Magez, W. (2015). *Itemconstructie, redactie en analyse van de CoVaT-CHC Basisversie*.
- Magez, W. (2016). *CoVaT-CHC Basisversie: Enkele psychometrische toelichten en achtergronden bij de publicatie 2015*. <http://www.thomasmore.be/psychodiagnostisch-centrum/covat-chc-basisversie->
- Magez, W., De Cleen, W., Bos, A., Rauws, G., Geerinck, K., & De Kerf, L. (2015). *Intelligentiemeting in nieuwe banen: De integratie van het CHC-model in de psychodiagnostische praktijk*.

- Magez, W., & Tierens, M. (2015). Introductie tot het project CoVaT-CHC Basisversie: De chronologie. *Intelligence*. <https://doi.org/10.1016/j.intell.2008.08.004>
- McDonald R. P. (1999). *Test theory: A unified treatment*. Mahwah, NJ: Lawrence Erlbaum.
- McGrew, K. S. (1997). Analysis of the major intelligence batteries according to a proposed comprehensive Gf-Gc framework. *Contemporary intellectual assessment: Theories, tests, and issues*. The Guilford Press.
- McGrew, K. S. (2009). CHC theory and the human cognitive abilities project: Standing on the shoulders of the giants of psychometric intelligence research. *Intelligence*, 37(1), 1–10. <https://doi.org/10.1016/j.intell.2008.08.004>
- McGrew, K. S., & Flanagan, D. P. (1998). The Intelligence Test Desk Reference (ITDR): Gf-Gc cross-battery assessment. *The intelligence test desk reference (ITDR): Gf-Gc cross-battery assessment*. Allyn & Bacon.
- Nelissen, W., Schittekatte, M., Fontaine, J., & Rigolle, F. Validation of the International Cognitive Ability Resource (ICAR) in a Flemish sample [Ongepubliceerde doctoraatscriptie]. Universiteit Gent.
- Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). McGraw-Hill.
- PDC Thomas More Antwerpen, & CAPvzw. (2018). *CoVaT- CHC Basisversie*.
- Petrill, S. A., Wilkerson, B., & Petrill, S. (2000). Intelligence and achievement: A behavioral genetic perspective. *Educational Psychology Review*, 12(2).
- Pezzuti, L., & Orsini, A. (2016). Are there sex differences in the Wechsler Intelligence Scale for Children - Forth Edition. *Learning and Individual Differences*, 45, 307–312. <https://doi.org/10.1016/j.lindif.2015.12.024>
- Pol, H. E. H., Schnack, H. G., Posthuma, D., Mandl, R. C. W., Baaré, W. F., Van Oel, C., Van Haren, N. E., Collins, D. L., Evans, A. C., Amunts, K., Bürgel, U., Zilles, K., De Geus, E., Boomsma, D. I., & Kahn, R. S. (2006). Genetic contributions to human brain morphology and intelligence. *Journal of Neuroscience*, 26(40), 10235–10242. <https://doi.org/10.1523/JNEUROSCI.1312-06.2006>

- Posthuma, D., De Geus, E., Baaré, W., Hulshoff Pol, H., Kahn, R., & Boomsma, D. (2002). The association between brain volume and intelligence is of genetic origin. *Nature Neuroscience*, 5(2), 83–84.
- Revelle, W., Dworak, E. M., & Condon, D. (2020). Cognitive ability in everyday life: The utility of open-source measures. *Current Directions in Psychological Science*, 29(4), 358–363. <https://doi.org/10.1177/0963721420922178>
- Rossee, Y. (2012). lavaan: An R Package for Structural Equation Modeling. In *Journal of Statistical Software*. <http://www.jstatsoft.org/v48/i02/>
- Roth, B., Becker, N., Romeyke, S., Schäfer, S., Domnick, F., & Spinath, F. M. (2015). Intelligence and school grades: A meta-analysis. *Intelligence*, 53, 118–137. <https://doi.org/10.1016/j.intell.2015.09.002>
- RStudio Team (2020). RStudio: Integrated Development for R. RStudio, PBC, Boston, MA URL <http://www.rstudio.com/>.
- Ryan, J. J., Carruthers, C. A., Miller, L. J., Souheaver, G. T., Gontkovsky, S. T., & Zehr, M. D. (2003). Exploratory factor analysis of the Wechsler Abbreviated Scale of Intelligence (WASI) in adult standardization and clinical samples. *Applied Neuropsychology*, 10(4), 252–256. https://doi.org/10.1207/s15324826an1004_8
- Salthouse, T. A. (2010). Influence of age on practice effects in longitudinal neurocognitive change. *Neuropsychology*, 24(5), 563–572. <https://doi.org/10.1037/a0019026>
- Schneider, W. J., & McGrew, K. S. (2012). The Cattell-Horn-Carroll model of intelligence. *Contemporary intellectual assessment: Theories, tests, and issues*. The Guilford Press.
- Schwartzman, A. E., Gold, D., Andres, D., Arbuckle, T. Y., & Chaikelson, J. (1987). Stability of intelligence: A 40-year follow-up. *Canadian Journal of Psychology*, 41(2), 244–256. <https://doi.org/10.1037/h0084155>
- Spearman, C. (1904). “General intelligence”, objectively determined and measured. *American Journal of Psychology*, 15, 201–293.

- Spinath, B., Spinath, F. M., Harlaar, N., & Plomin, R. (2006). Predicting school achievement from general cognitive ability, self-perceived ability, and intrinsic value. *Intelligence*, 34(4), 363–374. <https://doi.org/10.1016/j.intell.2005.11.004>
- Sundet, J. M., Barlaug, D. G., & Torjussen, T. M. (2004). The end of the Flynn effect? A study of secular trends in mean intelligence test scores of Norwegian conscripts during half a century. *Intelligence*, 32(4), 349–362. <https://doi.org/10.1016/j.intell.2004.06.004>
- Terman, L. M. (1916). The measurement of intelligence. *The measurement of intelligence*. Houghton, Mifflin and Company. <https://doi.org/10.1037/10014-000>
- Thurstone, L. L. (1931). Multiple factor analysis. *Psychological Review*, 38(5), 406–427. <https://doi.org/10.1037/h0069792>
- Tierens, M., & Van Huynegem, J. (2017). Intelligentiemeting in nieuwe banen: CoVaT-CHC. *Caleidoscoop*, 4(27).
- Watkins, M. W., & Beaujean, A. A. (2014). Bifactor structure of the Wechsler Preschool and Primary Scale of Intelligence-Fourth Edition. *School Psychology Quarterly*, 29(1), 52–63. <https://doi.org/10.1037/spq0000038>
- Weiss, E. M., Kemmler, G., Deisenhammer, E. A., Fleischhacker, W. W., & Delazer, M. (2002). *Sex differences in cognitive functions*. www.elsevier.com/locate/paid
- White, S. H. (2000). Conceptual foundations of IQ testing. *Psychology, Public Policy, and Law*, 6(1), 33–43. <https://doi.org/10.1037/1076-8971.6.1.33>
- Wilson, C. J., Bowden, S. C., Vannier, L.-C., Byrne, L. K., & Weiss, L. G. (2023). Measurement invariance of the Wechsler Intelligence Scale for children-Fifth Edition, in Australian and New Zealand and U.S. standardization samples. *Psychological Assessment*, 35(6), 510–521. <https://doi.org/10.1037/pas0001221>
- Young, S. R., & Keith, T. Z. (2020). An examination of the convergent validity of the ICAR16 and WAIS-IV. *Journal of Psychoeducational Assessment*, 38(8), 1052–1059. <https://doi.org/10.1177/0734282920943455>