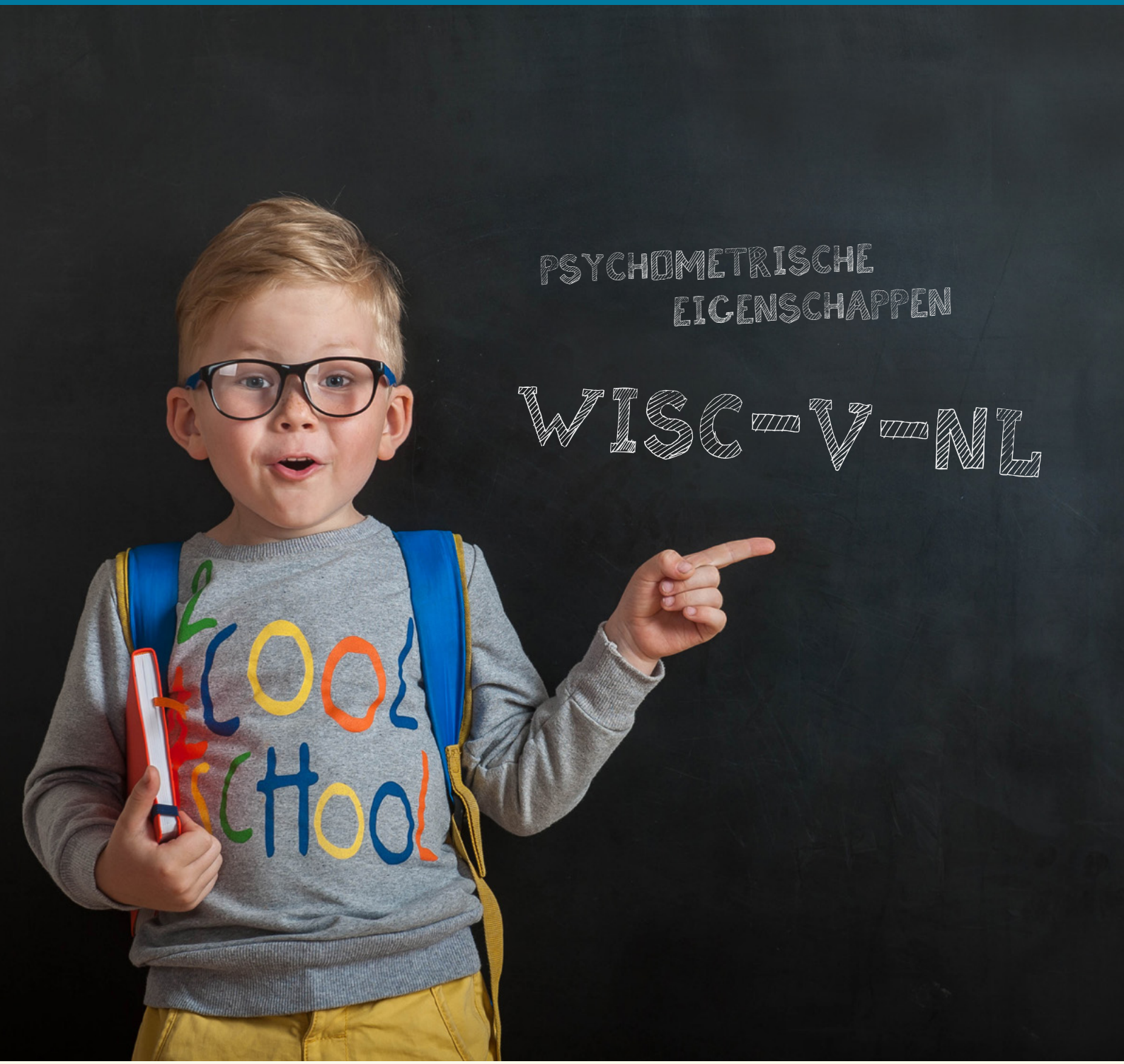


# WISC-V-NL

## Psychometrische eigenschappen

DEEL 2 VAN 3

Pearson

WWW.PEARSONCLINICAL.NL WWW.PEARSONCLINICAL.BE PEARSON BENELUX B.V.

# Inhoud

# 1 Samenvatting

Het normeringsonderzoek van de WISC-V-NL heeft plaatsgevonden van april 2016 tot juli 2017, onder 395 Vlaamse respondenten en 1038 Nederlandse respondenten. De steekproeven zijn gestratificeerd op sekse, regio, urbanisatiegraad, opleidingsniveau moeder, opleidingsniveau kind en etniciteit, voor Vlaanderen tevens op onderwijsnet. Het zijn representatieve steekproeven voor de populaties van Nederland en Vlaanderen, welke de basis vormen voor een tweetal normen; namelijk normen gebaseerd op alleen Nederlandse data en normen op basis van gecombineerde Vlaamse en Nederlandse data.

Algeheel onderzoek naar de betrouwbaarheid van de WISC-V-NL laat zien dat de betrouwbaarheid van het TIQ in de totale steekproef goed is (.95) voor belangrijke beslissingen op individueel niveau. Voor de indexen zijn de betrouwbaarheden voor de totale gecombineerde steekproef allen voldoende (Verbaal Begrip Index; .89) tot goed (Non-Verbale Index; .95). Wanneer gekeken wordt naar de betrouwbaarheden van de subtests van de gecombineerde steekproef, dan liggen deze tussen de onvoldoende (Begrijpen; .75) en goed (Gewichten; .91) voor belangrijke beslissingen op individueel niveau. Voor minder belangrijke beslissingen op individueel niveau is dit voldoende tot goed. Interpretatie van de prestaties op indexniveau is dus betrouwbaarder en heeft zodoende altijd de voorkeur boven het interpreteren van individuele prestaties op subtestniveau.
Bij de test-hertest wordt voor de gehele test-hertestgroep (N=81) voor het TIQ een gecorrigeerde correlatie van .95 voor het TIQ gevonden, een goed resultaat dat de stabiliteit van de test bevestigt. Voor de indexscores lopen de gecorrigeerde correlaties van voldoende (Verbaal Begrip Index en Verwerkingssnelheid Index; .76) tot goed (Non-Verbale Index en Algemene Vaardigheid Index; .92). De gecorrigeerde correlaties van de subtests liggen tussen de onvoldoende (Plaatjesreeksen; .62) en goed (Rekenen; .82), waarvan de meerderheid in de range van voldoende (Begrijpen, Blokpatronen, Figuur Samenstellen, Cijfers en Letters Nazeggen, Symbool Substitutie Coderen, Symbool Zoeken en Figuur Zoeken).
Uit onderzoek naar de interbeoordelaarsbetrouwbaarheid blijkt dat de gemiddelde intraclass Pearsoncorrelatie (McGraw & Wong, 1996; Shrout & Fleiss, 1979) tussen de drie beoordelaars voor Overeenkomsten .98 is, voor Woordenschat .99 en voor Begrijpen .98. Deze resultaten geven aan dat hoewel deze subtests een meer subjectieve scoring vereisen, er toch een hoge mate van consistentie tussen de scores is.

Om de validiteit van de WISC-V-NL te onderzoeken, is onder meer onderzoek gedaan naar de inhoudsvaliditeit, responsprocessen, interne structuur en intercorrelaties. Hieruit blijkt dat de algehele validiteit goed wordt ondersteund. Uit confirmatieve factoranalyse blijkt ook op basis van de Nederlandse en Vlaamse data dat opnieuw ondersteuning wordt gevonden voor een vijffactor-model dat eerder al werd gevonden in de Amerikaanse en andere internationale data. Onderzoek naar relaties met andere tests laat zien dat er sterke verbanden gevonden worden tussen het TIQ van de WISC-V-NL en de TIQ's van respectievelijk de WISC-III$^{NL}$, de WPPSI-III-NL en de WAIS-IV-NL en tussen overeenkomstige IQ-scores van de WISC-V-NL en de neventests. Daarnaast is over het algemeen ook sprake van een grote samenhang tussen de vergelijkbare indexen en subtests van de WISC-V-NL en eerder genoemde tests. Wat betreft de criteriumvaliditeit, wordt bewijs gevonden voor gelijktijdige criteriumvaliditeit. Een onderzoeksopzet waarbij de voorspellende kwaliteit van de WISC-V-NL gerelateerd wordt aan bijvoorbeeld opleidingsniveau en studiesucces in de toekomst zou de criteriumvaliditeit van de WISC-V-NL verder moeten onderbouwen.

# 2 Inleiding

In deze whitepaper wordt een samenvatting gegeven van het normeringsonderzoek dat gedaan is naar de WISC-V-NL en op basis waarvan de normen berekend zijn. Daarnaast worden delen van het onderzoek en de resultaten van het betrouwbaarheids- en validiteitsonderzoek globaal besproken. Dit is gedaan zodat toekomstige gebruikers van de WISC-V-NL zich kunnen informeren over de psychometrische eigenschappen van de WISC-V-NL voordat zij de Technische Handleiding aangeschaft hebben. In de tekst worden ook enkele aanbevelingen gedaan over hoe de gevonden resultaten geïnterpreteerd dienen te worden. Benadrukt moet worden dat deze whitepaper slechts een beknopte weergave is van het onderzoek. Voor het volledige onderzoek inclusief alle resultaten verwijzen we u naar de Technische Handleiding van de WISC-V-NL.

# 3 Normeringssteekproef

Normeringsdata voor de WISC-V-NL zijn van april 2016 tot juli 2017 in Nederland onder 1038 Nederlandse respondenten verzameld. De Vlaamse normeringsdata zijn verzameld tussen oktober 2016 en juni 2017 onder 395 Vlaamse respondenten. De uiteindelijke steekproef bevatte na weging 1396 personen tussen de 6;0 en 16;11 jaar, waarvan 1035 uit Nederland en 361 uit Vlaanderen.

In deze paragraaf wordt een beknopte weergave gegeven van de verzamelde steekproeven die ten grondslag liggen aan beide normeringen en de rechtvaardiging van de normconstructie. Een compleet overzicht vindt u in paragraaf 4.4 van de Technische Handleiding.

## Totstandkoming en kwaliteit steekproeven

Tijdens het verzamelen van de benodigde personen is gestratificeerd op land, leeftijd, sekse, regio, migratieachtergrond, urbanisatiegraad en opleidingsniveau van moeder en kind om tot een zo goed mogelijke afspiegeling van de doelpopulatie te komen. Na het verzamelen van alle afnames is de gehele steekproef vergeleken met streefpercentages, gebaseerd op de ten tijde van de steekproefsamenstelling recentste gegevens van respectievelijk het Centraal Bureau voor de Statistiek (CBS, 2015), van Statbel (2015) en informatie gebaseerd op de Vlaamse Onderwijsstatistieken (2015): Vlaams onderwijs in cijfers 2014-2015 en het Statistisch jaarboek van het Vlaams onderwijs over schooljaar 2014-2015 (het Vlaamse Ministerie van Onderwijs en Vorming, 2015). Dit is gedaan om te controleren of de representativiteit voldoende was en om een geringe weging uit te voeren om de steekproef zo goed mogelijk in overeenstemming te brengen met de populatie. De steekproeven van Nederland en Vlaanderen zijn daarbij apart beschouwd.

### Representativiteit Nederlandse steekproef

De verdeling van jongens en meisjes per leeftijdsgroep is goed te noemen, zowel voor als na weging. Uitzondering hierop is leeftijdsgroep 7:0-7:11 voor weging. Na weging is ook voor deze leeftijdsgroep de verdeling van jongens en meisjes goed met een maximale afwijking van 3.4%. Wat opleidingsniveau moeder betreft, komen de steekproefpercentages zowel voor als na weging goed overeen met de populatiecijfers. De verschillen zijn na weging ten hoogste 2.4% (oververtegenwoordiging opleidingsniveau hoog).
Wat betreft opleidingsniveau kind, zien we in de leeftijdsgroep 6-12 jaar een goede overeenkomst tussen de steekproef en populatie voor basisonderwijs en speciaal basisonderwijs, maar 1.7% te weinig kinderen uit het speciaal onderwijs. Kijken we echter naar de totale groep kinderen van 6 tot 17 jaar, dan komt het percentage uit het speciaal onderwijs precies overeen met de populatie. Wat betreft opleidingsniveau van kinderen in het secundair onderwijs zien we voor weging een te groot verschil tussen steekproefpercentage en populatiecijfer (9.7% ondervertegenwoordiging van bovenbouw vmbo/mbo). Dit heeft te maken met de bereidheid tot medewerking aan het onderzoek. Deze is bij de bovenbouw van havo en vwo hoger dan bij de bovenbouw van vmbo of bij het mbo. Na weging verdwijnt deze onder- en oververtegenwoordiging. Voor alle opleidingsniveaus geldt na weging dat het steekproefpercentage erg dicht ligt bij het populatiepercentage (met een maximaal verschil van 1.4%).

De verdeling van kinderen met een Nederlandse achtergrond versus kinderen met een migratie-achtergrond in de totale Nederlandse steekproef in vergelijking met de populatie is goed te noemen en het verschil met de populatiecijfers is minimaal (0.7%). Hier is geen weging op toegepast.
Ten slotte is er gestratificeerd op de variabelen regio en urbanisatie. De regio West (+4.1%) en in mindere mate de regio Noord (+3%) waren oververtegenwoordigd in de steekproef vóór weging en de regio Zuid (-5%) ondervertegenwoordigd. Na weging zijn deze verschillen minimaal; het grootste verschil is 0.5%.
Voor urbanisatie zien we dat stad voor weging oververtegenwoordigd was (+4.5%) in de steekproef en platteland ondervertegenwoordigd (-5.5%). Na weging is dit verschil geminimaliseerd en liggen de steekproefpercentages van stad, verstedelijkt en platteland heel dicht bij de populatiecijfers, met een maximaal verschil van 2.7%.

### Representativiteit Vlaamse steekproef

Voor weging was de verhouding jongens/meisjes in enkele leeftijdsgroepen niet geheel gelijkwaardig en daarom is hier een weging voor uitgevoerd. Na weging lag de verdeling van sekse in de aparte leeftijdsgroepen dichter bij de gewenste 50%-50%-verdeling, alleen in de groep 12- en 16-jarigen was de afwijking groter dan 5%.
De verdeling van opleidingsniveau moeder in de Vlaamse steekproef is na weging maximaal 2.5%. De Vlaamse steekproef correspondeert over de gehele steekproef heen, na weging, uitstekend met de gewenste indeling met betrekking tot het opleidingsniveau van de jongere (nergens meer dan 1% afwijkend van het streefcijfer).
Specifiek voor Vlaanderen werd ervoor gekozen om ook voor onderwijsnet (OGO, VGO, GO) te monitoren tijdens de steekproefsamenstelling. De percentageverschillen zijn acceptabel; er is niet voor gewogen.
Van de Vlaamse respondenten zijn tevens de nationaliteit en het geboorteland van hun ouders bevraagd. Op basis daarvan werden drie groepen onderscheiden (autochtoon, TMA-allochtoon, andere nationaliteit). Er was sprake van een lichte oververtegenwoordiging in de verzamelde data van kinderen met een andere nationaliteit (+3.3%). Dit is echter slechts een marginaal verschil.
Er bleek wat betreft regio een klein verschil te bestaan tussen het percentage werkelijk geteste kinderen en de streefcijfers. Het betrof een ondervertegenwoordiging van kinderen uit de regio Oost (-2.5%), samengaand met een ‘oversampling’ in regio West van 5.8% (waar ook de meeste testleiders vandaan kwamen). Ten behoeve van de constructie van de normen werd toch besloten een weging voor regio toe te passen.
Met betrekking tot urbanisatiegraad als variabele van de spreiding van de Vlaamse normgroep is gebleken dat de gevonden percentages geteste kinderen voldoende corresponderen met de streefpercentages. Voor deze variabele werd geen weging toegepast.

Concluderend kan gesteld worden dat de WISC-V-NL normeringssteekproef zowel voor Nederland als voor Vlaanderen een representatieve afspiegeling is van de beoogde doelgroep van de WISC-V-NL op alle stratificatievariabelen. Verschillen tussen de steekproefpercentages en populatiepercentages zijn voor weging in alle gevallen acceptabel tot zeer acceptabel. Na weging (waarbij nooit met meer dan factor 2 is gewogen, zoals aangegeven in het COTAN-beoordelingssysteem (Evers et al., 2010)), zijn deze verschillen verder geminimaliseerd en in alle gevallen zeer acceptabel te noemen.

## Twee normeringen

Na verzameling en weging van de Nederlandse en Vlaamse steekproef werd onderzocht welke verschillen er zijn tussen de resultaten van beide landen, mede om te bepalen in hoeverre het samenstellen van normen gebaseerd op een gecombineerde steekproef Vlaanderen en Nederland gerechtvaardigd is.
Er werd een serie MANOVA's uitgevoerd om de overeenstemming tussen de Nederlandse en Vlaamse ruwe scores op indexniveau te onderzoeken. Een MANOVA heeft, ten opzichte van T-toetsing, het voordeel dat er wordt gecontroleerd voor de kans op Type-1 fouten. In het model werden Leeftijd en Land opgenomen als onafhankelijke variabelen. Voor deze variabelen is gekozen, omdat er verwacht werd dat leeftijd een groot effect heeft op de resultaten en de overige variabelen representatief gestratificeerd zijn per leeftijdsgroep. Een optelsom van de ruwe scores van de primaire subtests onderliggend aan de vijf Indexscores en het Totaal IQ werden opgenomen als afhankelijke variabelen. De resultaten van deze analyses worden weergegeven in Tabel 1.
Uit de resultaten bleek dat geen van de vijf indexscores een significant effect liet zien voor de variabele Land, met een minimale p-waarde van .21 (VBI). Wanneer het TIQ werd bekeken (welke bestaat uit zeven subtests) bleek er wel sprake van een significant effect (p=.03). De effectgrootte van dit effect is echter klein volgens Cohen's criteria, namelijk eta2=.01. Dit betekent dat slechts 1% van de scorevariantie van het TIQ verklaard wordt door Land als variabele. Er worden zodoende geen bezwaren gevonden tegen het samenvoegen van beide datasets.

Op basis van de representatieve gewogen steekproeven zijn twee sets normen geconstrueerd met behulp van inferentieel normeren, een vorm van continue normering:

- een gecombineerde norm op basis van beide steekproeven;
- een Nederlandse norm op basis van alleen Nederlandse data.

Tabel 1 Analyse resultaten Manova's voor de variabelen Land (Nederland en Vlaanderen) en leeftijd

| Indexscore | Factor | Wilks' Lambda | *F*-waarde | *P*-waarde | Eta$^2$ |
|---|---|---|---|---|---|
| **VBI** | Land | 0.998 | 1.58 | 0.21 | 0.00 |
| | Leeftijd | 0.406 | 1044.39 | <.01** | 0.36 |
| | Leeftijd x Land | 0.998 | 1.24 | 0.29 | 0.00 |
| **VRI** | Land | 0.999 | 0.89 | 0.41 | 0.00 |
| | Leeftijd | 0.645 | 392.19 | <.01** | 0.20 |
| | Leeftijd x Land | 1.000 | 0.27 | 0.77 | 0.00 |
| **FRI** | Land | 1.000 | 0,12 | 0.89 | 0.00 |
| | Leeftijd | 0.670 | 351.60 | <.01** | 0.18 |
| | Leeftijd x Land | 1.000 | 0.23 | 0.79 | 0.00 |
| **Wgl** | Land | 0.999 | 0.93 | 0.39 | 0.00 |
| | Leeftijd | 0.644 | 393.72 | <.01** | 0.20 |
| | Leeftijd x Land | 0.999 | 0.89 | 0.41 | 0.00 |
| **Vsl** | Land | 0.999 | 0.38 | 0.69 | 0.00 |
| | Leeftijd | 0.535 | 603.98 | <.01** | 0.27 |
| | Leeftijd x Land | 0.999 | 0.76 | 0.47 | 0.00 |
| **TIQ** | Land | 0.989 | 2.22 | 0.03* | 0.01 |
| | Leeftijd | 0.357 | 358.21 | <.01** | 0.14 |
| | Leeftijd x Land | 0.992 | 1.65 | 0.12 | 0.00 |

* Significant op .05-niveau
** Significant op .01-niveau

## Kwaliteit normeringen

De standaardfout van het gemiddelde is één van de maten die aangeeft hoeveel de gemiddelden van steekproef tot steekproef kunnen verschillen. Hoe groter de standaardfout, des te meer de gemiddelden van steekproef tot steekproef kunnen variëren. Een grote standaardfout geeft aan dat het steekproefgemiddelde geen goede schatter is van het populatiegemiddelde. De standaardfout hangt onder andere af van de steekproefgrootte (Slotboom, 2008). De COTAN (Evers et al., 2010) gaat, in de richtlijnen voor subgroepsgrootte bij continue normering met acht subgroepen, voor de kwalificatie 'goed' uit van een maximale standaardfout van .75 op basis van een N van 400 bij klassiek normeren.

Op basis van de WISC-V-NL steekproefdata is de standaardfout van de som van zeven geschaalde subtestscores (die bijdragen aan het TIQ) per leeftijdsgroep berekend indien er sprake zou zijn geweest van klassiek normeren bij N = 300 en N = 400. Daarnaast is de standaardfout van de som van zeven geschaalde subtestscores (die bijdragen aan het TIQ) per leeftijdsgroep berekend voor het toegepaste WISC-V-NL normeringsmodel. Hiermee wordt nagegaan of de normgroepsgroottes voor beide gepresenteerde normgroepen (Nederland en Nederland en Vlaanderen gecombineerd) minimaal equivalent zijn aan de gewenste N bij klassiek normeren en daarmee acceptabel qua omvang zijn.

Uit de analyses blijkt dat de standaardfout in de Nederlandse steekproef in alle normgroepen onder die van N=300 ligt en in Nederland in 9 van de 11 groepen ook onder die van N=400. Dit is echter acceptabel, zolang de mate van verslechtering ten opzichte van klassiek normeren beperkt is en de winst bij de middelste groepen groot (Evers et al., 2010). In de gecombineerde steekproef ligt de standaardfout in alle normgroepen onder die van 400. Concluderend kan gesteld worden dat beide steekproefgroottes van voldoende omvang zijn. Daarnaast ligt de standaardfout in alle gevallen onder de standaardfout van .75, waar de COTAN van uit is gegaan bij de bepaling van de COTAN-richtlijnen. De groottes van de Nederlandse en de gecombineerde steekproeven zijn dus volgens de COTAN-richtlijnen 'goed' te noemen.

Er zijn weinig verschillen tussen beide normen, de gecombineerde norm is bruikbaar voor beide landen, maar er is vanuit sommige Nederlandse gebruikers vraag om een Nederlandse norm te gebruiken op basis van puur Nederlandse data. De grootte van de Nederlandse data staat dit kwalitatief gezien ook toe, de aparte Vlaamse dataset zou voor dit doeleinde qua normgroepsgrootte te klein zijn. Echter, de Nederlandse normen verschillen alleen op subtest-niveau van de gecombineerde normen, op indexniveau zijn de normen voor beide normeringen (gecombineerd en Nederland) gelijk omdat de resultaten hierbij geen verschillen lieten zien. De verschillen op subtestniveau zijn minimaal. Dit plaatst het belang van de afzonderlijke normen ook weer in perspectief: het TIQ en de indexen zijn de belangrijkste en betrouwbaarste uitkomsten van de test. Analyse op subtest- en itemniveau is minder betrouwbaar.

# 4 Betrouwbaarheid

Om de betrouwbaarheid van de WISC-V-NL te beoordelen, is onderzoek gedaan naar de interne consistentie, standaardmeetfouten, test-hertestbetrouwbaarheid en interbeoordelaars-betrouwbaarheid. Enkele van deze onderdelen worden in deze whitepaper besproken. Voor uitgebreide informatie over alle betrouwbaarheidsonderzoeken verwijzen we naar hoofdstuk 5 van de Technische Handleiding.
Wat betreft de interne consistentie, is de belangrijkste conclusie dat de betrouwbaarheid van het TIQ in de totale steekproef goed is (.95) voor belangrijke beslissingen op individueel niveau. Bij de test-hertest wordt voor het TIQ een gecorrigeerde correlatie van .95 gevonden, een goed resultaat dat de stabiliteit van de test bevestigt. De interbeoordelaarsbetrouwbaarheid van de verbale subtests is goed te noemen. Hieronder volgt een meer gedetailleerde beschrijving van een aantal betrouwbaarheidsonderzoeken.

## Interne consistentie

Voor het weergeven van de betrouwbaarheid van de WISC-V-NL zijn Guttman lambda2 en de gecorrigeerde Guttman lambda2 berekend (Guttman, 1945). Alle betrouwbaarheden zijn berekend op de grootst mogelijke dataset en dus gebaseerd op de gecombineerde Nederlandse en Vlaamse dataset.
Omdat de subtests Symbool Substitutie Coderen, Symbool Zoeken en Figuur Zoeken enkel snelheidsgerelateerde taakopdrachten omvatten en niet bestaan uit meerdere items, is het niet mogelijk om lambda2 voor deze subtests te berekenen. Voor Symbool Substitutie Coderen en Symbool Zoeken wordt daarom de split-half betrouwbaarheid gerapporteerd. Voor Figuur Zoeken geldt dat de betrouwbaarheden voor de subtestsscore en processcores (FZw en FZg) gebaseerd zijn op de resultaten van test-hertest betrouwbaarheden van twee groepen; een groep van 6-9 jaar en een groep van 10-16 jaar. De COTAN (Evers et al., 2010) geeft ook aan dat voor een test die primair verwerkingssnelheid meet de test-hertest betrouwbaarheid een goede maat is voor de interne consistentie.
In Tabel 2 worden de betrouwbaarheden weergegeven voor de subtests, processcores en indexscores, per leeftijdsgroep en voor de totale steekproef voor Nederland en Vlaanderen gecombineerd. Op basis van de betrouwbaarheidscoëfficiënten in Tabel 2 valt af te leiden dat de betrouwbaarheden in de meeste gevallen op basis van de COTAN-criteria als voldoende en goed kunnen worden aangemerkt. Voor de indexen liggen de betrouwbaarheden voor de totale gecombineerde steekproef tussen de .89 (Verbaal Begrip Index en Visueel Ruimtelijke Index) en .95 (Non-Verbale Index). De indexscores van de Verbaal Begrip Index en de Visueel Ruimtelijke Index kwalificeren daarmee als ‘voldoende’. De overige indexen hebben alle een waarde gelijk aan of boven de .90 en kwalificeren voor ‘goed’. In de afzonderlijke leeftijdsgroepen liggen de betrouwbaarheden voor de indexen tussen .82 en .96. Dit is op basis van de COTAN-richtlijnen goed te noemen. Deze betrouwbaarheden liggen over het algemeen hoger dan die van de aparte subtests die onderliggend zijn aan de indexen. Dit verschil is een algemeen waargenomen patroon en ontstaat doordat elke subtest slechts een klein deel van het cognitieve functioneren van het kind representeert, terwijl de indexscores de prestaties van het kind op een breder vlak van vaardigheden opsommen. De resultaten geven bovendien aan dat de betrouwbaarheid van het TIQ in de totale steekproef .95 is en voor de aparte leeftijds-groepen varieert tussen .94 en .96. Dit is dus zeer goed te noemen. Het TIQ kan aangemerkt

worden als het betrouwbaarste resultaat in het kader van beslissingen op individueel niveau. De Non-Verbale Index en Algemene Vaardigheid Index hebben eveneens hoge betrouwbaarheden, wat ze naast hun theoretische bruikbaarheid ook psychometrisch geschikt maakt als alternatieven voor het TIQ.

Wanneer gekeken wordt naar de betrouwbaarheden van de subtests van de gecombineerde steekproef, dan liggen deze tussen .80 (Blokpatronen) en .91 (Gewichten), met uitzondering van Begrijpen (.76). Deze waarden volgen hetzelfde patroon van hoge en minder hoge betrouwbaarheden als de Amerikaanse data. In de aparte leeftijdsgroepen worden iets lagere betrouwbaarheden gevonden, maar niet onder de .70, met uitzondering van Woordenschat bij 7-jarigen (.62) en Begrijpen bij 7- en 9-jarigen (.69). Deze zijn dus voldoende tot goed te noemen.

De betrouwbaarheidscoëfficiënten van de processcores voor de totale steekproef liggen tussen de .73 (Cijferreeksen Voorwaarts) en .85 (Blokpatronen deelscore). Indien men kijkt naar de aparte leeftijdsgroepen ziet men hetzelfde beeld, behalve voor Cijferreeksen Voorwaarts in de leeftijdsgroepen 7, 8 en 9 jaar. De lagere waarden in die groepen zouden kunnen zijn ontstaan door steekproeffluctuaties, doordat kinderen op deze leeftijd minder stabiel scoren of doordat deze leeftijdsgroepen minder items maken. De betrouwbaarheden in de aparte Nederlandse steekproef komen sterk overeen met de bevindingen van de gecombineerde steekproef. Daarmee doen beide steekproeven qua betrouwbaarheid niet voor elkaar onder.

Tabel 2 Betrouwbaarheidscoëfficiënten lambda per subtest-, proces- en indexscore per leeftijdsgroep (Nederland en Vlaanderen gecombineerd)

| Subtest-/ Proces-/ indexscore | leeftijdsgroep | | | | | | | | | | | totale steekproef $r_{xx}$[3] |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | |
| **OV** | .84 | .84 | .83 | .81 | .83 | .83 | .86 | .86 | .85 | .86 | .87 | .84 |
| **WS** | .81 | .62 | .77 | .76 | .77 | .81 | .83 | .80 | .83 | .87 | .88 | .80 |
| **BG** | .76 | .69 | .70 | .69 | .75 | .75 | .76 | .77 | .80 | .81 | .80 | .76 |
| **BP** | .80 | .81 | .81 | .81 | .80 | .80 | .76 | .81 | .80 | .80 | .79 | .80 |
| **FS** | .85 | .86 | .88 | .85 | .87 | .86 | .83 | .84 | .82 | .87 | .89 | .86 |
| **MR** | .83 | .89 | .87 | .83 | .86 | .85 | .82 | .84 | .81 | .81 | .80 | .84 |
| **GW** | .91 | .92 | .92 | .92 | .92 | .91 | .91 | .91 | .91 | .90 | .93 | .91 |
| **RE** | .88 | .89 | .88 | .89 | .88 | .86 | .86 | .88 | .89 | .89 | .90 | .88 |
| **CR** | .85 | .81 | .82 | .84 | .83 | .82 | .86 | .88 | .85 | .88 | .89 | .85 |
| **PR** | .84 | .83 | .84 | .81 | .85 | .86 | .87 | .82 | .85 | .84 | .80 | .84 |
| **CLN** | - | - | .90 | .84 | .75 | .78 | .84 | .80 | .80 | .82 | .78 | .82 |
| **SSC**[b] | .85 | .83 | .81 | .89 | .86 | .86 | .86 | .85 | .86 | .85 | .80 | .85 |
| **SZ**[b] | .87 | .89 | .85 | .90 | .89 | .91 | .91 | .82 | .80 | .97 | .86 | .87 |
| **FZ**[c] | .74 | .74 | .74 | .74 | .83 | .83 | .83 | .83 | .83 | .83 | .83 | .80 |
| **BPz** | .79 | .80 | .80 | .78 | .80 | .81 | .76 | .79 | .80 | .81 | .82 | .80 |
| **BPd** | .84 | .87 | .90 | .89 | .85 | .86 | .84 | .84 | .83 | .81 | .79 | .85 |
| **CRv** | .73 | .61 | .66 | .68 | .74 | .75 | .73 | .78 | .73 | .78 | .80 | .73 |
| **CRa** | .79 | .80 | .79 | .77 | .79 | .78 | .78 | .79 | .78 | .82 | .81 | .79 |
| **CRs** | .83 | .80 | .82 | .81 | .82 | .81 | .82 | .84 | .83 | .86 | .83 | .83 |
| **FZw**[c] | .71 | .71 | .71 | .71 | .70 | .70 | .70 | .70 | .70 | .70 | .70 | .70 |
| **FZg**[c] | .87 | .87 | .87 | .87 | .77 | .77 | .77 | .77 | .77 | .77 | .77 | .81 |
| **VBI** | .88 | .82 | .88 | .86 | .88 | .90 | .91 | .90 | .91 | .92 | .93 | .89 |
| **VRI** | .89 | .89 | .91 | .89 | .90 | .89 | .87 | .89 | .89 | .90 | .90 | .89 |
| **FRI** | .91 | .94 | .93 | .91 | .92 | .92 | .91 | .92 | .91 | .91 | .91 | .92 |
| **Wgl** | .89 | .88 | .88 | .87 | .90 | .90 | .91 | .91 | .91 | .90 | .90 | .90 |
| **Vsl** | .91 | .91 | .90 | .94 | .92 | .92 | .93 | .90 | .90 | .91 | .88 | .91 |
| **TIQ** | .95 | .94 | .95 | .95 | .95 | .95 | .95 | .96 | .96 | .96 | .96 | .95 |
| **KRI** | .93 | .94 | .93 | .93 | .93 | .92 | .91 | .94 | .94 | .94 | .95 | .93 |
| **AWI** | - | - | .91 | .90 | .87 | .89 | .91 | .91 | .90 | .91 | .91 | .90 |
| **NVI** | .95 | .95 | .95 | .95 | .95 | .95 | .94 | .96 | .96 | .95 | .95 | .95 |
| **AVI** | .94 | .93 | .94 | .94 | .94 | .94 | .94 | .95 | .95 | .95 | .95 | .94 |
| **CCI** | .93 | .93 | .92 | .93 | .94 | .93 | .94 | .94 | .94 | .93 | .93 | .93 |

[a] Gemiddelde betrouwbaarheidscoëfficiënten voor de totale steekproef zijn berekend met behulp van Fisher's z-transformatie.

[b] Voor de subtests SZ en SSC worden split-half betrouwbaarheden vermeld.

[c] Voor de subtest FZ en de processcores FZw en FZg worden test-hertest betrouwbaarheden vermeld.

## Test-hertestbetrouwbaarheid

De stabiliteit van de WISC-V-NL werd onderzocht door middel van het bepalen van de test-hertest betrouwbaarheid. Dit onderzoek werd uitgevoerd bij een heterogene groep van 81 kinderen. Op basis van de resultaten van deze test-hertest groep werden de correlaties tussen de eerste en tweede afname berekend per subtest, processcore en indexscore. In Tabel 3 wordt zowel het gemiddelde als de standaarddeviatie van beide afnames vermeld.
Uit de resultaten blijkt dat de gecorrigeerde correlatie tussen twee afnames van het TIQ .95 is; een goed resultaat dat de stabiliteit van de test bevestigt. De gecorrigeerde correlaties van de primaire en aanvullende indexen liggen tussen de .76 (Verbaal Begrip Index en Verwerkingssnelheid Index) en .92 (Non-Verbale Index en Algemene Vaardigheid Index). De lengte van het tijdsinterval en eventuele gebeurtenissen in de tussentijd beïnvloeden vermoedelijk deze betrouwbaarheden.
De gecorrigeerde correlaties van de subtests liggen tussen de .62 (Plaatjesreeksen) en .82 (Rekenen), en de correlaties tussen de processcores tussen de .71 (Cijferreeksen Voorwaarts, Cijferreeksen Sorteren en Figuur Zoeken Willekeurig) en .83 (Figuur Zoeken Gestructureerd). Deze waarden vallen (op basis van de COTAN-indeling) in de range van onvoldoende (Overeenkomsten, Woordenschat, Matrix Redeneren, Gewichten en Plaatjesreeksen) tot goed (Rekenen en Cijferreeksen), waarvan de meerderheid in de range van voldoende (Begrijpen, Blokpatronen, Figuur Samenstellen, Cijfers en Letters Nazeggen, Symbool Substitutie Coderen, Symbool Zoeken en Figuur Zoeken).
Daarnaast wordt uit de resultaten duidelijk dat er op alle subtests (met uitzondering van Rekenen en Cijferreeksen Sorteren) bij de tweede afname een hogere score behaald wordt. Dit oefen- of leereffect is het sterkst voor Symbool Substitutie Coderen, met een gemiddelde effectgrootte van .62 en op de processcore Figuur Zoeken Willekeurig (gemiddelde effectgrootte van .64).
Bij de indexen wordt het grootste effect gevonden op de Verwerkingssnelheid Index (.62).
Het TIQ gaat in deze steekproef met 5.8 IQ-punten omhoog bij de tweede afname.
Hoewel de effectgroottes klein tot gemiddeld zijn, wordt afgeraden om een kind op korte termijn na een testafname opnieuw te onderzoeken met de WISC-V-NL. Dit omdat er dan een vertekend beeld kan optreden.

Tabel 3 Test-hertest betrouwbaarheid van de subtest-. proces-, en indexscores

| Alle leeftijden Subtest-/Proces-/ indexscore | Eerste afname | | Tweede afname | | | | |
|---|---|---|---|---|---|---|---|
| | Gem. | *SD* | *Gem.* | *SD* | $r12^3$ | Gecorrigeerde $r^b$ | Standaard-verschil[c] |
| **OV** | 10.2 | 2.3 | 11.2 | 2.4 | .40 | .63 | .43 |
| **WS** | 10.8 | 3.1 | 11.1 | 2.6 | .68 | .68 | .10 |
| **BG** | 10.8 | 2.9 | 11.1 | 3.0 | .78 | .79 | .10 |
| **BP** | 10.1 | 2.3 | 11.1 | 2.3 | .51 | .70 | .43 |
| **FS** | 10.6 | 2.5 | 11.3 | 2.7 | .64 | .75 | .27 |
| **MR** | 10.4 | 2.7 | 11.4 | 2.6 | .53 | .63 | .38 |
| **GW** | 10.6 | 3.0 | 11.4 | 3.2 | .65 | .65 | .26 |
| **RE** | 11.1 | 2.7 | 10.9 | 2.8 | .75 | .82 | -.07 |
| **CR** | 11.6 | 3.0 | 11.5 | 2.7 | .79 | .80 | -.04 |
| **PR** | 10.8 | 2.9 | 10.9 | 2.5 | .60 | .62 | .04 |
| **CLN** | 10.9 | 2.7 | 11.5 | 2.9 | .68 | .76 | .21 |
| **SSC** | 10.6 | 2.9 | 12.5 | 3.2 | .71 | .73 | .62 |
| **SZ** | 10.2 | 3.1 | 12.0 | 3.5 | .73 | .71 | .54 |
| **FZ** | 10.0 | 2.8 | 11.5 | 2.8 | .77 | .79 | .54 |
| **BPz** | 10.3 | 2.0 | 11.2 | 2.3 | .50 | .78 | .42 |
| **BPd** | 9.9 | 2.4 | 11.0 | 2.4 | .56 | .73 | .46 |
| **CRv** | 10.7 | 2.6 | 11.2 | 2.6 | .56 | .71 | .19 |
| **CRa** | 11.1 | 2.6 | 11.4 | 2.6 | .68 | .76 | .12 |
| **CRs** | 11.4 | 3.1 | 11.1 | 2.8 | .74 | .71 | -.10 |
| **FZw** | 9.8 | 3.0 | 11.7 | 2.9 | .70 | .71 | .64 |
| **FZg** | 9.9 | 2.6 | 11.2 | 2.8 | .77 | .83 | .48 |
| **VBI** | 102.7 | 12.1 | 105.9 | 12.1 | .62 | .76 | .26 |
| **VRI** | 101.8 | 11.6 | 106.3 | 11.6 | .71 | .83 | .39 |
| **FRI** | 103.4 | 12.8 | 108.5 | 13.7 | .68 | .77 | .38 |
| **Wgl** | 106.6 | 14.3 | 107.0 | 12.5 | .84 | .85 | .03 |
| **Vsl** | 102.4 | 15.6 | 112.9 | 18.2 | .78 | .76 | .62 |
| **TIQ** | 105.1 | 11.1 | 110.9 | 12.6 | .91 | .95 | .49 |
| **KRI** | 104.9 | 14.6 | 106.4 | 15.1 | .80 | .81 | .10 |
| **AWI** | 106.2 | 14.2 | 108.1 | 14.8 | .81 | .86 | .13 |
| **NVI** | 104.9 | 12.0 | 111.0 | 13.3 | .87 | .92 | .48 |
| **AVI** | 103.1 | 10.5 | 107.9 | 12.4 | .83 | .92 | .42 |
| **CCI** | 105.6 | 15,9 | 112.0 | 16.3 | .87 | .85 | .40 |

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

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

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

## Interbeoordelaarsbetrouwbaarheid

Bij een aantal subtests kan het lastig zijn om op geheel gestandaardiseerde wijze te scoren, zoals bij de verbale subtests Overeenkomsten, Woordenschat en Begrijpen, omdat er meerdere juiste antwoorden mogelijk zijn. Bij deze subtests zijn uitgebreide voorbeeldantwoorden aanwezig in de Afname- en scoringshandleiding om de scoring toch zo gestandaardiseerd mogelijk uit te voeren. Uit onderzoek naar de interbeoordelaarsbetrouwbaarheid blijkt dat de gemiddelde intraclass Pearsoncorrelatie (McGraw & Wong, 1996; Shrout & Fleiss, 1979) tussen de drie beoordelaars voor Overeenkomsten .98 is, voor Woordenschat .99 en voor Begrijpen .98. Deze resultaten geven aan dat, hoewel deze subtests een meer subjectieve scoring vereisen, er toch een hoge mate van consistentie in de scores is. De Afname- en scoringshandleiding biedt dus voldoende informatie om de antwoorden van een kind op gestandaardiseerde wijze te scoren.

# 5 Validiteit

Om de validiteit van de WISC-V-NL te beoordelen, is onderzoek gedaan naar de inhouds-validiteit, responsprocessen, interne structuur, intercorrelaties, confirmatieve factoranalyses, relaties met andere tests, klinische en specifieke groepen, relatie met biografische gegevens en criteriumvaliditeit. Enkele van deze onderdelen worden in deze whitepaper besproken, voor uitgebreide informatie over alle validiteitsonderzoeken verwijzen we naar hoofdstuk 6 van de Technische Handleiding.
Resultaten van de confirmatieve factoranalyse laten zien dat het (in alle uitgaven van de WISC-V) gebruikte factormodel aansluit bij de gecombineerde Nederlandse en Vlaamse data. Onderzoek naar relaties met andere tests laat zien dat er sterke verbanden gevonden tussen het TIQ en de IQ-scores van respectievelijk de WISC-III$^{NL}$, de WPPSI-III-NL en de WAIS-IV-NL.

## Teststructuur

Er is niet één specifieke theorie die de structuur van de WISC-V-NL heeft bepaald. Het WISC-V-model is echter wel een weergave van hedendaagse structurele intelligentietheorieën zoals CHC (Catell-Horn-Carroll) en verdedigbare theoretische perspectieven en frameworks. Algemeen geaccepteerde structurele intelligentiemodellen gebaseerd op factoranalytische resultaten, zoals CHC-theorie, leveren overweldigend bewijs voor algemene intelligentie bovenaan een hiërarchisch model, en voor verschillende gerelateerde en onderscheidbare brede vaardigheden op het niveau daaronder. In sommige modellen bestaan de specifieke vaardigheden elk uit verschillende beperkte vaardigheden op het laagste niveau. Ook al komt het bewijsmateriaal uit structurele modellen niet exact samen, toch blijkt uit de meeste modellen dat vaardigheden voor verbaal begrip, fluid redeneren, werkgeheugen en verwerkingssnelheid, en visueel-ruimtelijke vaardigheden de belangrijkste onderdelen vormen, en dit zijn dan ook de vijf primaire index-scores die beschikbaar zijn voor de WISC-V. De namen van deze factoren verschillen afhankelijk van de taxonomie die door een groep onderzoekers wordt gebruikt; de CHC-taxonomie geeft namen voor deze constructen (d.w.z. respectievelijk Gc, Gv, Gf, Gsm en Gs). De Wechsler intelligentieschalen zijn als reactie op deze consistent waargenomen factoren verder ontwikkeld en de WISC-V-NL zet dit werk voort door middel van nieuwe maten voor werkgeheugen en een nieuwe indexscore voor werkgeheugen, en door aparte visueel-ruimtelijke indexscores en index-scores voor fluid redeneren. Ook zijn de maten voor verbaal begrip en verwerkingssnelheid verbeterd, terwijl er voor beide indexscores blijven bestaan.

## Factoranalyse

Het hiërarchische factormodel van de WISC-V-NL (met vijf Indexscores en één hogere-orde factor) is onderzocht aan de hand van structurele vergelijkingsmodellen (Jöreskog & Sörbom, 1993). Deze confirmatieve analyses toetsen of de verwachte structuur met vijf factoren vol-doende wordt gesteund door de data.

De ontwikkeling van de WISC-V-NL berustte op de theoretische aanname dat de schaal een schatting geeft van algemene cognitieve vaardigheid die naar voren komt in vijf cognitieve domeinen (verbaal begrip, visueel-ruimtelijk, fluide redeneren, werkgeheugen en verwerkings-

snelheid). Dit komt overeen met een tweede-orde factormodel met vijf eerste-orde factoren (de vijf vaardigheidsdomeinen) en een tweede-orde factor (algemene intelligentie, oftewel g). De drie nieuwe primaire en secundaire subtests (Figuur Samenstellen, Gewichten en Plaatjesreeksen) werden ontwikkeld met het uitdrukkelijke doel het afzonderlijk meten van respectievelijk visueel-ruimtelijke vaardigheden, fluide redeneren en werkgeheugenvaardigheden te versterken. De daadwerkelijke plaatsing van de nieuwe subtests binnen het vijffactormodel werd duidelijk nadat een serie confirmatieve factoranalyses op basis van de normatieve steekproef werd uitgevoerd. Nadat alle normeringsdata waren verzameld, werd een reeks confirmatieve factoranalyses uitgevoerd. Hierbij werd gebruikgemaakt van de gecombineerde Nederlandse en Vlaamse dataset.

De modellen die werden getoetst waren de volgende:
**Model 1:** een eenfactormodel, waarin alle subtests rechtstreeks laden op de g-factor algemene cognitieve vaardigheid.
**Model 2:** het historische tweefactormodel met aparte verbale en performale factoren.
**Model 3:** een driefactormodel met een gecombineerde factor verbaal begrip en auditief werkgeheugen (het vroegere VIQ), een gecombineerde factor visueel-ruimtelijk, fluide redeneren en visueel werkgeheugen (het vroegere PIQ) en een verwerkingssnelheidfactor.
**Model 4a - 4d:** vierfactormodellen welke identiek zijn in de wijze waarop de factoren verbaal begrip en verwerkingssnelheid gedefinieerd zijn, maar verschillen in de manier waarop de overige negen subtests worden ingedeeld.
**Model 5a - 5e:** variaties van een vijffactormodel dat verbaal begrip, visueel-ruimtelijk, fluide redeneren, werkgeheugen en verwerkingssnelheid bevat. Zij verschillen alleen in hun patronen van de ladingen voor Rekenen.

De strategie voor het onderzoeken van de factorstructuur van de WISC-V-NL begon met analyses van alle primaire en secundaire subtests in de gehele leeftijdsgroep. Nadat een optimaal factormodel werd vastgesteld in de set van alle primaire en secundaire subtests, werd de fit van dat model apart geëvalueerd in vijf aparte leeftijdsgroepen: 6-7, 8-9, 10-11, 12-13 en 14-16. Ten slotte is het optimale model getoetst met de tien primaire subtests om te bevestigen dat deze set factoren de correlaties tussen de gereduceerde set subtests bleef verklaren.

De resultaten laten zien dat ten minste vier factoren nodig zijn om een goede fit te verkrijgen, zo blijkt uit de resultaten op de CFI en RMSEA. Alle vijffactormodellen hebben een goede fit, waarbij twee modellen (5c en 5e) een iets betere fit hebben dan alle vierfactormodellen. Als wordt toegestaan dat Rekenen op zowel op fluide redeneren, werkgeheugen als verbaal begrip laadt (model 5e) treedt een significante verbetering op ten opzichte van model 5c, waarbij Rekenen op fluide redeneren en op werkgeheugen laadt. Dit vijffactormodel wordt daarnaast ondersteund in Amerikaans onderzoek naar klinische groepen, waar er een duidelijker onderscheid gemaakt kan worden tussen de Visueel Ruimtelijke Index en Fluid Redeneren Index (Chen, Zhang, Engi Raiford, Zhu, & Weiss, 2015; Wechsler, 2014). Ook in de correlaties met neventests is zichtbaar dat er vijf factoren onderscheiden kunnen worden (zie de Amerikaanse technische handleiding, Wechsler, 2014). Geconcludeerd kan worden dat de resultaten aangeven dat het (in alle uitgaven van de WISC-V) gebruikte factormodel aansluit bij de gecombineerde Nederlandse en Vlaamse data.

## Neventests

Een belangrijk aspect van het validatieonderzoek is het onderzoeken van de samenhang van het nieuwe instrument, in dit geval de WISC-V-NL, met reeds in gebruik zijnde tests die een vergelijkbaar construct meten en waarvan de validiteit zich reeds bewezen heeft. Dit type onderzoek levert informatie op over de begripsvaliditeit. Bij het ontwikkelen van een intelligentietest moeten gebruikers er immers op kunnen vertrouwen dat het construct intelligentie ook daadwerkelijk wordt gemeten. In dat kader is de WISC-V-NL vergeleken met andere intelligentietests, namelijk de WISC-III$^{NL}$, WPPSI-III-NL en de WAIS-IV-NL.
De steekproef bestaat uit kinderen van wie de ouders toestemming gaven om bij hun kind niet alleen een standaardafname van de WISC-V-NL uit te voeren in het kader van de normering, maar tevens een validatietest. Deze validatietests zijn zo veel mogelijk in counterbalanced order afgenomen, dat wil zeggen dat ongeveer de helft van de kinderen eerst met de WISC-V-NL is getest en de andere helft eerst met de validatietest. Op deze manier wordt zo goed mogelijk gecontroleerd voor mogelijke leereffecten.

### WISC-III$^{NL}$

De WISC-III$^{NL}$ (Wechsler, 2003b) is de Nederlandstalige bewerking van de *Wechsler Intelligence Scale for Children – Third Edition* (Wechsler, 1991).

Het vergelijkend onderzoek tussen WISC-III$^{NL}$ en WISC-V-NL is uitgevoerd bij 43 kinderen (22 jongens en 21 meisjes). De gemiddelde leeftijd van de kinderen bij de afname van de WISC-III$^{NL}$ was 11 jaar (SD van 3 jaar). Het tijdsinterval tussen beide afnames was gemiddeld 18 dagen, met een range van 2 tot 44 dagen. De verwachting is dat gemiddelde scores van de WISC-III$^{NL}$ hoger zullen zijn dan de gemiddelde WISC-V-NL-scores, onder andere vanwege het Flynn-effect. Tevens wordt verwacht dat er hoge tot middelmatige correlaties worden gevonden tussen subtests en indexen van beide tests die inhoudelijk en qua meetdomeinen gerelateerd zijn.
De gemiddelde scores van het Totaal IQ en de primaire indexen bevinden zich op beide tests in de gemiddelde range, zoals te zien is in Tabel 4. Zoals verwacht zijn alle gemiddelde scores op de indexen lager voor de WISC-V-NL dan voor de WISC-III$^{NL}$. Hierbij is de effectgrootte voor de Verbaal Begrip Index klein (.22) en zijn de effectgroottes voor alle andere indexen verwaarloosbaar. De gecorrigeerde correlatie van het Totaal IQ van de WISC-III$^{NL}$ met het Totaal IQ van de WISC-V-NL is hoog, namelijk $r$ = .87. Hieruit blijkt dat beide versies van de test sterk met elkaar samenhangen. Het gemiddelde Totaal IQ van de WISC-III$^{NL}$ ligt slechts 1.3 punt hoger dan het WISC-V-NL Totaal IQ. Op basis van het Flynn-effect is dit een onverwacht gering verschil. Op basis van het theoretisch verwachte Flynn-effect van 0.3 IQ-punten per jaar, wordt namelijk verwacht dat het ongeveer 4.5 IQ-punten zou betreffen. Zoals verwacht is de samenhang tussen de gerelateerde indexscores hoog, met gecorrigeerde correlaties tussen $r$ = .72 en $r$ = .83. Deze gegevens ondersteunen de begripsvaliditeit van de WISC-V-NL op basis van de gemeten constructen van de WISC-III$^{NL}$.

Tabel 4 Correlaties tussen de WISC-V-NL en de WISC-III$^{NL}$ (Nederland en Vlaanderen gecombineerd)

| Subtest-/ Proces-/ indexscore | WISC-V-NL | | | WISC-III$^{NL}$ | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Gem.[3] | *SD*[3] | *n* | Gem.[3] | *SD*[3] | *n* | *r*12[b] | Gecorrigeerde *r*12[b] | Standaard-verschil[c] |
| **OV** | 10.3 | 2.8 | 42 | 11.8 | 3.4 | 42 | .69 | .69 | .48 |
| **WS** | 10.3 | 2.6 | 42 | 10.6 | 3.1 | 42 | .75 | .80 | .10 |
| **BG** | 10.2 | 3.2 | 41 | 10.9 | 3.0 | 41 | .69 | .68 | .23 |
| **BP** | 10.0 | 3.4 | 43 | 10.1 | 3.1 | 43 | .78 | .77 | .03 |
| **RE** | 9.7 | 2.3 | 43 | 10.0 | 3.2 | 43 | .72 | .73 | .11 |
| **CR** | 10.1 | 3.1 | 41 | 11.0 | 3.4 | 41 | .78 | .76 | .28 |
| **SSC** | 10.8 | 3.1 | 40 | 10.9 | 3.4 | 40 | .81 | .77 | .03 |
| **SZ** | 10.6 | 3.2 | 43 | 11.6 | 3.8 | 43 | .80 | .80 | .28 |
| **VBI-VIQ** | 101.9 | 13.2 | 43 | 103.9 | 16.9 | 43 | .78 | .81 | .13 |
| **VRI-PIQ** | 100.5 | 15.3 | 42 | 102.3 | 14.7 | 42 | .74 | .72 | .12 |
| **VSI-VSI** | 103.6 | 16.0 | 41 | 105.9 | 18.0 | 41 | .82 | .80 | .14 |
| **TIQ-TIQ** | 101.7 | 13.7 | 43 | 103.0 | 17.2 | 43 | .88 | .87 | .08 |
| **VBI-VBI** | 100.9 | 11.5 | 42 | 104.0 | 15.9 | 42 | .78 | .83 | .22 |
| **VRI-POI** | 99.1 | 16.4 | 42 | 99.1 | 16.5 | 42 | .80 | .77 | .00 |

*Noot.* VIQ = WISC-III$^{NL}$ Verbaal IQ, PIQ, PIQ = WISC-III$^{NL}$ Performaal IQ, PIQ - WISC-III$^{NL}$ Perceptuele Organisatie Index. Correlaties werden berekend voor beide afnamevolgorden volgens een 'counterbalanced' onderzoeksopzet en gecorrigeerd voor de variabiliteit binnen de WISC-V-NL normeringssteekproef (Guilford & Fruchter, 1978).

[a] De vermelde waarden in de Gemiddelde en SD-kolommen betreffen gemiddelde waarden op basis van beide afnamevolgorden.

[b] Gemiddelde correlaties over beide afnamevolgorden zijn berekend met behulp van Fisher's z-transformatie.

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

### WPPSI-III-NL

De WPPSI-III-NL (Wechsler, 2009) is de Nederlandse bewerking van de *Wechsler Preschool and Primary Scale of Intelligence – Third Edition* (Wechsler, 2002). De test is geschikt voor kinderen van 2.5 tot 8 jaar en meet het algemene intelligentieniveau.

Het validiteitsonderzoek is uitgevoerd bij 49 kinderen (25 jongens en 24 meisjes). De gemiddelde leeftijd van de kinderen bij de eerste afname van de WPPSI-III-NL of WISC-V-NL was 6.8 jaar (SD 0.5). Het tijdsinterval tussen beide afnames was gemiddeld 36 dagen, met een range van 2 tot 63 dagen. Verwacht wordt dat er over het geheel genomen hoge tot middelmatige correlaties worden gevonden tussen subtests en indexen van beide tests die inhoudelijk en qua meet-domeinen gerelateerd zijn. Zoals te zien is in Tabel 5 liggen de gemiddelde scores van het Totaal IQ en de indexen alle in de gemiddelde range. Alle gemiddelde scores op de indexen zijn lager voor de WISC-V-NL dan voor de WPPSI-III-NL, hetgeen in overeenstemming is met de bevindingen in de Amerikaanse correlatiestudie. De effectgroottes zijn in alle gevallen klein. De gecorrigeerde correlatie van het Totaal IQ van de WPPSI-III-NL met het Totaal IQ van de WISC-V-NL is hoog: $r = .78$. Het gemiddelde Totaal IQ van de WPPSI-III-NL ligt meer dan 4 punten hoger dan het Totaal IQ van de WISC-V-NL. Een mogelijke verklaring hiervoor kan zijn dat bij 62% van de kinderen eerst de WPPSI-III-NL werd afgenomen en bij 38% van de kinderen eerst de WISC-V-NL. Er kan dus sprake zijn van een leereffect dat zichtbaar wordt in het hogere resultaat op de WPPSI-III-NL.
Zoals verwacht is de samenhang tussen de gerelateerde indexscore middelmatig tot hoog (tussen $r = .67$ en $r = .78$). Deze gegevens laten een goede overeenstemming tussen beide instrumenten zien en zijn in lijn met eerder gerapporteerde resultaten in het Amerikaanse

correlatie-onderzoek naar de WISC-V. Hiermee wordt ondersteuning gevonden voor de begripsvaliditeit van de WISC-V-NL.

Tabel 5 Correlaties tussen de WISC-V-NL en de WISC-III-NL (Nederland en Vlaanderen gecombineerd)

| Subtest-/ Proces-/ indexscore | WISC-V-NL | | | WPPSI-III-NL | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Gem.[3] | *SD*[3] | *n* | Gem.[3] | *SD*[3] | *n* | *r*12[b] | Gecorrigeerde *r*12[b] | Standaard-verschil[c] |
| **OV** | 10.2 | 2.4 | 45 | 11.1 | 2.2 | 45 | .56 | .62 | .39 |
| **WS** | 9.5 | 2.5 | 41 | 10.4 | 2.6 | 41 | .64 | .73 | .35 |
| **BG** | 9.7 | 2.8 | 42 | 9.6 | 2.4 | 42 | .43 | .47 | .04 |
| **BP** | 10.2 | 2.8 | 49 | 11.4 | 2.8 | 49 | .77 | .79 | .43 |
| **MR** | 10.7 | 3.2 | 45 | 12.5 | 2.7 | 45 | .59 | .63 | .61 |
| **SSC** | 10.7 | 3.4 | 48 | 10.7 | 3.1 | 48 | .49 | .43 | .00 |
| **SZ** | 10.4 | 2.9 | 48 | 10.8 | 3.0 | 48 | .57 | .59 | .14 |
| **VBI-VIQ** | 99.6 | 10.8 | 43 | 101.6 | 12.2 | 43 | .59 | .67 | .17 |
| **VRI-PIQ** | 102.7 | 12.2 | 45 | 108.1 | 13.7 | 45 | .70 | .76 | .42 |
| **FRI-PIQ** | 103.4 | 15.1 | 49 | 110.0 | 14.9 | 49 | .69 | .69 | .44 |
| **VSI-VS** | 102.6 | 14.6 | 43 | 105.5 | 12.3 | 43 | .73 | .77 | .21 |
| **TIQ** | 102.1 | 12.7 | 45 | 106.2 | 12.5 | 45 | .72 | .78 | .33 |

*Noot.* VIQ = WPPSO-III-NL Verbaal IQ, PIQ, PIQ = WPPSI-III-NL Performaal IQ, VS = WISC-III-NL Verwerkingssnelheid. Correlaties werden berekend voor beide afnamevolgorden volgens een 'counterbalanced' onderzoeksopzet en gecorrigeerd voor de variabiliteit binnen de WISC-V-NL normeringssteekproef (Guilford & Fruchter, 1978).

[a] De vermelde waarden in de Gemiddelde en SD-kolommen betreffen gemiddelde waarden op basis van beide afnamevolgorden.

[b] Gemiddelde correlaties over beide afnamevolgorden zijn berekend met behulp van Fisher's z-transformatie.

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

### WAIS-IV-NL

Ten slotte is er een vergelijkend onderzoek gedaan met de WAIS-IV-NL. De WAIS-IV-NL (Wechsler, 2012a) is de Nederlandse bewerking van de *Wechsler Adult Intelligence Scales – Fourth Edition* (Wechsler, 2008). De test is geschikt voor personen tussen 16 en 85 jaar en meet het algemene intelligentieniveau.

Het onderzoek is uitgevoerd bij 39 kinderen (16 jongens en 23 meisjes). De gemiddelde leeftijd van de kinderen bij de afname van de WAIS-IV-NL was 16.5 jaar (*SD* 0.3). Het tijdsinterval tussen beide afnames was gemiddeld 36 dagen, met een range van 7 tot 127 dagen.
Zoals te zien is in Tabel 6 bevinden de gemiddelde scores voor het Totaal IQ en de indexen zich alle in de gemiddelde range. De gecorrigeerde correlatie tussen het Totaal IQ van de WAIS-IV-NL met het Totaal IQ van de WISC-V-NL is hoog, namelijk $r = .89$. De gemiddelde Totaal IQ-scores blijken nauwelijks van elkaar te verschillen (1.0 IQ-punt), met een verwaarloosbare effectgrootte van .06.
Zoals verwacht is de samenhang tussen de gerelateerde indexscores ook middelmatig tot hoog, tussen $r = .63$ en $r = .89$. Deze gegevens laten opnieuw een goede overeenstemming tussen beide instrumenten zien en zijn wederom in lijn met eerder gerapporteerde resultaten in het Amerikaanse correlatie-onderzoek naar de WISC-V (Wechsler, 2014). Hiermee wordt ondersteuning gevonden voor de begripsvaliditeit van de WISC-V-NL.

Tabel 6 Correlaties tussen de WISC-V-NL en de WAIS-IV-NL (Nederland en Vlaanderen gecombineerd)

| Subtest-/ Proces-/ indexscore | WISC-V-NL | | | WAIS-IV-NL | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Gem.[3] | *SD*[3] | *n* | Gem.[3] | *SD*[3] | *n* | *r*12[b] | Gecorrigeerde *r*12[b] | Standaard-verschil[c] |
| **OV** | 10.0 | 2.7 | 36 | 10.6 | 2.8 | 36 | .76 | .78 | .22 |
| **WS** | 10.8 | 3.0 | 36 | 9.8 | 2.5 | 36 | .70 | .73 | .36 |
| **BG** | 11.2 | 2.9 | 31 | 10.9 | 3.1 | 31 | .61 | .69 | .10 |
| **BP** | 11.0 | 2.6 | 36 | 11.4 | 3.4 | 36 | .73 | .74 | .13 |
| **FS** | 10.4 | 3.6 | 39 | 10.5 | 3.1 | 39 | .72 | .69 | .03 |
| **MR** | 10.0 | 3.6 | 36 | 9.9 | 2.8 | 36 | .58 | .56 | .03 |
| **GW** | 10.9 | 2.9 | 34 | 10.8 | 3.6 | 34 | .72 | .73 | .03 |
| **RE** | 10.0 | 2.5 | 38 | 9.7 | 3.4 | 38 | .75 | .77 | .10 |
| **CR** | 10.4 | 2.2 | 39 | 11.2 | 2.6 | 39 | .73 | .80 | .33 |
| **CLN** | 10.2 | 3.1 | 34 | 10.2 | 3.3 | 34 | .64 | .63 | .00 |
| **SSC** | 11.3 | 2.2 | 37 | 10.5 | 2.7 | 37 | .65 | .73 | .33 |
| **SZ** | 11.6 | 3.2 | 37 | 10.9 | 3.4 | 37 | .83 | .81 | .21 |
| **FZ** | 9.4 | 3.1 | 33 | 10.8 | 3.2 | 33 | .32 | .37 | .44 |
| **VBI-VBI** | 101.9 | 13.9 | 34 | 102.3 | 13.5 | 36 | .71 | .74 | .03 |
| **VRI-PRI** | 103.1 | 16.2 | 39 | 102.6 | 16.2 | 39 | .84 | .83 | .03 |
| **FRI-PRI** | 102.6 | 15.1 | 38 | 103.2 | 15.8 | 38 | .69 | .69 | .04 |
| **Wgl-Wgl** | 100.3 | 10.9 | 39 | 102.5 | 14.0 | 39 | .59 | .63 | .18 |
| **Vsl-Vsl** | 107.4 | 11.0 | 36 | 103.7 | 14.8 | 36 | .81 | .87 | .28 |
| **TIQ** | 104.5 | 11.9 | 35 | 103.8 | 13.3 | 35 | .85 | .89 | .06 |

*Noot.* PRI = WAIS-IV-NL Perceptueel Redeneren Index. Correlaties werden berekend voor beide afnamevolgorden volgens een 'counterbalanced' onderzoeksopzet en gecorrigeerd voor de variabiliteit binnen de WISC-V-NL normeringssteekproef (Guilford & Fruchter, 1978).

[a] De vermelde waarden in de Gemiddelde en SD-kolommen betreffen gemiddelde waarden op basis van beide afnamevolgorden.

[b] Gemiddelde correlaties over beide afnamevolgorden zijn berekend met behulp van Fisher's z-transformatie.

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

## Criteriumvaliditeit

Om de (gelijktijdige) criteriumvaliditeit van de WISC-V-NL te kunnen vaststellen is onderzocht hoe goed de WISC-V-NL kan discrimineren tussen een controlegroep en twee groepen waarvan respectievelijk een lage score (een groep met verstandelijk beperkten, N = 58) en een hoge score (hoogbegaafden, N = 27) verwacht wordt. Concluderend kan gesteld worden dat het onderscheidend vermogen van de WISC-V-NL voor kinderen met een verstandelijke beperking uitstekend is en voor hoogbegaafde kinderen goed. Wel moet worden opgemerkt dat de hier gepresenteerde resultaten slechts bewijs bieden voor *gelijktijdige* criteriumvaliditeit. Een onderzoeksopzet waarbij de *voorspellende* kwaliteiten van de WISC-V-NL gerelateerd worden aan bijvoorbeeld opleidingsniveau en studiesucces in de toekomst zou de criteriumvaliditeit van de WISC-V-NL verder moeten onderbouwen.

# 6 Referenties


Allen, M. J., & Yen, W. M. (2002). *Introduction to measurement theory*. Long Grove, IL: Waveland Press.

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. doi: 10.1037/a0017767

Bodin, D., Pardini, D. A., Burns, T. G., & Stevens, A. B. (2009). Higher order factor structure of the WISC-IV in a clinical neuropsychological sample. *Child Neuropsychology, 15*, 417-424. doi: 10.1080/09297040802603661.

Bowden, S. C., Weiss, L. G., Holdnack, J. A., & Lloyd, D. (2006). Age-related invariance of abilities measured with the Wechsler Adult Intelligence Scale-III. *Psychological Assessment, 18*(3), 334-339.

Centraal Bureau voor de Statistiek (CBS). (2015). Retrieved March 2015 from http://www.cbs.nl.

Chen, H., Keith, T. Z., Weiss, L., Zhu, J., & Li, Y. (2010). Testing for multigroup invariance of second-order WISC-IV structure across China, Hong Kong, Macau, and Taiwan. *Personality and Individual Differences, 49*, 677-682. doi: 10.1016/j.paid.2010.06.004.

Chen, H., Zhang, O., Engi Raiford, S., Zhu, J., & Weiss, L. G. (2015). Factor invariance between genders on the Wechsler Intelligence Scale for Children – Fifth Edition. *Personality and Individual Differences, 86*(2015), 1-5.

Cohen, B. H. (1996). *Explaining psychological statistics*.Pacific Grove, CA: Brooks & Cole.

Evers, A., Lucassen, W., Meijer, R., & Sijtsma, K. (2010). *COTAN beoordelingssysteem voor de kwaliteit van tests.* NIP: Amsterdam.

Guilford, J. P. (1954). *Psychometric methods* (2nd ed.). New York: McGraw-Hill.

Guttman, L. (1945). A basis for analyzing test-retest reliability. *Psychometrika, 10*(4), 255-282.

Haertel, E. H. (2006). Reliability. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 65-110). Westport, CT: Praeger.

Joreskog, K. G., & Sorbom, D. (1993). *LISREL 8: User's reference guide*. Chicago, IL: Scientific Software International.

Kaufman, A. S., Lichtenberger, E. O., & McLean, J. (2001). Two- and three-factor solutions of the WAIS-III. *Assessment, 8*(3), 267-280.

Keith, T. Z., Fine, J. G., Taub, G. E., Reynolds, M. R., & Kranzler, J. H. (2006). Higher order, multi-sample, confirmatory factor analysis of the Wechsler Intelligence Scale for Children – Fourth Edition: What does it measure? *School Psychology Review, 35*(1), 108-127.

Magnusson, D. (1967). *Test theory.* Reading, MA: Addison-Wesley.

Nunnally, J., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York: McGraw-Hill.

Silver, N. C., & Dunlap, W. P. (1987). Averaging correlation coefficients: Should Fishers z transformation be used? *Journal of Applied Psychology, 72*(1), 146-148.

Statistics Belgium (Statbel) (2015). Retrieved March 2015 from http://statbel.fgov.be/nl/statistieken/Cijfers.

Strube, M. J. (1988). Some comments of the use of magnitude-of-effect estimates. *Journal of Counseling Psychology, 35*(3), 342-345.

Vlaamse Ministerie van Onderwijs en Vorming (2015). *Onderwijsstatistieken schooljaar 2014-2015*. Retrieved March 2015 from https://onderwijs.vlaanderen.be/nl/onderwijsstatistieken.

Ward, L. C., Bergman, M. A., & Hebert, K. R. (2011). WAIS-IV subtest covariance structure: Conceptual and statistical considerations. *Psychological Assessment, 24*(2), 328-340. doi: 10.1037/a0025614.

Ward, L. C., Ryan, J. J., & Axelrod, B. N. (2000). Confirmatory factor analyses of the WAIS-III standardization data. *Psychological Assessment, 12*(3), 341-345.

Watkins, M. W. (2010). Structure of the Wechsler Intelligence Scale for Children – Fourth Edition among a national sample of referred students. *Psychological Assessment, 22*(4), 782-787.

Watkins, M. W., Wilson, S. M., Kotz, K. M., Carbone, M. C., & Babula, T. (2006). Factor structure of the Wechsler Intelligence Scale for Children – Fourth Edition among referred students. *Educational and Psychological Measurement, 66*, 975-983. doi: 10.1177/0013164406288168.

Wechsler, D. (1949). *Wechsler Intelligence Scale for Children*. New York, NY: The Psychological Corporation.

Wechsler, D. (1967). *Wechsler Preschool and Primary Scale of Intelligence*. New York, NY: The psychological Corporation.

Wechsler, D. (1991). *Wechsler Intelligence Scale for Children* (3rd ed.). San Antonio, TX: The Psychological Corporation.

Wechsler, D. (2002). *Wechsler Preschool and Primary Scale of Intelligence* (3rd ed.). San Antonio, TX: Pearson.

Wechsler, D. (2003a). *Wechsler Intelligence Scale for Children* (4th ed.). San Antonio, TX: Pearson.

Wechsler, D. (2003b). *Wechsler Intelligence Scale for Children, derde editie; Nederlandstalige bewerking.* Amsterdam: Pearson Benelux B.V.

Wechsler, D. (2008). *Wechsler Adult Intelligence Scale* (4th ed.). Bloomington, MN: Pearson.

Wechsler, D. (2009). *Wechsler Preschool and Primary Scale of Intelligence, derde editie; Nederlandstalige bewerking.* Amsterdam: Pearson Benelux B.V.

Wechsler, D. (2012a). *Wechsler Adult Intelligence Scale, vierde editie; Nederlandstalige bewerking.* Amsterdam: Pearson Benelux B.V.

Wechsler, D. (2012b). *Wechsler Preschool and Primary Scale of Intelligence* (4th ed.). Bloomington, MN: Pearson.

Wechsler, D. (2014). *Wechsler Intelligence Scale for Children* (5th ed.); WISC-V. Bloomington, MN: Pearson.

Weiss, L. G., Keith, T. Z., Zhu, J., & Chen, H. (2013a). WAIS-IV clinical validation of the four- and fivefactor interpretive approaches [Special edition]. *Journal of Psychoeducational Assessment, 31*(2), 94-113. doi: 10.1177/0734282913478030.

Weiss, L. G., Keith, T. Z., Zhu, J., & Chen, H. (2013b). WISC-IV and clinical validation of the four- and five-factor interpretive approaches [Special edition]. *Journal of Psychoeducational Assessment, 31*(2), 114-131. doi: 10.1177/0734282913478032.

**Pearson Benelux B.V.**
Gatwickstraat 1
1043 GK Amsterdam

t: +31 (0)20 581 5500
e: info-nl@pearson.com

www.pearsonclinical.nl
www.pearsonclinical.be

Pearson