

Na veel kritiek op de validiteit van de in 2000 verschenen Nederlandstalige bewerking van de WAIS-III, werden vorig jaar nieuwe normen en een technische handleiding gepubliceerd. In onderstaande bijdrage

worden deze tegen het licht gehouden. De auteurs concluderen dat we omzichtig moeten zijn met de

*Don van Ravenzwaaij en
Ronald Hamel*

Forum

De Nederlandstalige WAIS-III na hernormering

De WAIS (Wechsler Adult Intelligence Scale) is een van de meest vooraanstaande intelligentietests. In de oorspronkelijke taal zijn er drie versies van deze test op de markt gekomen, de WAIS, de WAIS-R en de WAIS-III (Wechsler, 1997). De eerste Nederlandstalige versie van de WAIS verscheen in 1970. De tweede versie, de WAIS-R, is nooit in het Nederlands vertaald. De WAIS-III is de meest recente versie en verscheen in 2000 (Wechsler, 2000). Uit divers onderzoek blijkt dat de nieuwste versie van de oorspronkelijke test, de WAIS-III, zeer goed bruikbaar is als diagnostisch instrument. De betrouwbaarheidsgegevens zijn uitstekend (Wechsler, 1997; Zhu & Tulskey, 2000) en ook de validiteit is ruim voldoende (Saklofske, Hildebrand & Gorsuch, 2000; Ryan & Paolo, 2001; Van der Heijden & Donders, 2003). Met de Nederlandse versie bleek het een stuk minder goed gesteld (zie onder anderen Tellegen, 2003; Van der Laan & Oswald, 2001; Derksen & Katzko, 2002). Harcourt Test Publishers onderkende het probleem en heeft begin 2005 nieuwe steekproefgegevens gepubliceerd in een technisch rapport hernormering (Kooij, Rolffhus, Wilkins, Yang & Zhu, 2005). In oktober 2005 is een nieuwe technische handleiding uitgegeven (Wechsler, 2005).

De onderhavige bijdrage onderzoekt de validiteit van de Nederlandstalige bewerking van de WAIS-III als diagnostisch instrument na hernormering. Hierbij worden de gerapporteerde veranderingen gelegd naast wat onveranderd is gebleven. Voor de evaluatie van de validiteit wordt gekeken naar:

- de factorstructuur;
- de subtest-intercorrelaties;
- de convergente validiteit;
- de criteriumvaliditeit;
- de kwaliteit van de testnormen.

In de eerste paragraaf komen de eerste vier punten aan bod onder het kopje 'Validiteit'. In de tweede paragraaf worden de testnormen behandeld. Hierna volgt een conclusie over de validiteit van de Nederlandse bewerking van de WAIS-III. Ter afsluiting komen de maatschappelijke implicaties van dit alles voor het gebruik van de WAIS-III aan de orde.

interpretatie van resultaten van de Nederlandse WAIS-III, vooral omdat de test niet voldoet aan de criteria die gelden voor gebruik in diagnostische settings.

Validiteit

Het onderzoek naar de factorstructuur van de WAIS-III is uitgevoerd met zowel exploratieve als confirmatieve factoranalyse. De hernormering heeft een verwaarloosbare invloed gehad op de waarden die uit

de beide factoranalyses zijn gekomen. Bij de confirmatieve factoranalyse werden vier modellen getest, waarbij de beste fit werd bereikt met een vierfactormodel met als factoren: Verbaal Begrip, Perceptuele Organisatie, Werkgeheugen en Verwerkingssnelheid (Wechsler, 2000; Wechsler, 2005). Ook exploratieve factoranalyse gaf een vierfactormodel, met als kanttekening dat de subtest Rekenen (analoog aan de Amerikaanse subtest Arithmetic) niet duidelijk op één van de vier voorgestelde factoren laadt. De subtest laadt evenredig op de factoren Verbaal Begrip (0,31), Perceptuele Organisatie (0,30) en Werkgeheugen (0,26). Dit probleem is verre van triviaal, stelt Tellegen (2003), aangezien als gevolg hiervan twee van de vier veronderstelde factoren gerepresenteerd worden door slechts twee subtests. Werkgeheugen bestaat alleen uit Cijferreeksen en Cijfers en Letters Nazeggen, Verwerkingssnelheid is de combinatie van Symbool Zoeken en Symbool Substitutie-Coderen. Dit is, volgens Tellegen, te weinig om een factor mee te definiëren. Is die stelling echter wel houdbaar? Een factor die gebaseerd is op twee *individuele itemscores*, is slecht gedefinieerd. Indien de betrouwbaarheid van een subtest echter voldoende is en deze bevat voldoende items, is zelfs één subtest voldoende om een factor mee te definiëren. Wel lijkt het wenselijk om Rekenen niet alleen als indicator voor de factor Werkgeheugen te hanteren, maar ook voor de factoren Verbaal Begrip en Perceptuele Organisatie.

De subtestintercorrelaties variëren na hernormering van .30 tot .79 (Wechsler, 2005). Over het geheel genomen is dit redelijk. Wel is het jammer dat het gevonden correlatiepatroon noch in de handleiding, noch in het technisch rapport hernormering besproken wordt; de veronderstellingen worden dus niet geëvalueerd. Concluderend kan gesteld worden dat het met de subtestintercorrelaties wel goed zit,

maar dat er het een en ander schort aan de rapportage en evaluatie van de onderzoeksgegevens.

Er worden in de handleiding drie correlaties met andere tests besproken. Enerzijds wordt de Nederlandse WAIS-III gecorreleerd met de Nederlandstalige (oorspronkelijke) WAIS. Deze correlatie wordt vergeleken met de correlatie tussen de Amerikaanse WAIS-III en de WAIS-R. Anderzijds wordt gekeken naar de samenhang van de WAIS-III met twee andere tests: de KAIT (Kaufman Adolescent and Adult Intelligence Test; Kaufman & Kaufman, 1993) en de K-SNAP (Kaufman Short Neuropsychological Assessment Procedure; Kaufman & Kaufman, 1994). De KAIT is een intelligentietest, de K-SNAP meet cognitief functioneren. Beide mogen dus geacht worden hoog te correleren met de WAIS-III.

Op de subtests Onvolledige Tekeningen, Plaatjes Ordenen en Figuur Leggen na, is er een aanzienlijk verband tussen de oorspronkelijke versies en de nieuwe versies. De reden dat de versies van deze drie subtests zo'n lage correlatie vertonen, zou gelegen zijn in het feit dat beide bij de ontwikkeling van de WAIS-III rigoureus herzien zijn (Wechsler, 2000). Het feit dat de Amerikaanse correlaties hoger liggen, is te wijten aan het feit dat de WAIS-R in Nederland nooit is verschenen. Er zit dus een groter tijdsinterval tussen de twee Nederlandse versies dan tussen de laatste twee Amerikaanse versies (Wechsler, 2000). Bovendien zouden de WAIS en de WAIS-R te veel van elkaar verschillen en daarom niet op een dergelijke manier naast elkaar gezet kunnen worden. Dit neemt echter niet weg dat de correlatie tussen Onvolledige Tekeningen van de WAIS-III en die van de oorspronkelijke WAIS .33 bedraagt. Het feit dat de validiteit van de oorspronkelijke WAIS-III onvoldoende was, zou een verklaring kunnen zijn voor deze lage correlatie. Desalniettemin is deze waarde laag, het is tenslotte dezelfde subtest met dezelfde meetpretenties.

De correlaties tussen de WAIS-III enerzijds en de KAIT en de K-SNAP anderzijds pakken eveneens ongunstig uit. De correlaties tussen de diverse schalen van de WAIS-III en de KAIT variëren van .24 tot .51 (tussen de WAIS-III en de Amerikaanse KAIT .73 en .84). De auteurs wijten dit verschil aan het feit dat de Nederlandstalige KAIT nog experimenteel was, maar het betreft hier twee bewerkingen, dus het zou ook kunnen liggen aan de Nederlandstalige WAIS-III. De correlaties tussen de diverse schalen van de WAIS-III en de K-SNAP liggen tussen de .38 en de .52 (tussen de WAIS-III en de Amerikaanse K-SNAP .58 en .67). De WAIS-III, de KAIT en de K-SNAP zijn verschillende tests met een verschillende meetpretentie, zodat een perfecte correlatie niet verwacht kan worden. Het grote verschil in correlaties tussen de Nederlandse en Amerikaanse versies is evenwel opmerkelijk. De convergente validiteit van de WAIS-III is in Nederland een stuk slechter dan in Amerika en dat zou niet het geval moeten zijn. Ook is het jammer dat er geen gegevens over de divergente validiteit worden gerapporteerd.

De gegevens van de criteriumvaliditeit van de Nederlandse versie van de WAIS-III worden in de handleiding volledig gebaseerd op Amerikaans onderzoek, op één studie na

(over 28 volwassenen die gediagnosticeerd zijn als mentaal geretardeerd). Aangezien het voor het aantonen van de criteriumvaliditeit niet volstaat om de gegevens van een andere versie over te nemen, voldoet dit niet aan de kwaliteitsnormen van de Cotan (1999). Weinig verrassend is het oordeel van de Cotan (2002) over de criteriumvaliditeit van de WAIS-III dan ook 'onvoldoende'.

De begripsvaliditeit wordt volgens de Cotan (2002) als voldoende beoordeeld, maar Tellegen velt terecht het oordeel 'onvoldoende'. De factorstructuur verdient verdere verfijning, aangezien Rekenen niet duidelijk onder één factor valt. De convergente validiteit is slecht en onderzoek naar de divergente validiteit ontbreekt volledig.

Normen

Over de normen van de WAIS-III is een hoop te doen geweest. In de handleiding worden de volgende statistieken van de gekozen steekproef gerapporteerd (Wechsler, 2000):

- de steekproef bestond uit 745 individuen;
- er waren zeven leeftijdsgroepen, die als volgt ingedeeld waren: 16-20 jaar (12%), 21-25 jaar (13%), 26-35 jaar (19%), 36-50 jaar (15%), 51-65 jaar (18%), 66-75 jaar (13%) en 76-85 jaar (10%);
- de steekproef bestond uit 56% vrouwen en 44% mannen.

Verder is geprobeerd om zowel voor opleidingsniveau als geografische spreiding een zo correct mogelijke verhouding aan te houden.

Bij vergelijking van de testresultaten op de WAIS-III met resultaten op de oorspronkelijke WAIS komt een aantal merkwaardigheden naar boven. Zo worden er voor drie individuele proefpersonen grote verschillen gerapporteerd tussen scores op de WAIS-III en de oude WAIS (Van der Laan & Oswald, 2001). De handleiding verdedigt zichzelf op voorhand door te stellen dat de verschillen tussen de scores op beide tests worden veroorzaakt door het Flynn-effect (Wechsler, 2000). Een alternatieve verklaring is dat bovenstaande verschillen niet alleen worden veroorzaakt door het Flynn-effect, maar ook doordat hoogopgeleiden sterk oververtegenwoordigd zouden zijn in de normgroep van de WAIS-III (Tellegen, 2002). Bij een dergelijke oververtegenwoordiging is het weinig verwonderlijk dat de standaard voor de WAIS-III een stuk hoger is komen te liggen en dit zou betekenen dat de cognitieve vaardigheden van individuen bij wie de test wordt afgenomen, structureel te laag worden ingeschat. Derksen en Katzko (2002) stellen dat de diverse opleidingsniveaus op zich redelijk goed gerepresenteerd zijn, maar dat de steekproef eenvoudigweg veel te klein is, waardoor het weinig verrassend is dat het scorepatroon geen goede afspiegeling van de Nederlandse en Vlaamse bevolking geeft.

Hoe zit het nu met de nieuwe steekproefgegevens? Het probleem van de oververtegenwoordiging van hoogopgeleiden lijkt ondervangen, de steekproef is echter nog kleiner geworden: de steekproef bestaat nu uit slechts 588 individuen. Bovendien zijn er nu tien leeftijdsgroepen in plaats van

zeven. Voor berekening van de betrouwbaarheden zijn de leeftijdsgroepen 16-17 en 18-19 samengenomen, aangezien de groepen apart niet groot genoeg waren voor fatsoenlijke analyse (Wechsler, 2005). Zelfs gecombineerd levert dit echter een groep op van slechts 40 proefpersonen, hetgeen een breed 95% betrouwbaarheidsinterval geeft. Dit wordt in de handleiding verantwoord door te wijzen op de gehanteerde regressiemethode van continue normering (Zachary & Gorsuch, 1985). Deze methode is echter niet bedoeld als panacee voor kleine of verkleinde steekproeven. Zachary en Gorsuch (1985) beweren ook helemaal niet dat deze methode kan compenseren voor het gebruik van kleine steekproeven. Continue normering is daarentegen bedoeld om 'rare sprongen' van het IQ als gevolg van kleine verschuivingen in leeftijd glad te strijken ('smoothing'). Continu normeren is een methode om de vergelijking van iemands ruwe score met zijn of haar normgroep niet te laten beïnvloeden door het beperkte aantal groepen. Het is *niet* bedoeld als oplossing voor geringe omvang van die groepen. De oorspronkelijke Nederlandstalige WAIS had een steekproef van 2169 individuen. Het lijkt wenselijk om voor de WAIS-III een steekproef van vergelijkbare grootte te gebruiken.

Conclusie

De validiteit van de WAIS-III is niet goed. Weliswaar vindt het voorgestelde vierfactormodel steun, maar de subtest Rekenen vertroebelt het beeld, aangezien deze niet duidelijk op één factor laadt. Verder is de convergente validiteit laag, ontbreekt onderzoek naar de divergente validiteit in de handleiding van de WAIS-III en ontbreken gegevens over de criteriumvaliditeit van de *Nederlandse* versie van de WAIS-III. Op grond van het onderzoek naar de normering van de WAIS-III lijkt de nieuwe steekproef meer noodzaak dan luxe. Op het eerste gezicht lijkt de nieuwe steekproef te klein. Het blijft afwachten hoe het oordeel van de Cotan uitpakt.

Met een slechte validiteit en te summere steekproefgegevens zal het weinig verrassend zijn dat het gebruik van de huidige versie van de Nederlandse WAIS-III riskant is en dat resultaten met de grootste zorgvuldigheid geïnterpreteerd dienen te worden. Aangezien bepaalde instanties, zoals de Zwakzinnigenzorg of de Sociaal Pedagogische Dienstverlening, als indicatie voor hulpverlening vaak een IQ-criterium hanteren (in dit geval een IQ van maximaal 70) is het van vitaal belang dat een intelligentietest de ware score van een individu zo dicht mogelijk kan benaderen (Van der Laan & Oswald, 2001). De WAIS-III voldoet op dit moment nog onvoldoende aan deze eis. De ingevoerde hernormeringen zouden een belangrijke stap in de richting kunnen zijn van een betere test, een grotere steekproef is echter wenselijk.

Verder lijkt ook een betere uitwerking van de voorgestelde factorstructuur niet overbodig, bijvoorbeeld door de subtest Rekenen als indicator te gebruiken voor zowel Verbaal Begrip, Perceptuele Organisatie en Werkgeheugen. Een andere noodzakelijke verbetering zou het toevoegen van deugdelijke onderzoeksresultaten zijn ten behoeve

van de predictieve waarde op het gebied van academische vaardigheden, arbeidssucces en de diagnosticering van diverse klinische syndromen. Dit is nodig om de WAIS-III met enig vertrouwen te kunnen gebruiken bij personeelsselectie, schoolkeuze en dergelijke. Voorts zou het nuttig zijn onderzoek te verrichten naar het gebruik van de WAIS-III bij allochtone groepen. Hoewel dit in de oorspronkelijke handleiding (Wechsler, 2000) aangekondigd werd, blijkt in de nieuwe technische handleiding (Wechsler, 2005) hiervan nog geen sprake.

Een laatste aanpassing die verre van overbodig lijkt, betreft een nauwkeurige (her)evaluatie van de handleiding. De toekomst moet uitwijzen of de Nederlandse versie van de WAIS-III aan de genoemde criteria kan voldoen. Het is in de regel niet wenselijk om IQ-scores die verkregen zijn met een intelligentietest absoluut te nemen en zeer omzichtig te zijn bij gebruik in diagnostiek, personeelsselectie en dergelijke. In de praktijk dienen testcores door psychologen altijd geïnterpreteerd te worden met behulp van een betrouwbaarheidsinterval. Helaas maken sommige instanties slechts gebruik van de gerapporteerde IQ-score. Op grond van het bovenstaande raden we psychologen aan bij de interpretatie van pure IQ-scores extra zorgvuldigheid te betrachten zolang de WAIS-III nog niet grondig is herzien.

D. van Ravenzwaaij, BSc, en dr. R. Hamel zijn beiden verbonden aan de Afdeling Psychologie van de Universiteit van Amsterdam.
Correspondentieadres: <D.vanRavenzwaaij@uva.nl>.

Noot

De auteurs danken Harry Vorst en Jelte Wicherts voor hun commentaar op een eerdere versie van het manuscript.

Literatuur

- Cotan (1999). *Beoordelingssysteem voor de kwaliteit van tests*. Amsterdam: NIP.
- Cotan (2002). *Cotan-beoordeling WAIS-III*. <http://www.testresearch.nl/wais/cotan.html>.
- Derksen, J.J.L. & Katzko, M. (2002). WAIS-III: enkele kritische notities. *De Psycholoog*, 37, 460-462.
- Heijden, P. van der & Donders, J. (2003). A confirmatory factor analysis of the WAIS-III in patients with traumatic brain injury. *Journal of Clinical and Experimental Neuropsychology*, 25, 59-65.
- Kaufman, A.S. & Kaufman, N.L., (1993). *Kaufman Adolescent & Adult Intelligence Test*. Circle Pines: American Guidance Service, Inc.
- Kaufman, A.S. & Kaufman, N.L., (1994). *Kaufman Short Neuropsychological Assessment Procedure*. Circle Pines: American Guidance Service, Inc.
- Kooij, A.P., Rolfhus, E., Wilkins, C., Yang, Z. & Zhu, J. (2005). *Technisch rapport hernormering*. Lisse: Harcourt Test Publishers.
- Laan, E.C. van der & Oswald, H.L. (2001). WAIS-III in discussie. *De Psycholoog*, 36, 677-678.
- Ryan, J.J. & Paolo, A.M. (2001). Exploratory factor analysis of the WAIS-III in a mixed patient sample. *Archives of Clinical Neuropsychology*, 16, 151-156.
- Saklofske, D.H., Hildebrand, D.K. & Gorsuch, R.L. (2000). Replication of the factor structure of the Wechsler Adult Intelligence Scale – Third Edition with a Canadian sample. *Psychological Assessment*, 12, 436-439.
- Tellegen, P.J. (2002). De kwaliteit van de normen van de WAIS-III. *De Psycholoog*, 37, 463-465.
- Tellegen, P.J. (2003). De betrouwbaarheid en validiteit van de WAIS-III NL. *De Psycholoog*, 38, 128-132.

- Wechsler, D. (1997). *WAIS-III administration and scoring manual*. San Antonio, TX: The Psychological Corporation.
- Wechsler, D. (2000). *WAIS-III Nederlandstalige bewerking. Technische handleiding*. Lisse: Swets & Zeitlinger.
- Wechsler, D. (2005). *WAIS-III Nederlandstalige bewerking. Technische handleiding*. Lisse: Harcourt Test Publishers.
- Zachary, R. A. & Gorsuch, R. L. (1985). Continuous norming: Implications for the WAIS-R. *Journal of Clinical Psychology, 41*, 86-94.
- Zhu, J. & Tulskey, D.S. (2000). Co-norming the WAIS-III and WMS-III: is there a test-order effect on IQ and memory scores? *The Clinical Neuropsychologist, 14*, 461-467.

Summary

On the Dutch WAIS-III after re-norming

D. van Ravenzwaaij, R. Hamel

In this paper, the validity of the Dutch version of the WAIS-III after re-norming is assessed. The structure of the factorial design, the subtest-intercorrelations, the convergent validity, the criterion validity and the new test norms are used as indicators. It is concluded that the validity of the Dutch version of the WAIS-III is insufficient. It does not meet the criteria for application in diagnostic settings.

Reactie op de bijdrage van Shirin Slabbers, april 2006

In de nieuwe rubriek 'juridisch nieuws' in het aprilnummer 2006 van *De Psycholoog* behandelt gezondheidsrechtjuriste Shirin Slabbers de recente verlenging van de bewaartermijn voor dossiers in de WGBO tot vijftien jaar. Met recht een belangrijk nieuwtje! In de beroepsethische adviespraktijk van het NIP komen veel vragen binnen over het dossier, zoals de samenstelling en bewaartermijn.

In het artikel spreekt de auteur de mening uit dat de bepaling over de bewaartermijn in de *Beroepscode 1998* van het NIP 'niet altijd juist' zou zijn, een opvatting die niet wordt toegelicht of onderbouwd. Om onduidelijkheid te voorkomen, wil ik er graag op wijzen dat de betreffende bepaling (artikel III 2.5.9) uitdrukkelijk wordt toegelicht in de *Toelichting op de Beroepscode 1998*. Daarin wordt inderdaad als uitgangspunt genomen dat een op naam gesteld dossier niet langer bewaard wordt dan noodzakelijk is voor het doel waarvoor het dossier is aangelegd. Hoe verhoudt die bepaling zich nu tot de genoemde bewaartermijn in de WGBO?

De codebepaling over de bewaartermijn is rechtstreeks gebaseerd op de Wet Bescherming Persoonsgegevens (WBP), die regels stelt voor het verwerken en bewaren van persoons-

gegevens. Bij het opstellen van de huidige beroepscode is gekozen voor deze formulering, omdat psychologen werkzaam zijn op velerlei werkgebieden en niet alleen in de gezondheidszorg. De *Beroepscode* als geheel is immers geschreven voor alle psychologen, ongeacht hun concrete werkterrein! De WBP dwingt de psycholoog in feite om een bewaartermijn te kiezen die niet langer dan noodzakelijk is in verhouding tot het professionele doel van de psycholoog. De *Toelichting* wijst de psychologen die als behandelaar of therapeut werkzaam zijn in de gezondheidszorg, er vervolgens met nadruk op dat zij zich hebben te houden aan de bewaartermijn van (tot voor kort) tien jaar zoals gesteld in de WGBO. En nu dus vijftien jaar.

De tekst van de *Toelichting* is rechtstreeks te downloaden van de pagina Beroepsethiek van www.psynip.nl. Op dezelfde pagina is recent de rubriek 'veel gestelde vragen' opgenomen. Als eerste onderwerp wordt daarin het dossier behandeld.

Mr Rosalinde Visser, medewerker beroepsethiek van het NIP, e-mail <ethiek@psynip.nl>.