

PDF hosted at the Radboud Repository of the Radboud University Nijmegen

The following full text is a publisher's version.

For additional information about this publication click this link.

<http://hdl.handle.net/2066/54742>

Please be advised that this information was generated on 2021-06-17 and may be subject to change.

De WISC-III bij kinderen met ernstige gedragsproblemen, ontwikkelingsproblemen en/of psychiatrische problemen

Henk Oosterbaan · Gert Kroes · van Brechtje Gent ·
Eric E. J. De Bruyn

Oosterbaan, H., Kroes, G., Van Gent, B., & De Bruyn, E. J. (2006). *The wisc-III in children with serious behavioural problems, development problems and/or psychiatric problems*, Kind en Adolescent, 27 (1), 56-69.

Summary: Research was conducted into how the recently standardised intelligence test wisc-III^{nl} compares with its predecessor, the wisc-R, in children with serious behavioural problems, development problems and/or psychiatric problems. The research group consisted of 59 children of elementary school age, who attended a cluster-4 school, and who were also receiving support. These children had taken the wisc-R test in the past; they were now retested using the WISC-III^{nl}. The IQ scores of both test versions appeared to match well, apart from a slight drop in verbal IQ in the WISC-III^{nl}. In this research group, the factor structure of the wisc-III^{nl} was found to be more similar to the factor structure found in the national standard group of the wisc-R, than to the factor structure to be found in the national standard group of the wisc-III^{nl}.

Keywords: wisc-III^{nl}, elementary school age, cluster-4 school.

Samenvatting Onderzocht is hoe de recent genormeerde intelligentietest wisc-III^{nl} zich verhoudt tot zijn voorganger, de wisc-R, bij kinderen met ernstige gedragsproblemen, ontwikkelingsproblemen en/of psychiatrische problemen. De onderzoeksgroep bestond uit 59 kinderen op de basisschoolleeftijd, die een cluster-4 school bezochten en tevens jeugdhulpverlening kregen. In het verleden was bij hen de wisc-R afgenomen; nu werden zij hertest met de wisc-III^{nl}. De IQ-scores van beide testversies bleken goed overeen te komen, een lichte daling van het verbale IQ van de wisc-III^{nl} daargelaten. De factorstructuur van de wisc-III^{nl} leek bij deze onderzoeksgroep meer op de factorstructuur die is aangetroffen bij de landelijke normeringsgroep van de

wisc-R, dan op de factorstructuur die is gevonden bij de landelijke normeringsgroep van de wisc-III^{nl}.

De Wechsler Intelligence Scale for Children (Wechsler, 1949) en zijn revisies behoren wereldwijd tot de meest gebruikte kinderintelligentietests. In 1986 verscheen een genormeerde Nederlandse versie, de wisc-R (De Bruyn, Vander Steene, & Van Haasen, 1986). Dit instrument werd onlangs opnieuw genormeerd en qua vorm en inhoud aan de huidige tijd aangepast. Zo maakten zwart-witplaten plaats voor gekleurde platen, werden gedateerde teksten in opgaven vervangen door actuelere teksten, werden aan enkele subtests items toegevoegd, en kwam er een geheel nieuwe subtest – Symbolen Vergelijken – bij. Deze meest recente versie van de wisc kwam onder de naam wisc-III^{nl} (Kort e.a., 2002) op de markt.

De wisc-III^{nl} heeft in de vakpers veel stof doen opwaaien. Vooral Tellegen (2002a, 2002b) leverde forse kritiek. Zijn belangrijkste bezwaren betreffen de representativiteit van de normeringssteekproef bij de oudere kinderen (oververtegenwoordiging van havo/vwo-leerlingen) en de grootte van het leeftijdsinterval die in de normtabellen wordt gehanteerd (voor alle leeftijdsgroepen geldt een interval van vier maanden, wat volgens Tellegen bij de jongere kinderen die zich aan de rand van het interval bevinden, tot vertekeningen kan leiden). Toen de commotie gaandeweg ook de landelijke pers bereikte werd zelfs besloten de test enige tijd uit de handel te nemen en aanvullend normeringonderzoek te doen. Met de publicaties van herziene normtabellen door het Nederlands Instituut van Psychologen (NIP; nip Dienstencentrum, 2003) en in de *Handleiding en verantwoording* (Kort e.a., 2005) leek de storm wat geluwd te zijn. Eind 2005 kreeg de test een

positieve beoordeling van de Commissie Testaangelegenheden Nederland (Nederlands Instituut van Psychologen, 2005). De uitgangspunten bij de testconstructie, alsook de kwaliteit van het testmateriaal en van de handleiding werden ‘goed’ bevonden; normen, betrouwbaarheid en begripsvaliditeit zijn als ‘voldoende’ beoordeeld (met de aantekening dat interpretatie op subtestniveau wordt afgeraden vanwege de lage betrouwbaarheid van enkele subtests); alleen de criteriumvaliditeit heeft het predicaat ‘onvoldoende’ gekregen, aangezien hierover nog te weinig onderzoek beschikbaar is.

In het traject dat leidt tot geïndiceerde jeugdhulp of tot plaatsing binnen het speciaal onderwijs wordt vrijwel altijd intelligentieonderzoek verricht. En is het kind eenmaal geplaatst, dan ontstaat na verloop van tijd vaak behoefte aan herhalingsonderzoek. Aangenomen mag worden dat in de diagnostische praktijk de WISC-R geleidelijk aan plaats zal maken voor de WISC-III^{nl}. Daardoor zullen er de komende jaren kinderen zijn bij wie eerst de WISC-R en later de WISC-III^{nl} is afgenomen. De vraag doet zich nu voor in hoeverre beide WISC-versies, wanneer ze bij klinische populaties worden toegepast, gelijkwaardig zijn. Testversies zijn gelijkwaardig als ze dezelfde uitspraken toelaten over het niveau en de betekenis van de behaalde testcores. Met ‘niveauverschillen’ worden de verschillen bedoeld tussen de normscores die op beide versies zijn behaald, en met ‘betekenis’ de interpretatie van de subtests vanuit onderliggende intelligentiedimensies of -factoren.

Verschillen tussen de scores die een kind op twee opeenvolgende versies van een intelligentietest behaalt, kunnen door meerdere factoren veroorzaakt zijn. Voor ons onderzoek zijn het *testeffect*, het *Flynn-effect*, en het *groepsspecifieke effect* de meest relevante. Het *testeffect* of *leereffect* heeft betrekking op de prestatieverbetering die zich voordoet doordat het kind de test al eerder gedaan heeft. Het *Flynn-effect* heeft betrekking op het gegeven dat het absolute prestatieniveau van de bevolking in de loop der tijd blijkt toe te nemen. Als gevolg daarvan zal, naarmate de tijd vordert, het IQ van kinderen die met hetzelfde instrument worden getest steeds hoger worden ingeschat, zolang de test niet opnieuw wordt genormeerd. Hierover is al in 1975 gepubliceerd door Kaufman, naar aanleiding van de vervanging van de WISC (Wechsler, 1949) door de WISC-R (Wechsler, 1974), en later ook door Sattler (1982). Het Flynn-effect ontleent zijn naam echter aan de Nieuw-Zeelander James Flynn (1987), die aan de hand van literatuuronderzoek aantoonde dat het gemiddelde IQ binnen één generatie 5 tot 25 punten kan stijgen. Flynn (1999) heeft, aan de hand van diverse IQ-tests en over verschillende leeftijds-groepen, aangetoond dat het gemiddelde IQ met ongeveer 0.3 punt per jaar toeneemt. Het *groepsspecifieke effect* ten slotte heeft betrekking op de mogelijkheid dat het

reviseren van een test een andere invloed heeft op de prestaties van bepaalde (klinische) groepen dan op die van de normeringssteekproef.

Zimmerman en Woo-Sam (1996) toonden aan dat het testeffect verwaarloosbaar klein is indien het tijdsinterval tussen de afname van de WISC-R en de afname van de WISC-III ten minste negen maanden bedraagt. Zij verzamelden de IQ-scores op de WISC-R en de WISC-III uit zestien gepubliceerde en ongepubliceerde onderzoeken, van in totaal 1000 kinderen die verwezen waren naar de jeugdhulpverlening of het speciaal onderwijs. Het tijdsinterval tussen de afname van beide tests varieerde van 9 tot 43 maanden, met een gemiddelde van 29 maanden. Zimmerman en Woo-Sam vonden een gemiddeld verschil van 6.4 punten tussen het totaal IQ van de WISC-R (99.8) en het totaal IQ van de WISC-III (93.4). Bij het verbale IQ bedroeg het verschil 5.6 punten en bij het performale IQ 5.5 punten. Zimmerman en Woo-Sam schreven deze verschillen geheel toe aan het Flynn-effect en niet aan een testeffect, aangezien ze even groot bleken te zijn in de onderzoeken waarin eerst de WISC-III en vervolgens de WISC-R was afgenomen, als in de onderzoeken waarin eerst de WISC-R en daarna de WISC-III was afgenomen. Ook in de klinische praktijk wordt meestal de vuistregel gehanteerd dat een intelligentietest na negen maanden opnieuw mag worden afgenomen, zonder dat rekening hoeft te worden gehouden met een test-effect (zie o.a. De Bruyn, Heinrichs, & Oosterbaan, 1979).

Een voorbeeld van een groepsspecifiek effect vinden we bij Mahone en collega's (2003). Zij onderzochten de verschillen in perfoormaal IQ tussen WISC-R en WISC-III bij 122 kinderen met ADHD. Bij de ene helft van deze kinderen werd de WISC-R afgenomen en bij de andere helft de WISC-III. Bij 46 ‘normale’ kinderen werd eenzelfde procedure toegepast. De groepen bij wie de WISC-R en de groepen bij wie de WISC-III werd afgenomen werden gematcht op sekse en verbaal IQ. Vervolgens werden de performale IQ's op de WISC-R en de WISC-III met elkaar vergeleken. Bij de ADHD-kinderen bleek het performale IQ op de WISC-III gemiddeld 5.7 punten lager te zijn dan het performale IQ op de WISC-R; bij de ‘normale’ kinderen werd geen verschil tussen beide testversies aangetroffen. De onderzoekers zagen hierin een bevestiging van hun hypothese dat bepaalde wijzigingen in de WISC-III – zoals een stringenter toekenning van bonuspunten voor responsnelheid – meer invloed hebben op de prestaties van ADHD-kinderen dan op die van ‘normale’ kinderen.

Pesch en Ponsioen (2004) waren de eersten die in Nederland publiceerden over onderzoek met de WISC-III^{nl} bij een klinische groep. Zij namen bij 75 licht verstandelijk gehandicapte jongeren eerst de WISC-R en daarna de WISC-III^{nl} af. Bij de WISC-III^{nl} bleek het totale IQ gemiddeld 6.6 punten, het verbale IQ 7.3 punten, en het performale IQ 5.2 punten lager te zijn dan bij de WISC-R. Pesch en Ponsioen

vinden het opmerkelijk dat het Flynn-effect zich bij hun onderzoeksgroep in sterkere mate voordoet bij het verbale dan bij het performale IQ, aangezien het Flynn-effect bij verbale tests doorgaans juist minder uitgesproken is.

Drie jaar na het verschijnen van de WISC-III^{nl} is de oorspronkelijke handleiding vervangen door de *Handleiding en verantwoording* (Kort e.a., 2005). Hierin (pagina 68 e.v.) wordt uitvoerig melding gemaakt van een onderzoek bij ruim 300 Vlaamse kinderen die waren getest door revalidatiecentra en CLB's (Centrum voor Leerlingen Begeleiding, de Vlaamse pendant van de schooladviesdienst). Bij deze kinderen was eerst de WISC-R en daarna (gemiddeld 3.0 jaar later) de WISC-III^{nl} afgenomen. Verreweg de meeste kinderen (93%) waren op de basisschoolleeftijd en het merendeel (ongeveer 70%) volgde buitengewoon onderwijs. Bij de totale groep daalde het totale IQ 2.83 punten, daalde het verbale IQ 4.85 punten, en steeg het performale IQ 0.29 punten. Nadere analyse liet zien dat de verschillen het grootst waren bij de minder begaafde kinderen (IQ tussen 70 en 90); bij die selectie daalde het verbale IQ ongeveer 4 punten en bleef het performale IQ nagenoeg gelijk. Bij de kinderen met een gemiddelde intelligentie (IQ tussen 90 en 110) waren de verschillen miniem: hun verbale WISC-III^{nl}IQ bleek 1 punt lager uit te vallen dan hun verbale WISC-RIQ, terwijl hun performale WISC-III^{nl}IQ 2 punten hoger was dan hun performale WISC-RIQ.

Ons onderzoek nu is in de eerste plaats gericht op de vraag hoe de WISC-III^{nl} en de WISC-R zich qua scoreniveau tot elkaar verhouden bij normaal begaafde kinderen op de basisschoolleeftijd, bij wie sprake is van ernstige gedragsproblemen, ontwikkelingsproblemen en/of psychiatrische problemen. Op grond van de Angelsaksische literatuur over het Flynn-effect bij klinische populaties zouden wij lagere IQ scores op de WISC-III^{nl} dan op de WISC-R mogen verwachten; en de verschillen tussen de performale IQ's zouden groter moeten zijn dan de verschillen tussen de verbale IQ's. Nederlands en Vlaams onderzoek bij klinische populaties waarin de WISC-R en de WISC-III^{nl} met elkaar worden vergeleken is vooralsnog schaars. De eerdergenoemde onderzoeken van Pesch en Ponsioen (2004) en van Kort en collega's (2005, p. 68 e.v.) suggereren dat een Flynn-effect vooral zal optreden bij het verbale IQ, en niet of in geringere mate, bij het performale IQ.

Onze tweede onderzoeksvraag is gericht op de factorstructuren van beide testversies. De literatuur hierover leert ons het volgende. Alle versies van de WISC bevatten normtabellen waarmee behalve een totaal IQ ook een verbaal IQ en een performaal IQ kunnen worden bepaald. Onderzoek heeft keer op keer het bestaan van deze beide factoren bevestigd. Daarnaast zijn er ook andere factorindelingen gangbaar. Zo bevat de WISC-R normtabellen voor drie factoren: Verbaal Begrip, Perceptuele Organisatie

en Vrijheid van Afleidbaarheid. Bij de Amerikaanse en Engelse WISC-III werden vier factoren gevonden, te weten Verbaal Begrip, Perceptuele Organisatie, Vrijheid van Afleidbaarheid en Verwerkingssnelheid. Deze laatste factor, die in eerdere versies nog niet voorkwam, bestaat uit de subtests Substitutie en Symbolen Vergelijken. Bij een exploratieve factoranalyse op de Nederlandse normeringsgroep van de WISC-III^{nl}, zo vermeldt de handleiding, zijn wel de 'Amerikaanse' factoren Verbaal Begrip, Perceptuele Organisatie, en Verwerkingssnelheid aangetroffen, maar bleek de factor Vrijheid van Afleidbaarheid te ontbreken. In de handleiding van de WISC-III^{nl} wordt tevens vermeld dat de factorstructuur die is aangetroffen bij de Franse normeringsgroep van de WISC-III, overeenkomt met de factorstructuur die bij de Nederlandse normeringsgroep van de WISC-III^{nl} is gevonden.

De factorstructuur van de algemene populatie waarop een test is genormeerd hoeft niet noodzakelijkerwijs ook te passen bij specifieke populaties. In hun eerdergenoemde onderzoek met de WISC-III^{nl} bij licht verstandelijk gehandicapte jongeren vonden Pesch en Ponsioen (2004) bij exploratieve factoranalyse niet drie, maar vier factoren. De eerste, tweede en vierde factor kwamen overeen met de in de landelijke steekproef van de WISC-III^{nl} gevonden factoren Verbaal Begrip, Perceptuele Organisatie en Verwerkingssnelheid. De derde factor bestond bij de groep licht verstandelijk gehandicapte jongeren uit de subtests Rekenen en Cijferreeksen. Die twee subtests vormen bij de WISC-R, samen met Substitutie, de factor Vrijheid van Afleidbaarheid. Watkins en Kush (2002) toetsten de indeling in vier factoren van de Engels/Amerikaanse WISC-III bij 1200 jeugdigen met leerproblemen. Zij vonden wel de drie factoren Verbaal Begrip, Perceptuele Organisatie en Verwerkingssnelheid, maar niet de factor Vrijheid van Afleidbaarheid. Omdat bovendien de predictieve validiteit van de factoren Verwerkingssnelheid en Vrijheid van Afleidbaarheid twijfelachtig is, pleiten Watkins en Kush ervoor deze beide factoren slechts met grote terughoudendheid te gebruiken.

De literatuur laat dus zien dat bij sommige specifieke doelgroepen een factorstructuur wordt aangetroffen die wel overeenkomt met de factorstructuur van de landelijke normeringsgroep, terwijl bij andere specifieke doelgroepen een afwijkende factorstructuur wordt gevonden. Daarom kunnen wij over de factorstructuur van de WISC-III^{nl} bij onze doelgroep nauwelijks voorspellingen doen. Door de relatief beperkte omvang van onze onderzoeksgroep hebben we gekozen voor de exploratieve factoranalyse, een beschrijvende methode die in het WISC-onderzoek op ruime schaal wordt toegepast en ook in de reeds geciteerde onderzoeken van Pesch en Ponsioen (2004) en van Kort en collega's (2005) is uitgevoerd.

Ten slotte, de door ons onderzochte klinische groep komt uit een populatie waarvan het klinische profiel aanmerkelijk verschilt van de door de Pesch en Ponsioen (2004) en Kort en collega's (2005) onderzochte klinische groepen. Specifiek voor onze klinische populatie is de combinatie van normale intelligentie en ernstige ontwikkelingsproblemen (zie hieronder de subparagraaf 'Deelnemers'). Ook voor deze populatie moet worden nagegaan of er sprake is van een Flynn-effect en of de factorstructuur van de nieuwe versie overeenkomt met die van de normpopulatie. Het door ons uitgevoerde onderzoek levert hier toe een eerste aanzet.

Methodede

Deelnemers

De onderzoeksgroep bestond uit 59 kinderen (54 jongens en 5 meisjes) op de basisschoolleeftijd, die in behandeling waren bij Stichting De Waarden in Nijmegen. De Waarden is een multifunctionele organisatie voor jeugdhulpverlening. Aan de instelling zijn twee zogenoemde cluster-4 scholen verbonden, de Paedologisch Instituutschool en de Roelant Berk&Beukschool voor ZMOK-onderwijs. Om een kind op een cluster-4 school te kunnen plaatsen moet een bevoegd diagnosticus, aan de hand van een psychiatrisch classificatiesysteem (DSM-IV of ICD-10; American Psychiatric Association, 2000; World Health Organization, 1992), het kind een classificatie hebben gegeven. Dit betekent voor onze onderzoeksgroep dat alle 59 kinderen 'een DSM-classificatie hebben'. Op de Paedologisch Instituutschool zitten voornamelijk kinderen met autisme spectrumstoornissen, en op de Roelant Berk&Beukschool voornamelijk kinderen met ADHD of met een oppositio-neel-opstandige gedragsstoornis. Voor het onderzoek werden die kinderen geselecteerd van wie een compleet WISC-R protocol aanwezig was, en wier ouders toestemming hadden gegeven om de WISC-III^{nl} af te nemen. Twee kinderen waarvan een volledig WISC-R protocol aanwezig was, vielen uit omdat hun ouders geen toestemming voor afname van de WISC-III^{nl} gaven. Van de 59 deelnemende kinderen bezochten 38 de Paedologisch Instituutschool en 21 de Roelant Berk&Beukschool. De jeugdhulpverlening bestond bij 31 kinderen uit residentiële behandeling en bij 28 kinderen uit (deeltijd)dagbehandeling. Op het moment waarop de WISC-III^{nl} werd afgenomen waren de kinderen tussen 7;6 en 13;0 jaar oud. Hun gemiddelde leeftijd bedroeg 10;2 jaar.

Procedure

De WISC-R was reeds voor aanvang van het onderzoek afgenomen; bij sommige kinderen voorafgaand aan, bij andere kinderen tijdens hun plaatsing op De Waarden. De WISC-III^{nl} is afgenomen door twee gekwalificeerde orthopedagogen en gescoord aan de hand van de herziene normtabellen (nip Dienstencentrum, 2003). Tussen de afname van de WISC-R en die van de WISC-III^{nl} lag bijna twee jaar (gemiddeld 23 maanden, bereik 9-54 maanden, *SD* 11 maanden). Deze relatief grote tijdsspanne maakt aannemelijk dat de metingen voldoende onafhankelijk zijn afgenomen. Bovendien garandeert het feit dat bij elk kind beide tests door verschillende personen zijn afgenomen dat geen 'tester-effect' kan optreden.

Meetinstrument

Bij zowel de WISC-R als de WISC-III^{nl} worden behalve een totaal IQ, een verbaal IQ en een performaal IQ, ook drie factoren onderscheiden. De factoren van de WISC-R worden respectievelijk aangeduid met F1 (bestaande uit de subtests Informatie, Overeenkomsten, Woordenschat en Begrijpen), P3 (Onvolledige Tekeningen, Blokpatronen, Figuur Leggen en Doolhoven) en F3 (Rekenopgaven, Cijferreeksen en Substitutie). De drie factoren van de WISC-III^{nl} dragen de namen Verbaal Begrip (Informatie, Overeenkomsten, Woordkennis en Begrijpen), Perceptuele Organisatie (Onvolledige Tekeningen, Plaatjes Ordenen, Blokpatronen en Figuur Leggen), en Verwerkingssnelheid (Substitutie en Symbolen Vergelijken).

Resultaten

Verschillen in scores tussen de WISC-R en de WISC-III^{nl}

In tabel 1 zijn de gemiddelden en standaarddeviaties op de WISC-R en de WISC-III^{nl}, alsook de verschilcores tussen beide tests, naast elkaar gezet. Rekening houdend met het aantal uitgevoerde toetsen is de uitkomst als volgt. De tests blijken voor wat betreft het totale IQ en het performale IQ nauwelijks van elkaar te verschillen. Wel treffen we een significant verschil aan bij het verbale IQ (na een Bonferroni-correctie ligt de significantiegrens bij $p = .017$); dit valt bij de WISC-III^{nl} bijna 4.5 punten lager uit dan bij de WISC-R. Deze uitkomsten komen vrij nauwkeurig overeen met de verschillen tussen WISC-R en de WISC-III^{nl} die gerapporteerd worden in de handleiding van de WISC-III^{nl} (Kort e.a., 2005) voor een onderzoek bij ruim 300 Vlaamse kinderen.

Tabel 1 De gemiddelden en standaarddeviaties van scores op de WISC-R en de WISC-III^{nl}, de verschillen tussen de gemiddelden, en de correlaties tussen de scores op beide testversies ($N = 59$)

Standaardscore	WISC-R		WISC-III ^{nl}		WISC-R – WISC-III ^{nl}			
	M	SD	M	SD	Vershil	t	p*	r
Totaal IQ	94.30	15.46	92.88	14.36	1.42	-1.31	.195	.85
Verbaal IQ	97.72	16.15	93.23	14.80	4.48	-3.80	.000	.83
Performaal IQ	92.20	17.28	94.30	16.20	-2.10	1.50	.138	.79
<i>Verbale subtests</i>								
Informatie	9.70	3.26	9.48	3.11	0.22	.53	.595	.51
Overeenkomsten	10.17	3.28	9.63	2.86	0.53	1.49	.142	.60
Rekenen	8.87	3.26	8.27	3.18	0.60	1.76	.084	.66
Woordkennis	9.90	3.38	8.27	3.16	1.12	3.37	.001	.69
Begrijpen	10.13	3.07	8.57	3.11	1.57	3.73	.000	.45
Cijferreeksen	8.92	3.32	8.53	3.37	0.38	.85	.398	.46
<i>Performale subtests</i>								
Onvolledige Tekeningen	8.70	3.28	9.82	3.34	-1.12	-2.61	.011	.50
Substitutie	8.43	2.79	8.12	2.38	0.32	.87	.389	.41
Plaatjes Ordenen	10.02	3.03	9.88	3.64	0.13	.30	.764	.49
Blokpatronen	8.27	3.71	9.27	3.44	-1.00	-3.54	.001	.82
Figuur Leggen	7.90	3.57	8.97	3.44	-1.07	-2.61	.012	.59
Symbolen Vergelijken	-	-	8.35	2.73	-	-	-	-
Doolhoven	10.53	3.37	9.27	3.10	1.25	2.82	.007	.45

*Het verschil tussen de WISC-R en WISC-III scores is getoetst met de t-test voor gepaarde waarnemingen.

In dat onderzoek bedroeg de gemiddelde daling ten opzichte van de WISC-R 2.83 punten voor het totale IQ en 4.85 IQ-punten voor het verbale IQ, en liet het performale IQ een lichte stijging van 0.29 punten zien. Van de verbale subtests laten Woordkennis en Begrijpen een significante scoredaling zien. Bij de performale subtests is er alleen bij Blokpatronen een significant verschil (significantiengrens is $p < .005$ na Bonferroni-correctie) te zien ten opzichte van de WISC-R.

Tabel 1 bevat tevens de Pearson-correlaties (r) tussen de twee testversies. Die blijken voor wat betreft de IQ-scores te variëren van .79 tot .85. De correlaties voor de subtests zijn over het algemeen lager (variërend van .41 voor Substitutie tot .82 voor Blokpatronen). Ze zijn vergelijkbaar met de correlaties die gerapporteerd zijn voor het eerdergenoemde Vlaamse onderzoek (zie Kort e.a., 2005), waarbij de correlaties tussen de IQ-scores varieerden van .81 tot .88 voor de IQ-scores, en die tussen de subtests van .41 (Doolhoven) tot .71 (Woordenschat). De correlaties tussen de IQ-scores zijn hoog te noemen. Als we deze IQ-intercorrelaties corrigeren voor de onbetrouwbaarheid van de betrokken tests variëren ze voor onze onderzoeksgroep van .89 tot .92.

Het Flynn-effect kan ook worden getoetst aan de hand van de correlatie tussen het tijdsinterval tussen de afname van de WISC-R en de WISC-III^{nl} enerzijds, en het verschil tussen de scores op beide testversies anderzijds. Op grond van de gevonden niveaoverschillen voorspelden wij een

negatieve correlatie tussen het tijdsinterval en het scoreverschil bij het verbale IQ. Deze voorspelling wordt bevestigd ($r = -.28$, $p = .035$). Bij het performale IQ en bij het totale IQ verwachtten wij geen statistisch significant verband te zullen aantreffen. Ook dit blijkt inderdaad het geval te zijn, zowel bij het totale IQ ($r = -.12$, $p = .367$) als bij het performale IQ ($r = .08$, $p = .525$). Deze correlatie-analyse toont dus aan dat, net als de analyse van de niveaoverschillen, binnen onze onderzoeksgroep alleen bij het verbale IQ sprake is van een Flynn-effect.

Factorstructuur

Tabel 2 geeft de resultaten weer van de exploratieve factoranalyse op de WISC-III^{nl} bij de steekproef van De Waarden. Deze worden vergeleken met die van de landelijke normeringsgroep (Kort e.a., 2005). Bij onze steekproef blijkt een driefactorenmodel de onderliggende structuur het best weer te geven; hiermee wordt 59.9% van de totale variantie verklaard.

Anders dan bij de landelijke steekproef blijkt de derde factor in onze steekproef niet uit twee, maar uit vier subtests te bestaan: naast Substitutie en Symbolen Vergelijken maken namelijk ook Rekenen en Cijferreeksen hier deel van uit. Deze derde factor komt grotendeels overeen met de derde factor die bij de Amerikaanse en Engelse normeringsgroep gevonden werd en door de Amerikaanse onderzoekers Vrijheid van Afleidbaarheid

Tabel 2 Factorloadingen van de exploratieve factoranalyse op de WISC-III^{nl} bij De Waarden (N = 59) en bij de landelijke normgroep (N = 1239)*

Subtests	Verbaal Begrip		Perceptuele Organisatie		Vrijheid van Afleidbaarheid	Verwerkings-snelheid
	De Waarden	Normgroep	De Waarden	Normgroep	De Waarden	Normgroep
Informatie	.71	.72				
Overeenkomsten	.85	.70				
Rekenen	.53	.56			.58	
Woordkennis	.81	.82				
Begrijpen	.72	.66				
Cijferreeksen					.69	
Onv. Tekeningen	.63		.45	.45		
Substitutie			.52		.64	.71
Plaatjes Ordenen			.62	.44		
Blokpatronen			.66	.72		
Figuur Leggen			.75	.69		
Symbolen Vergelijken			.66		.48	.71
Doolhoven			.72			
Percentage verklaarde variantie	25.1%	21.7%	21.7%	14.8%	13.0%	9.9%

* Alleen ladingen > .40 zijn weergegeven.

werd genoemd. In navolging van de Amerikaans/Engelse WISC-III-onderzoekers hebben wij onze derde factor eveneens Vrijheid van Afleidbaarheid genoemd. Een verschil met de Amerikaans/Engelse factorstructuur is dat in onze steekproef de derde factor naast Rekenen, Cijferreeksen en Substitutie ook de subtest Symbolen Vergelijken bevat. Verder blijkt bij onze steekproef de factor Perceptuele Organisatie uit meer subtests te bestaan dan bij de landelijke normgroep het geval is: niet alleen Onvolledige Tekeningen, Plaatjes Ordenen, Blokpatronen en Figuur Leggen, maar ook Substitutie, Symbolen Vergelijken en Doolhoven laden hoog op deze factor. Opmerkelijk is dat in onze steekproef de subtests Substitutie en Symbolen Vergelijken zowel op de factor Perceptuele Organisatie als op de factor Vrijheid van Afleidbaarheid hoog laden.

De door ons aangetroffen factorstructuur komt sterk overeen met de factorstructuur van de WISC-R (uiteraard met uitzondering van de nieuwe subtest Symbolen Vergelijken), zoals die in eerder onderzoek bij een soortgelijke klinische populatie werd aangetroffen (vgl. De Bruyn, Kroes, & Oosterbaan, 1990).

Discussie

In dit onderzoek werd nagegaan of er sprake is van significante niveauverschillen bij de overgang van de WISC-R naar de WISC-III^{nl} – zoals we die op grond van het Flynn-effect zouden kunnen verwachten – bij een groep kinderen

op de basisschoolleeftijd met leer-, ontwikkelings-, en gedragsproblemen. Tevens werd de factorstructuur van de WISC-III^{nl} bij deze groep kinderen onderzocht. Alleen voor het verbale IQ werd een significant verschil gevonden tussen de WISC-R en de WISC-III^{nl}; bij het totale IQ en performale IQ troffen wij geen significante verschillen aan. Dit impliceert dat, indien de WISC-III^{nl} na de WISC-R wordt afgenomen, rekening dient te worden gehouden met een daling van het verbale IQ ter grootte van vier of vijf punten. Deze daling kan worden toegeschreven aan het Flynn-effect. De negatieve correlatie die wij aantroffen tussen enerzijds het tijdsinterval tussen de afname van beide tests en anderzijds de scoreverschillen tussen beide tests, onderstreept deze conclusie. Op subtestniveau bleken de scores op de subtests Woordkennis, Begrijpen en Doolhoven significant te dalen; de scores op de subtests Onvolledige Tekeningen, Blokpatronen en Figuur Leggen daarentegen waren bij de WISC-III^{nl} significant hoger dan bij de WISC-R. De sterke samenhang die wij vonden tussen de totale IQ-scores op de WISC-R en de WISC-III^{nl} wijst erop dat beide testversies dezelfde constructen meten.

Dat bij onze onderzoeksgroep het verbale IQ daalt, terwijl het performale IQ gelijk blijft, sluit niet aan op Angelsaksisch onderzoek naar de relatie tussen de WISC-R en de WISC-III bij klinische groepen. Zo constateerden Zimmerman en Woo-Sam (1997) in een literatuuronderzoek met betrekking tot diverse klinische groepen, en Mahone en collega's (2003) in een onderzoek bij ADHD-kinderen, een sterkere daling van het performale IQ dan het verbale IQ. Onze resultaten corresponderen echter

opmerkelijk goed met die van het qua opzet vergelijkbare onderzoek bij Vlaamse kinderen, bij wie eveneens eerst de WISC-R en later de WISC-III^{nl} was afgenomen (Kort e.a., 2005). Ook bij deze Vlaamse groep bleek de discrepantie tussen de verbale IQ's veel groter te zijn dan die tussen de performale IQ's.

De WISC-III^{nl} werd na een behandelperiode van gemiddeld twee jaar afgenomen. Dit in aanmerking genomen, zou het verschil in niveaudaling bij het performale IQ tussen ons onderzoek en andere gepubliceerde onderzoeken veroorzaakt kunnen zijn door een behandel-effect. Wanneer er inderdaad sprake is geweest van een behandel-effect, dan zou dit moeten resulteren in een verband tussen de duur van de (behandel)periode tussen beide metingen en de mate waarin het performale IQ toeneemt. Een dergelijk verband hebben wij echter niet kunnen aantonen ($r = .11, p = .21$). Overigens dient te worden opgemerkt dat ons onderzoek correlatieel van aard is. Een directe toetsing van een behandelhypothese vergt een experimentele opzet. Onze resultaten geven slechts indirecte steun aan de veronderstelling dat het weinig waarschijnlijk is dat een behandel-effect een door het Flynn-effect veroorzaakte daling van het performale IQ heeft geneutraliseerd.

Exploratieve factoranalyse bracht bij onze onderzoeksgroep drie te onderscheiden factoren aan het licht, die samen ongeveer de helft van de variantie in de subtestscores verklaren. Opvallend is dat bij deze onderzoeksgroep, anders dan bij de landelijke steekproef, de derde factor bestaat uit vier subtests: behalve Substitutie en Symbolen Vergelijken omvat hij ook Rekenen en Cijferreeksen. De door ons gevonden derde factor lijkt dus meer op de derde factor die in de Amerikaans/Engelse normeringsgroep is aangetroffen, dan op de derde factor uit de Nederlandse normeringsgroep. Hij blijkt ook sterk overeen te komen met de derde factor van de WISC-R, die ook wel Vrijheid van Afleidbaarheid wordt genoemd. Dit suggereert dat het verstandig kan zijn om bij het interpreteren van de WISC-III^{nl} bij kinderen met ernstige gedragsproblemen, ontwikkelingsproblemen en/of psychiatrische problemen de WISC-R factor Vrijheid van Afleidbaarheid te handhaven.

Samenvattend constateren we dat bij onze onderzoeksgroep de IQ-scores die met beide WISC-versies verkregen worden redelijk met elkaar overeenkomen, een lichte daling van het verbale IQ van de WISC-III^{nl}, die kan worden toegeschreven aan de nieuwe normering, daargelaten. Ook de correlaties tussen de IQ's zijn hoog. Voorts blijkt de WISC-III^{nl} bij onze onderzoeksgroep een factorstructuur op te leveren die meer correspondeert met de driefactorenstructuur van de WISC-R dan met de in de handleiding van de WISC-III^{nl} gepresenteerde structuur.

Er moet wel worden aangetekend dat onze onderzoeksgroep betrekkelijk klein was en uit voornamelijk jongens

bestond, die behandeld werden binnen één instelling (het oververtegenwoordigd zijn van jongens in de jeugdhulpverlening en in het speciaal onderwijs is overigens een algemeen fenomeen, dat al zeer lang bekend is). Om na te gaan in hoeverre onze bevindingen gelden voor kinderen met ernstige gedragsproblemen, ontwikkelingsproblemen en/of psychiatrische problemen in het algemeen, is replicatieonderzoek op een grotere en meer representatieve steekproef nodig. Onderzoek bij grotere en meer representatieve klinische steekproeven zou daarnaast de mogelijkheid creëren om – bij gebleken verschillen tussen de exploratief gevonden factorstructuur van klinische en normeringsgroepen – de robuustheid van deze verschillen ook via toetsende factoranalyse te onderzoeken.

Literatuur

- American Psychiatric Association (2000). *Diagnostic and Statistical Manual of Mental disorders (4th ed.)-Text Revision*. Washington DC: American Psychiatric Association.
- De Bruyn, E. E. J., Heinrichs, W., & Oosterbaan, H. (1979). Niveauverschillen op niet-genormeerde Nederlandse versies van de intelligentietests HAWIK en WISC-R. *Tijdschrift voor Orthopedagogiek*, 18, 464-469.
- De Bruyn, E. E. J., Kroes, G., & Oosterbaan, H. (1990). De WISC-R en herhalingsonderzoek bij kinderen met ernstige gedrags- en leerproblemen. *Tijdschrift voor Orthopedagogiek*, 29, 599-607.
- De Bruyn, E. E. J., Vander Steene, G., & Van Haasen, P. P. (Red.) (1986). *Wechsler Intelligence Scale for Children-Revised (WISC-R)*. Nederlandsestalige uitgave en verantwoording. Lisse: Swets & Zeitlinger.
- Flynn, J. R. (1987). Massive IQ gains in 14 nations: What IQ tests really measure. *Psychological Bulletin*, 101, 171-191.
- Flynn, J. R. (1999). Searching for justice. The discovery of IQ gains over time. *American Psychologist*, 54, 5-20.
- Kaufman, A. S. (1975). Factor analysis of the WISC-R at eleven age levels between 6½ and 16½ years. *Journal of Consulting and Clinical Psychology*, XLIII, 135-147.
- Kort, W., Schittekatte, M., Compaan, E. L., Bosmans, M., Bleichrodt, N., Vermeir, G. e.a. (2002). WISC-III^{nl}. Handleiding. Nederlandse bewerking. London: The Psychological Corporation.
- Kort, W., Schittekatte, M., Dekker, P. H., Verhaeghe, P., Compaan, E. L., Bosmans, M. e.a. (2005). WISC-III^{nl}. Handleiding en verantwoording. Nederlandse bewerking. London: Harcourt Assessment.
- Mahone, E. M., Miller, T. L., Koth, C. W., Mostofsky, S. H., Goldberg, M. C., & Denckla, M. B. (2003). Differences between WISC-R and WISC-III performance scale among children with ADHD. *Psychology in Schools*, 40, 4, 331-340.
- Nederlands Instituut van Psychologen (2005). *Documentatie van tests en testresearch in Nederland, aanvulling 2005/03*. Amsterdam: Nederlands Instituut van Psychologen (NIP)/Boom Testuitgevers.
- NIP Dienstencentrum (2003). *Errata en normtabellen WISC-III^{nl}. Oktober 2003*. Amsterdam: NIP Dienstencentrum.
- Pesch, W., & Ponsioen, A. (2004). Flinterdunne en flagrante Flynn-effecten bij licht verstandelijk gehandicapte kinderen. Aanbevelingen voor het gebruik van de WISC-III. *De Psycholoog*, 39, 64-68.
- Sattler, J. S. (1982). *Assessment of children's intelligence and special abilities*. Boston: Allyn and Bacon.

- Tellegen, P. (2002a). De wisc-iii^{nl}. Een illusie armer. *De Psycholoog*, 37, 607-610.
- Tellegen, P. (2002b). Afname van de waais-iii of wisc-iii. Verantwoord en verstandig? *De Psycholoog*, 37, 677-679.
- Watkins, M. W., & Kush, J. C. (2002). Confirmatory factor analysis of the wisc-iii for students with learning disabilities. *Journal of Psychoeducational Assessment*, 20, 4-19.
- Wechsler, D. (1949). *Manual for the Wechsler Intelligence Scale for Children*. New York: Psychological Corporation.
- Wechsler, D. (1974). *Manual for the Wechsler Intelligence Scale for Children-Revised*. New York: Psychological Corporation.
- World Health Organization (1992). *International Statistical Classification of Diseases and Related Health Problems, 1989 Revision*. Geneva: World Health Organization.
- Zimmerman, I. L., & Woo-Sam, J. M. (1996). Is retesting with the wisc-iii a defensible procedure? *Perceptual and motor skills*, 82, 349-350.
- Zimmerman, I. L., & Woo-Sam, J. M. (1997). Review of the criterion-related validity of the wisc-iii: The first five years. *Perceptual and Motor Skills*, 85, 531-546.