



Het Flynn-effect

Vergelijking van het Flynn-effect bij normaal begaafde kinderen en verstandelijk beperkten

Cynthia Abbink & Mayke Prinsen
794892 836757

**Master thesis
Universiteit van Tilburg
Psychologie en Geestelijke Gezondheid
Kinder- en Jeugdpsychologie**

Begeleider UvT: Dr. J. Scheirs
Begeleider Cello: Dr. J. Blok
Tweede beoordelaar UvT: Dr. M. Feltzer

Juni 2008

Samenvatting

James Flynn stelde in 1984 vast dat de prestaties op intelligentietests over de tijd heen stijgen, ongeveer 3 – 5 IQ-punten per decennium. Dit zogeheten Flynn-effect maakt hernormering van intelligentietests noodzakelijk, zodat weer een gemiddelde IQ-score van 100 en een standaarddeviatie van 15 worden gevonden. Het huidige onderzoek diende als vervolgstudie op het onderzoek van Nijman (2007). Deze studie onderzocht het Flynn-effect bij verstandelijk beperkten. Hoewel het Flynn-effect in de algemene populatie vrijwel altijd werd teruggevonden, gaven onderzoeken naar het Flynn-effect bij mensen met een verstandelijke beperking uiteenlopende resultaten. De uiteenlopende resultaten zouden het gevolg kunnen zijn van een bodemeffect. In tegenstelling tot andere onderzoeken maakte Nijman (2007) daarom gebruik van de SON 2,5 – 7 en de SON-R 2,5 – 7, omdat deze beter aansluiten bij de cognitieve capaciteiten van de verstandelijk beperkten. Nijman (2007) vond een significant Flynn-effect bij verstandelijk beperkten. In het huidige onderzoek is het Flynn-effect onderzocht bij normaal begaafde kinderen, ook middels de SON 2,5 – 7 en de SON-R 2,5 – 7. Het onderzoek vond plaats op een reguliere basisschool (N=70). Om een optimale vergelijking te kunnen maken met Nijman (2007) bestond de onderzoeksgroep uit kinderen van 4 – 6 jaar oud. Hiermee werd aangesloten bij de mentale leeftijd van de groep verstandelijk beperkten uit het onderzoek van Nijman (2007). In het huidige onderzoek werd een significant Flynn-effect bij normaal begaafde kinderen gevonden. De sterkte van het Flynn-effect bij normaal begaafde kinderen was echter kleiner dan dat bij verstandelijk beperkten. Ook binnen de groep normaal begaafde kinderen bleek het Flynn-effect sterker te zijn bij de lagere IQ-scores. In tegenstelling tot Nijman (2007) werd er geen leereffect gevonden. Mogelijke verklaringen voor de gevonden resultaten en eventuele aanbevelingen voor verder onderzoek komen ter discussie.

Keywords: Flynn-effect, intelligentie, normaal begaafd, kinderen, verstandelijk beperkten, SON-tests

Inleiding

Onderzoeken naar en opvattingen over intelligentie hebben de afgelopen eeuw een grote ontwikkeling doorgemaakt. Zowel over de definitie als de aard van intelligentie bestaan verschillende opvattingen (De Zeeuw, Dekker, & Resing, 2004). Intelligentie wordt door Wechsler (1974, in Kievit, Tak, & Bosch, 2002) gedefinieerd als de algemene capaciteit van een individu om de wereld om hem heen te begrijpen en ermee om te gaan. Intelligentie wordt hier gezien als een globale entiteit, die door meerdere dingen bepaald wordt en meerdere aspecten heeft. De eerste succesvolle intelligentietest kwam in 1905 tot stand door Binet en Simon. Deze intelligentietest was bedoeld om kinderen met een te laag cognitief niveau voor het lager onderwijs te detecteren (Ingalls, 1978).

Flynn-effect

Bij de bespreking van intelligentietests is het Flynn-effect een belangrijk verschijnsel. Eerder longitudinaal onderzoek liet zien dat de intelligentie nagenoeg een mensenleven lang constant blijft. James Flynn (1984, 1987, 1998, 1999, 2000) maakte met zijn onderzoeken echter duidelijk dat de intelligentie gemeten met eenzelfde test en dezelfde normgegevens, niet zo constant is (Pesch & Ponsioen, 2004). Het intelligentiequotiënt blijkt per decennium ongeveer 3 punten te stijgen. Deze stijging van het IQ op gestandaardiseerde tests staat bekend als het *Flynn-effect*. Enkele recente onderzoeken rapporteerden een einde aan de stijging van het IQ of zelfs een daling van het IQ (Sundet, Berlaug, & Torjussen, 2004; Teasdale & Owen, 2005; Teasdale & Owen, 2008). Desondanks werd het Flynn-effect in de meeste onderzoeken gerepliceerd (Lynn & Hampson, 1986; Truscott & Frank, 2001; Uterwijk, 2001; Must, Must, & Raudik, 2003; Sanborn, Truscott, Phelps, & McDougal, 2003; Pesch & Ponsioen, 2004; Nettelbeck & Wilson, 2004; Wicherts et al., 2004). Door de stijging van het IQ is het noodzakelijk om intelligentietests om de 15 tot 20 jaar te hernormeren, zodat de IQ-scores op intelligentietests in de populatie weer normaal verdeeld zijn, met een gemiddelde van 100 en een standaarddeviatie van 15 (Kanaya, Ceci, & Scullin, 2005). Hoe ouder een test is, hoe groter het gevonden Flynn-effect zal zijn. Zonder rekening te houden met het Flynn-effect, bestaat er de kans dat mensen worden over- of onderschat, terwijl de daling of stijging op scores ook een weerspiegeling kan zijn van prestaties

op de test over de tijd heen (Sanborn et al., 2003). Het Flynn-effect kan door twee dingen aan het licht komen. Bij een herhaalde meting met dezelfde test op twee verschillende meetmomenten zal het IQ toenemen. Bij het inwisselen van een oude testversie voor een nieuwere versie, zal het IQ op de recent genormeerde test lager uitvallen door een correctie van de normen (Pesch & Ponsioen, 2004).

Het Flynn-effect op gestandaardiseerde intelligentietests werd wereldwijd teruggevonden. De omvang van het Flynn-effect varieert echter wel over verschillende tests en subtests. Het grootste effect in de stijging van het IQ werd teruggevonden bij tests die ontworpen zijn om non-verbaal redeneren en abstract probleemoplossend vermogen te toetsen. Opvallend is dat de toename op IQ-scores hoger is bij *fluid intelligence*, dan bij *crystalized intelligence* (Wicherts et al., 2004). Fluid intelligence bevat aspecten van de intelligentie die in een persoon vastliggen, terwijl crystalized intelligence cognitieve vaardigheden bevat die door ervaringen of leren zijn bepaald (Kievit, Tak, & Bosch, 2002). In de literatuur wordt performale intelligentie vooral als fluid intelligence gezien, terwijl verbale intelligentie wordt gezien als crystalized intelligence. Op de Wechsler intelligentietests treedt het Flynn-effect sterker op voor het performale dan voor het verbale IQ (Nettelbeck & Wilson, 2004; Wicherts et al., 2004).

Over zowel de oorzaak als aard van het Flynn-effect heerst nog onduidelijkheid. Is het zo dat mensen daadwerkelijk slimmer worden of is het Flynn-effect slechts een artefact? Flynn (1984) gaat er vanuit dat het stijgende IQ niet betekent dat mensen werkelijk intelligenter worden. Hij concludeert dat een IQ-test voornamelijk aangeeft hoe goed iemand is in het oplossen van abstracte problemen, een methodologisch artefact. Veel onderzoekers gaan echter wel uit van een daadwerkelijke stijging van de intelligentie (Greenfield, 1998; Lynn, 1990, 1998). Voor het bestaan van een reële normatieve IQ-stijging over tijd bestaan verschillende verklaringen. Veel genoemde verklaringen zijn: de verbeterde kwaliteit van de omgeving, betere voeding, betere gezondheid, kleinere gezinnen, verbeterde medische zorg, scholings- en technologische vooruitgang en genetische effecten (Nettelbeck & Wilson, 2004; Mingroni, 2007). De genetische oorzaak voor het Flynn-effect wordt aangeduid met de term *heterosis*, maar ook wel met de term hybride groeikracht. Heterosis is een genetisch effect dat voortkomt uit het feit dat mensen uit verschillende genetische subpopulaties zich samen voortplanten (Mingroni, 2007). Het effect bij deze hybride groeikracht is dat de nakomeling het gemiddelde van een of meer eigenschappen van de beide ouders overtreft. Een andere oorzakelijke hypothese voor het Flynn-

effect is die van Greenfield. Blootstelling aan frequente en intensieve stimuli in de vorm van televisie, computers, videospelletjes en andere grafische representaties, zouden resulteren in een verbetering van visuele en spatiële vaardigheden. Het zou daardoor niet gaan om een stijging van het IQ, maar om een verandering in cognitieve vaardigheden (Bradmetz & Mathy, 2006).

Verstandelijke beperking

De intelligentie speelt niet alleen een belangrijke rol bij de diagnostische evaluatie van normaal begaafde kinderen, maar ook bij kinderen en volwassenen met een verstandelijke beperking. Het intelligentiequotiënt is een van de belangrijkste determinanten bij de screening, classificatie en behandeling voor mensen met een verstandelijke beperking (Rosen, Stallings, Floor, & Nowakiwska, 1968). Het IQ wordt hierbij gebruikt om de mogelijkheden en beperkingen van een persoon met een verstandelijke beperking te bepalen. Het gebruik van intelligentietests voor mensen met een verstandelijke beperking kan echter problemen met zich meebrengen. Het testmateriaal, de instructies en het moeilijkheidsniveau van de opgaven zijn vaak niet in overeenstemming met de mogelijkheden van mensen met een verstandelijke beperking. Bovendien zijn de normeringen bij een IQ-score onder de 70 minder betrouwbaar. Een IQ-score van 50 – 60 wordt daarom vaak als ondergrens voor afname gehanteerd (Tellegen, 2002a). Mensen met een verstandelijke beperking scoren juist vaak onder deze grens.

Door de Amerikaanse Associatie voor Mentale Retardatie (AAMR) wordt de volgende definitie van een verstandelijke beperking gegeven: “Een verstandelijke beperking verwijst naar een substantiële beperking in het huidige functioneren. Het wordt gekenmerkt door een significant beneden gemiddeld intellectueel functioneren (een IQ-score beneden de 70), samengaan met gerelateerde beperkingen in twee of meer van de volgende gebieden: communicatie, zelfverzorging, zelfstandig wonen, sociale vaardigheden, gebruik van de maatschappij, zelfsturing, gezondheid en veiligheid, academische vaardigheden, vrije tijd en werk. Een verstandelijke beperking moet optreden vóór de leeftijd van 18 jaar” (AAMR, in Resing & Blok, 2002). Aan de hand van IQ-scores worden verschillende gradaties van een verstandelijke beperking onderscheiden. Een IQ-score van 50 – 69 wijst op een lichte verstandelijke beperking, waarbij de mentale leeftijd ligt tussen de 7 en de 11 jaar. Een IQ-score van 35 – 49 wijst op een matige verstandelijke beperking, waarbij de mentale leeftijd ligt tussen

de 4 en de 7 jaar. Een IQ-score van 20 – 34 wijst op een ernstige verstandelijke beperking, waarbij de mentale leeftijd ligt tussen de 2 en de 4 jaar. Een IQ-score onder de 20 betreft een diepe verstandelijke beperking, waarbij de mentale leeftijd tot 2 jaar loopt (Resing & Blok, 2002).

Mensen met een verstandelijke beperking functioneren mentaal en emotioneel niet op het niveau dat gelijk is aan hun kalenderleeftijd. Het gebruik van intelligentietests bij deze groep mensen moet dan ook afgestemd worden op de leeftijd van het mentaal en emotioneel functioneren. Daarom wordt er bij deze groep mensen geen IQ vastgesteld, maar een referentieleeftijd. De referentieleeftijd, ook wel mentale leeftijd, is de leeftijd waarop met de verkregen ruwe score een IQ van 100 behaald zou worden (Winkel, 1999). Bij de referentieleeftijd wordt niet gekeken naar hoe de prestaties afwijken van leeftijdgenoten, maar juist met welke leeftijd de prestaties overeenkomen. Op deze manier kunnen mensen met een verstandelijke beperking goed met anderen vergeleken worden (Tellegen, 2002a). Mensen met een verstandelijke beperking kunnen bepaalde strategieën niet gebruiken, die normaal begaafde mensen van dezelfde leeftijd wel kunnen hanteren (Bray, 1987). Over het algemeen hebben veel verstandelijk beperkten ook problemen met de taalvaardigheid en taalontwikkeling (Blount, 1968). De SON-tests hebben dan ook grote voordelen voor de diagnostiek bij mensen met een verstandelijke beperking. De voordelen zijn het niet-verbale karakter, afwezigheid van testen op kennis, geen tijdslimieten, adaptieve procedures waardoor niet te moeilijk wordt getest en het geven van feedback tijdens de afname (Tellegen, 2002a).

De SON-tests

Intelligentietests behoren tot de meest gebruikte instrumenten voor de psychologische diagnostische evaluatie van kinderen (Tellegen & Laros, 1993). Het Flynn-effect is veelvuldig onderzocht met behulp van de Wechsler-schalen. De Wechsler-schalen zijn goede intelligentietests, maar ze doen een zwaar beroep op taal. Het meten van de intelligentie of het Flynn-effect is daarom moeilijker bij kinderen met taalproblemen, zoals autochtonen of doven, of mensen met een verstandelijke beperking. Non-verbale intelligentietests hebben in een dergelijke situatie de voorkeur. Er zijn al non-verbale intelligentietests, zoals de RAVEN of TONI (Test of Nonverbal Intelligence). Deze tests hebben echter een specifieke inhoud, waardoor de resultaten

niet gegeneraliseerd kunnen worden naar een breder gebied van de intelligentie. De SON-tests (Snijders-Oomen Niet-verbale intelligentietest) onderzoeken wel een breed spectrum van de intelligentie, zonder een beroep te doen op taal (Tellegen & Laros, 1993). In 1943 promoveerde Nan Snijders-Oomen op de Snijders-Oomen niet-verbale intelligentietest. In 1958 werd deze test herzien en genormeerd voor zowel dove als horende kinderen. De SON 2,5 – 7 werd in 1996 herzien tot de SON-R 2,5 – 7 (Van Drunen, 2003). De SON-tests zijn instrumenten die individueel bij kinderen afgenomen kunnen worden. De tests zijn kindvriendelijk en meer geschikt voor kinderen die snel gedemotiveerd raken (Tellegen, 1993). Tijdens de afname van de tests kan er namelijk feedback worden gegeven over de antwoorden van de participant (Tellegen, Winkel, Wijnberg-Williams, & Laros, 1998). Het materiaal van de SON-tests ziet er aantrekkelijk uit en door de afbreekregels hoeft de participant niet onnodig veel moeilijke opgaven te maken (Wijnands, 1997).

Het Flynn-effect bij mensen met een verstandelijke beperking

Flynn (1998) gaat er vanuit dat het Flynn-effect op alle IQ-niveaus even groot is. Spitz (1983) vond in zijn onderzoek bij verstandelijk beperkten een significant Flynn-effect bij de vergelijking van de WISC met de WISC-R. Veel onderzoekers vonden echter afwijkende Flynn-effecten bij mensen met lagere IQ-niveaus. Het grootste Flynn-effect werd meestal gevonden in de lagere IQ-niveaus (Lynn & Hampson, 1986). Ook Bolen, Aichinger, Hall en Webster (1995) vonden bij de vergelijking van de WISC-R en de WISC-III bij verstandelijk beperkte kinderen, dat het Flynn-effect het grootst was bij de ernstig verstandelijk beperkten. Spitz (1989) daarentegen gaf aan dat het Flynn-effect kleiner was, naarmate men meer boven of onder het gemiddelde IQ scoorde. Bovendien werden de intelligentietests bij verstandelijk beperkten vaak toegepast aan de hand van de kalenderleeftijd. Bij verstandelijk beperkten kan echter beter de referentieleeftijd gehanteerd worden om een duidelijk beeld te krijgen van het niveau van cognitief functioneren. Onderzoek naar het Flynn-effect bij mensen met een verstandelijke beperking bracht dus uiteenlopende resultaten naar voren. Een verklaring zijn wellicht de verbale tekorten van mensen met een verstandelijke beperking. In vergelijking met de normale populatie hebben verstandelijk beperkten vaak een specifiek tekort aan verbale en/of talige vaardigheden (Blount, 1968). De

resultaten op verbale tests kunnen verschillen van de resultaten op non-verbale tests (Grossman, 1983).

Veel onderzoek naar het Flynn-effect bij mensen met een verstandelijke beperking is uitgevoerd met behulp van verbale intelligentietests, zoals de Wechsler-schalen. Vanwege bovenstaande tekorten van mensen met een verstandelijke beperking verdienen non-verbale intelligentietests de voorkeur. Voor zover bekend werd in voorgaand onderzoek naar het Flynn-effect bij verstandelijk beperkten geen gebruik gemaakt van de SON-tests of andere non-verbale tests, met uitzondering van het onderzoek van Nijman (2007).

Het Flynn-effect bij normaal begaafde kinderen

Resing en Nijland (2002) onderzochten het Flynn-effect bij kinderen van gemiddeld tot hoog opgeleide ouders, door in het jaar 2000 verzamelde LDT-scores (Leidse Diagnostische Test) te vergelijken met scores van de normeringsteekproef uit 1975. Zij vonden een toename van gemiddeld 10 tot 12 IQ-punten in de door Flynn voorspelde richting. Resing en Tunteler (2007) gaven aan dat de omvang van het Flynn-effect mogelijk ook samenhangt met de sociaal-economische status van de participant. Zij maakten hierbij gebruik van dezelfde normeringsteekproef als Resing en Nijland (2002). Bij kinderen met een hoge sociaal-economische status vonden zij een groter Flynn-effect (18 punten), dan bij kinderen met een gemiddelde sociaal-economische status (8 punten).

In een Vlaamse studie werd het Flynn-effect onderzocht door scores op de WISC-R^{NL} uit 1980 te vergelijken met scores op de WISC-III^{NL} uit 2000. Ongeveer tweederde van de kinderen scoorde een hoger IQ op de WISC-R^{NL} dan op de WISC-III^{NL}. Bij kinderen met een gemiddeld IQ werd geen significant Flynn-effect gevonden. In het reguliere onderwijs was er ook nauwelijks een verschil te zien tussen de scores van de intelligentiemeting met de WISC-R^{NL} en WISC-III^{NL}. Bij kinderen met een lichte verstandelijke handicap was er wel een significant gemiddeld verschil zichtbaar en ook in het Buitengewoon Lager Onderwijs (Vlaanderen) kwam een verschil in scores naar voren. Kortom, waar bij een gemiddeld IQ geen significant Flynn-effect werd gevonden, was dit wel aanwezig bij kinderen met een lager IQ of een (lichte) verstandelijke beperking (Kort et al., 2005).

Zoals eerder beschreven, verwachtte Flynn dat het effect een lineair beloop heeft. Ieder decennium zal het IQ ongeveer 3 punten stijgen. Sundet et al. (2004) betwijfelden dit. Als de optimale kwaliteit van het leven (gezondheid, opvoeding, onderwijs, etc.) is bereikt, dan zal de intelligentie wellicht niet meer 'groeien'. Het Flynn-effect blijft echter wel uitgesproken aanwezig voor lager begaafden (Kort et al., 2005).

Om de grootte van het Flynn-effect op de SON-tests te vergelijken tussen mensen met een verstandelijke beperking en normaal begaafde mensen, stelde Nijman (2007) voor dat het nuttig zou zijn een controlegroep te creëren van normaal begaafde kinderen. Nijman (2007) onderzocht het Flynn-effect bij mensen met een verstandelijke beperking met een gemiddelde referentieleeftijd tussen de 3,5 – 6 jaar. Met behulp van de SON 2,5 – 7 en de SON-R 2,5 – 7 vond zij een significant Flynn-effect. Het huidige onderzoek heeft als controle het Flynn-effect onderzocht bij normaal begaafde kinderen in de leeftijd van 4 – 6 jaar. Er is gekozen voor deze leeftijdsgroep, omdat de referentieleeftijden in de onderzoeksgroep van Nijman (2007) van 3,5 – 6 jaar liepen. Het is interessant om te kijken of het Flynn-effect naar voren komt, ondanks dat vele onderzoekers een daling of einde aan het effect rapporteren. Uit het huidige onderzoek zal naar voren komen of er bij normaal begaafde kinderen ook sprake is van een Flynn-effect of dat er een daling of einde wordt gerapporteerd. De volgende vraagstellingen komen in dit onderzoek aan de orde:

- Is er bij de vergelijking van de SON 2,5 – 7 en de SON-R 2,5 – 7 bij normaal begaafde kinderen sprake van een significant verschil tussen de resultaten op beide tests? Op basis van verschillende onderzoeken wordt een significant verschil (Flynn-effect) verwacht tussen de resultaten op beide tests bij normaal begaafde kinderen.
- Indien er een significant verschil is tussen de resultaten op de SON 2,5 – 7 en de SON-R 2,5 – 7, wordt de grootte van dit verschil dan beïnvloed door de volgorde van testafname? Er wordt verwacht dat het verschil in de resultaten op de SON 2,5 – 7 en de SON-R 2,5 – 7 wordt beïnvloed door de volgorde van testafname. Ondanks counterbalancing kunnen er leereffecten optreden. Wanneer de moeilijkste (meest recente) testversie het eerst wordt afgenomen, zal dit ertoe leiden dat de score op de tweede versie, die ook zonder leereffect een hogere score geeft, extra verhoogd wordt. Het verschil tussen beide versies wordt hier dus groter. Wanneer de makkelijkste (oudste) versie het eerst wordt afgenomen, leidt de aanwezigheid van leereffecten tot het tegengestelde, namelijk een relatief hoge score op de tweede test die niettemin lager

uitkomt dan de makkelijke versie. Het verschil tussen beide versies wordt hier dus kleiner (Nijman, 2007).

- Wordt er bij normaal begaafde kinderen een even sterk Flynn-effect gevonden als bij verstandelijk beperkten? Bij het beantwoorden van deze vraag werd gebruik gemaakt van de onderzoeksdata van Nijman (2007). De verwachting is dat er geen significant verschil gevonden wordt. Het Flynn-effect bij normaal begaafde kinderen en verstandelijke beperkten is gelijk aan elkaar.

- Welke verklaringen van het Flynn-effect blijken houdbaar (en welke juist niet), op basis van de resultaten van dit onderzoek?

Methode

Participanten

Het onderzoek werd uitgevoerd op basisschool St. Antonius Abt te Eindhoven. Voorafgaand is aan de basisschool toestemming gevraagd om het onderzoek te mogen uitvoeren. Hierbij heeft een gesprek plaatsgevonden met de coördinator van de onderbouw, waarmee afspraken zijn gemaakt over het uitvoeren van het onderzoek op de school. Op school was een groep van 100 kinderen aanwezig die in de doelgroep van het onderzoek vielen. De doelgroep van dit onderzoek bestond uit normaal begaafde kinderen in de leeftijd van 4 – 6 jaar. Vervolgens werden de ouders via een brief benaderd, waarin zij toestemming gaven voor de deelname van hun kind aan het onderzoek. Deelname aan het onderzoek werd door 30 ouders niet toegestaan. Aan het onderzoek namen 70 kinderen deel (N=70), waarvan 28 meisjes en 42 jongens. De gemiddelde leeftijd was 5.3 jaar, met een standaarddeviatie van .64 jaar. Doordat het onderzoek op een basisschool plaatsvond, is het niet gelukt om kinderen tussen de 3,5 – 4 jaar in het onderzoek mee te nemen. Hierom werd er voor gekozen om de leeftijdsgrens van 3,5 – 6 jaar te verleggen naar 4 – 6 jaar. Voor de vergelijking van het Flynn-effect bij normaal begaafde kinderen en verstandelijk beperkten werd gebruik gemaakt van de data uit het onderzoek van Nijman (2007). De groep verstandelijk beperkten in dit onderzoek bestond uit 32 personen (N=32), waarvan 21 mannen en 11 vrouwen.

Design

Er werd gekozen voor een within-subjects design. De participanten doorlopen hierbij beide tests. Op deze manier wordt het verschil in intelligentie tussen groepen als verklaring uitgesloten. Bij het gebruik van een within-subjects design worden de storende variabelen beperkt, doordat de participantvariabelen bij beide testversies gelijk zijn. Een nadeel van het within-subjects design is het optreden van leereffecten. Er werd gebruik gemaakt van counterbalancing, om ervoor te zorgen dat leereffecten geen verklaring vormden voor het gevonden verschil tussen beide tests. Bij counterbalancing worden de participanten aselekt verdeeld over twee groepen. De ene groep kreeg in het onderzoek eerst de SON 2,5 – 7 en vervolgens de SON-R 2,5 – 7 aangeboden. De andere groep kreeg eerst de SON-R 2,5 – 7 en vervolgens de SON 2,5 – 7 aangeboden. Testvolgorde werd als factor meegenomen, om aan uitspraken over het leereffect te kunnen voldoen. Beide tests werden conform de handleidingen afgenomen. Voorafgaand aan het onderzoek waren geen IQ-scores van de kinderen bekend.

Materiaal

De intelligentie werd in dit onderzoek gemeten met behulp van gestandaardiseerde non-verbale intelligentietests: de SON 2,5 – 7 (Snijders & Snijders-Oomen, 1975) en de SON-R 2,5 – 7 (Tellegen et al., 1998). De SON 2,5 – 7 bevat vijf subtests: Sorteren, Mozaïek, Combineren, Geheugen en Natekenen. De SON-R 2,5 – 7 bevat zes subtests: Mozaïeken, Categorieën, Puzzels, Analogieën, Situaties en Patronen. Zie Bijlage 1 voor een kort overzicht van de inhoud van de subtests. Afname van de SON 2,5 – 7 duurt ongeveer 50 minuten (Van Drunen, 2003) en afname van de SON-R 2,5 – 7 duurt ongeveer 50 – 60 minuten (Tellegen et al., 1998). Een bijzonder aspect van de SON-R 2,5 – 7 is het geven van feedback tijdens de afname. Nadat de participant een item had beantwoord, werd door de testleider aangegeven of dit antwoord goed of fout was. De items van de subtests zijn naar oplopende moeilijkheid gerangschikt. Hierdoor kon een instaprocedure gehanteerd worden die gerelateerd is aan de leeftijd en vaardigheden van het kind (Tellegen et al., 1998).

Ondanks dat de subtests in de twee testedities op een andere manier ingedeeld zijn, komen veel onderdelen van de SON 2,5 -7 op een aangepaste manier terug in de SON-R 2,5 – 7. Daarom

wordt er vanuit gegaan dat beide tests inhoudelijk vergelijkbaar zijn. De gemiddelde betrouwbaarheid (interne consistentie) van de SON 2,5 – 7 bedraagt .92 (Snijders & Snijders-Oomen, 1975). Tellegen et al. (1998) gaven aan dat de interne consistentie van de SON-R 2,5 – 7 gemiddeld .85 bedraagt. De test-hertest correlatie bedraagt .79. Tevens beoordeelden Evers, Van Vliet-Mulder en Groot (2000) de SON-R 2,5 – 7 op de uitgangspunten van de testconstructie, kwaliteit van het testmateriaal, kwaliteit van de handleiding, de normen, begripsvaliditeit en de criteriumvaliditeit als goed.

Procedure

De participanten werden aselekt ingedeeld in twee groepen, waarbij de ene groep eerst de SON 2,5 – 7 aangeboden kreeg en de andere groep de SON-R 2,5 – 7. De participanten werden getest in twee lege klaslokalen op basisschool St. Antonius Abt, zodat storende variabelen uit de omgeving werden geminimaliseerd. De onderzoekers haalden uit de groepen 1 en 2 steeds twee kinderen die de tests aflegden. Vervolgens werd at random bepaald of het kind de SON 2,5 – 7 of de SON – R 2,5 – 7 kreeg. Bij de volgende testafname kreeg het kind automatisch de andere versie van de SON-test. De tests werden niet direct na elkaar afgenomen, maar op twee verschillende dagen. De periode tussen de testafnames bedroeg gemiddeld 4 weken, om een leereffect zoveel mogelijk te minimaliseren. Nijman (2007) hanteerde in haar onderzoek een tijdsinterval van 7 tot 21 dagen. Om de onderzoeksvragen te kunnen beantwoorden werd zowel gebruik gemaakt van het IQ als de referentieleeftijd. De referentieleeftijd werd gebruikt om een vergelijkende analyse uit te kunnen voeren met het onderzoek van Nijman (2007).

Resultaten

Onderstaand worden de gevonden resultaten per onderzoeksvraag weergegeven.

Tabel 1. Kruistabel van gemiddelde IQ-scores behaald op de SON 2,5 – 7 en de SON-R 2,5 – 7 bij beide testvolgordes.

	IQ-score SON 2,5 – 7	IQ-score SON-R 2,5 – 7	
Testvolgorde 1: Eerst de SON 2,5 – 7	Gemiddeld 109.2 S.D. = 9.88	Gemiddeld 101.4 S.D. = 11.02	Totaal gemiddeld 105.3
Testvolgorde 2: Eerst de SON-R 2,5 – 7	Gemiddeld 115.6 S.D. = 11.75	Gemiddeld 102.9 S.D. = 11.55	Totaal gemiddeld 109.3
	Totaal gemiddeld 112.4	Totaal gemiddeld 102.2	

Door middel van een tweeweg-ANOVA werden onderzoeksvraag 1 en 2 onderzocht. Tabel 1 is een kruistabel van de gemiddelde IQ-scores behaald op de SON 2,5 – 7 en de SON-R 2,5 – 7 bij beide testvolgordes.

- Onderzoeksvraag 1: Is er bij de vergelijking van de SON 2,5 – 7 en de SON-R 2,5 – 7 bij normaal begaafde kinderen sprake van een significant verschil tussen de resultaten op beide tests? De gemiddelde IQ-score behaald op de SON 2,5 – 7 bedroeg 112.4, met een standaarddeviatie van 11.26. De gemiddelde IQ-score op de SON-R 2,5 – 7 bedroeg 102.2, met een standaarddeviatie van 11.23. Met deze ANOVA is gekeken of het in de steekproef gevonden verschil tussen de gemiddelde IQ-scores op de SON 2,5 – 7 en de gemiddelde IQ-score op de SON-R 2,5 – 7 statistisch significant is. Hieruit bleek een significant hoofdeffect van testversie ($F_{(1,68)} = 37.25$; $p = 0.00$). Op de SON 2,5 – 7 werd gemiddeld significant hoger gescoord dan op de SON-R 2,5 – 7. Het verwachte Flynn-effect werd gevonden en liep in de voorspelde richting.

- Onderzoeksvraag 2: Indien er een significant verschil is tussen de resultaten op de SON 2,5 – 7 en de SON-R 2,5 – 7, wordt de grootte van dit verschil dan beïnvloed door de volgorde van testafname? Er is onderzocht of er sprake is van een leereffect. Een leereffect houdt in dat de scores op een test toenemen als gevolg van het feit dat er de afname van een soortgelijke test aan vooraf is gegaan. In dit gedeelte van het onderzoek werd testversie (IQ-scores op de SON 2,5 – 7 en de SON-R 2,5 – 7) gebruikt als within-subjects factor en testvolgorde als between-subjects factor. Er bleek geen significant interactie-effect te zijn tussen testversie en testvolgorde ($F_{(1,78)} = 2.17$; $p = 0.15$). De volgorde van testafname was niet significant van invloed op de testresultaten. Als er gekeken wordt naar de gemiddelde IQ-scores, dan lijkt de IQ-score op de SON 2,5 – 7 hoger te zijn wanneer eerst de SON-R 2,5 – 7 wordt afgenomen. De SON-R 2,5 – 7 profiteert echter aanzienlijk minder wanneer eerst de SON 2,5 – 7

wordt afgenomen. Dit differentiële leereffect, van de moeilijke versie wordt meer geleerd dan van de makkelijke, wordt duidelijk als gekeken wordt naar de totale gemiddelden van de testvolgordes (zie ook Tabel 1). Bij testvolgorde 2 (eerst SON-R 2,5 – 7) werd gemiddeld 4 punten hoger gescoord dan bij testvolgorde 1 (eerst SON 2,5 – 7).

Tabel 2. Kruistabel van de gemiddelde referentieleeftijden behaald op de SON 2,5 – 7 en de SON-R 2,5 – 7 bij beide onderzoeksgroepen.

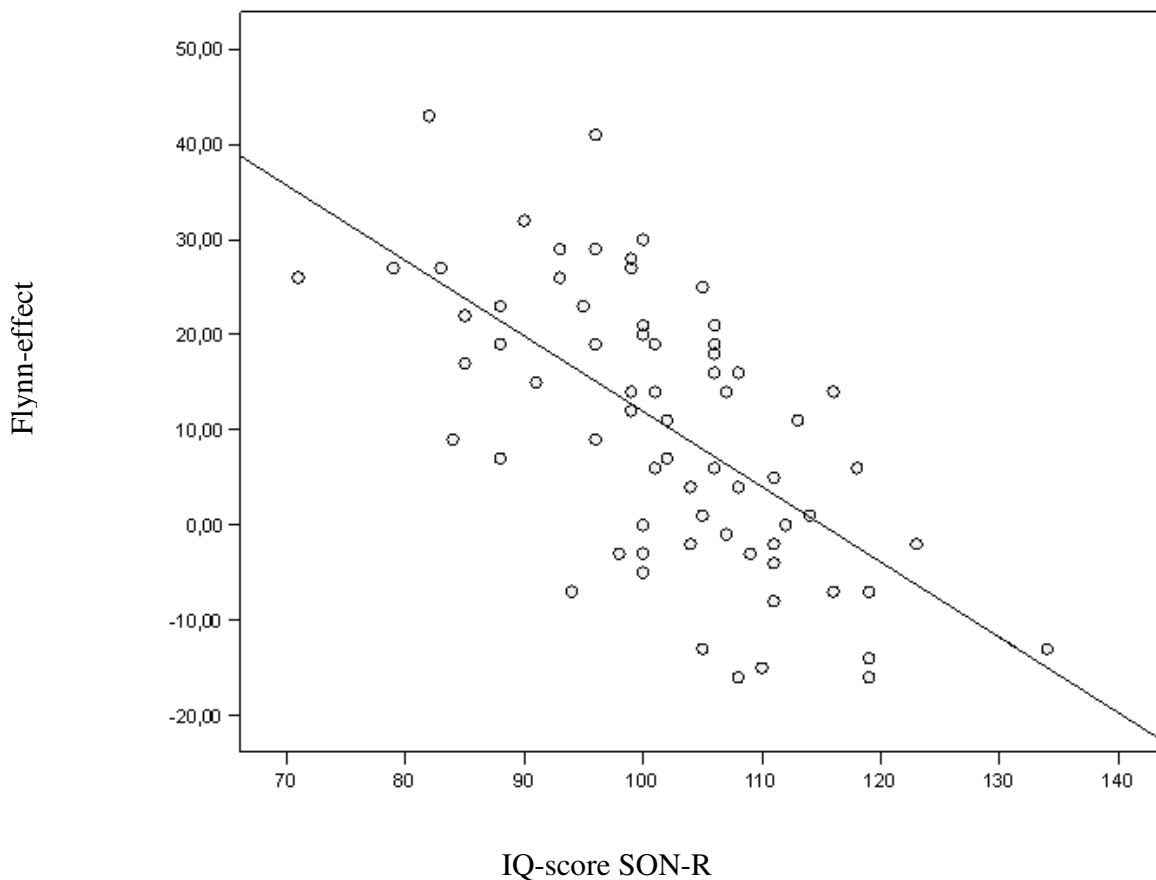
	Referentieleeftijd SON 2,5 – 7	Referentieleeftijd SON-R 2,5 – 7	
Onderzoeksgroep 1: Verstandelijk beperkten	Gemiddelde 5.1 S.D. = .17	Gemiddelde 4.5 S.D. = .18	Totaal gemiddeld 4.8
Onderzoeksgroep 2: Kinderen	Gemiddelde 5.8 S.D. = .84	Gemiddelde 5.9 S.D. = .83	Totaal gemiddeld 5.8
	Totaal gemiddeld 5.6	Totaal gemiddeld 5.5	

Onderzoeksvraag 3: Wordt er bij normaal begaafde kinderen een even sterk Flynn-effect gevonden als bij verstandelijk beperkten? Deze onderzoeksvraag werd beantwoord met behulp van de gegevens en data van Nijman (2007). In dit onderzoek kwam een significant Flynn-effect naar voren. Om een vergelijkende analyse te kunnen uitvoeren met de data van Nijman (2007) werden de IQ-scores op de SON 2,5 – 7 en de SON-R 2,5 – 7 omgezet naar referentieleeftijden. Tabel 2 geeft een overzicht van de gemiddelde referentieleeftijden, zowel voor de groep verstandelijk beperkten als voor de groep kinderen. Net als bij de IQ-scores, moet ook tussen de gemiddelde referentieleeftijden een verschil zichtbaar worden. Er is sprake van een Flynn-effect wanneer de gemiddelde referentieleeftijd op de SON 2,5 – 7 (significant) hoger is dan die op de SON-R 2,5 – 7. Bij de groep verstandelijk beperkten was dit verschil het duidelijkst zichtbaar, met een gemiddelde referentieleeftijd van 5.1 op de SON 2,5 – 7 en een gemiddelde referentieleeftijd van 4.5 op de SON-R 2,5 – 7. Bij de groep kinderen was dit verschil kleiner en was het zelfs tegengesteld van teken. De gemiddelde referentieleeftijd van 5.9 op de SON-R 2,5 – 7 was groter dan de gemiddelde referentieleeftijd van 5.8 op de SON 2,5 – 7.

Door middel van een mixed between – within subjects ANOVA werd onderzocht of de resultaten op de SON 2,5 – 7 en de SON - R 2,5 – 7 bij normaal begaafde kinderen significant verschilden van de resultaten op de SON 2,5 – 7 en de SON - R 2,5 – 7 bij verstandelijk beperkten. In dit gedeelte van het onderzoek werd testversie (gemiddelde referentieleeftijd op de

SON 2,5 – 7 en de SON-R 2,5 – 7) gebruikt als within-subjects factor en onderzoeksgroep als between-subjects factor. Er bleek een significant interactie-effect te zijn tussen testversie en onderzoeksgroep ($F_{(1,98)} = 33.21$; $p = 0.00$). De onderzoeksgroep bleek een significante invloed te hebben op de grootte van het verschil tussen de testversies (zie Tabel 2). Het Flynn-effect bij normaal begaafde kinderen was kleiner dan het Flynn-effect bij verstandelijk beperkten. Bij de onderzoeksgroep van normaal begaafde kinderen lieten de referentieleeftijden een aan de verwachting tegengesteld resultaat zien. Waar het Flynn-effect wel werd gevonden bij IQ-scores, bleek het niet aanwezig wanneer er werd gekeken naar referentieleeftijden. Verder was er een significant hoofdeffect van testversie ($F_{(1,98)} = 13.18$; $p = 0.00$). De gemiddelde referentieleeftijd van 5.6 op de SON 2,5 – 7 was significant hoger dan de gemiddelde referentieleeftijd van 5.5 op de SON-R 2,5 – 7. Dit houdt in dat wanneer werd gekeken naar de totaal gemiddelde referentieleeftijden, onderzoeksgroep buiten beschouwing gelaten, er wel sprake was van een significant Flynn-effect.

Het Flynn-effect blijkt dus sterker bij verstandelijk beperkten, de lager begaafden. Om na te gaan of dit verschil ook binnen de groep van normaal begaafde kinderen aanwezig is, werden onderstaande aanvullende analyses uitgevoerd. Daarbij werd bekeken of in de door ons onderzochte groep kinderen de grootte van het effect afhankelijk is van de intelligentie. Om die vraag te beantwoorden werd de correlatie berekend tussen het IQ, zoals gemeten door de SON-R, en het verschil tussen de twee testversies. Hieruit bleek een significante correlatie ($r = -.63$, met $p = 0.00$; $df = 68$). In Figuur 1 wordt deze samenhang grafisch weergegeven.



Figuur 1. Puntenwolk die het lineaire verloop van het Flynn-effect laat zien, met de best-passende lijn erdoorheen.

De conclusie is: hoe lager de intelligentie, hoe sterker het Flynn-effect. Aangezien niet alle kinderen de SON-R als eerste kregen, kan het leereffect van invloed zijn geweest op de score op de SON-R 2,5 – 7. Deze invloed is echter gering, omdat het gevonden leereffect minimaal was. Desondanks werd de samenhang tussen de sterkte van het Flynn-effect en IQ-scores op de SON-R ook getoetst met alleen de kinderen die eerst de SON-R kregen. Ook hier bleek sprake van een significante negatieve lineaire samenhang ($r = -.65$, met $p = 0.00$; $df = 33$). Hoe lager de IQ-scores, hoe sterker het Flynn-effect. Correlaties beschrijven echter uitsluitend een lineair verband. In de literatuur wordt vaak aangegeven dat het Flynn-effect het sterkst is in de extremen van de normale verdeling. Om na te gaan of er in dit onderzoek ook sprake was van een U-vormig verband, werden de gemiddelden nader bekeken. De IQ-scores werden hierbij in drie groepen verdeeld (70 – 90, 91 – 110, 111>). De grootte van het Flynn-effect in de lage IQ-groep

(70 – 90) bedroeg gemiddeld 22.9, in de midden-IQ-groep (91 – 110) gemiddeld 10.8 en in de hoge IQ-groep (111>) gemiddeld -6. Ook hier komt dus naar voren dat de sterkte van het Flynn-effect afhankelijk is van het IQ en daalt naarmate het IQ stijgt.

Discussie

Zoals voorspeld, is uit dit onderzoek gebleken dat er een significant verschil bestaat tussen de gemiddelde IQ-scores behaald op de SON 2,5 – 7 en de SON-R 2,5 – 7. In lijn met eerder onderzoek werd er een significant Flynn-effect gevonden (Lynn & Hampson, 1986; Truscott & Frank, 2001; Uterwijk, 2001; Must, Must, & Raudik, 2003; Sanborn, Truscott, Phelps, & McDougal, 2003; Pesch & Ponsioen, 2004; Nettelbeck & Wilson, 2004; Wicherts et al., 2004).

In tegenstelling tot de verwachting, kwam in dit onderzoek geen leereffect naar voren. Er werd echter wel een differentieel leereffect gevonden. Een eerdere afname van de SON-R 2,5 – 7 laat de gemiddelde IQ-score op de SON 2,5 – 7 stijgen. Kinderen scoorden op de ‘makkelijke’ test hoger wanneer deze vooraf werd gegaan door de ‘moeilijke’ testversie. In eerdere onderzoeken naar het Flynn-effect werd een leereffect niet gerapporteerd. Dit komt doordat men gebruikt maakte van herhaalde metingen over een periode van tientallen jaren (Resing & Tunteler, 2007; Wicherts et al., 2004; Nettelbeck & Wilson, 2004; Teasdale & Owen, 2005; Must, Must, & Raudik, 2003). In het huidige onderzoek werden daarentegen twee verschillende testversies kort na elkaar afgenomen. De afwezigheid van een leereffect kan impliceren dat men bij toekomstig onderzoek geen jaren meer hoeft te wachten bij twee vergelijkbare testafnames. Kanttekening hierbij is, dat een eerste testafname met de moeilijke versie de score op de tweede testafname met de makkelijke versie kan doen stijgen. Vervolgonderzoek is echter wenselijk.

In vergelijking met Nijman (2007) is het gevonden Flynn-effect bij normaal begaafde kinderen minder sterk aanwezig dan bij verstandelijk beperkten. Verschillende onderzoeken hebben zich gericht op de sterkte van het Flynn-effect. Hier kwam geen eenduidige verklaring naar voren. Sommige onderzoekers rapporteerden een sterker Flynn-effect bij de hogere IQ-scores (Spitz, 1989; Graf & Hinton, 1994), anderen juist bij de lagere IQ-scores (Lynn & Hampson, 1986; Teasdale & Owen, 1989; Lynn & Cooper, 1993; Colom, Lluís-Font, & Andrés-Pueyo, 2004). Aanvullende analyses in dit onderzoek toonden aan dat het Flynn-effect bij normaal begaafde kinderen het sterkst is bij de lagere IQ-scores.

Uit de resultaten van het huidige onderzoek en dat van Nijman (2007) kan geconcludeerd worden dat voor zowel normaal begaafde kinderen als mensen met een verstandelijk beperking een Flynn-effect wordt gevonden. In tegenstelling tot het onderzoek van Nijman (2007) werd in dit onderzoek echter geen leereffect gevonden. Dit kan komen door een groter tijdsinterval tussen de testafnames (7 tot 21 dagen vs. 4 weken). Personen die de tests kort na elkaar maken, profiteren meer van de vorige testafname door middel van een leereffect, dan het geval is bij personen waarbij de periode tussen beide testafnames langer is. In het huidige onderzoek was het tijdsinterval voor iedere proefpersoon gelijk, terwijl dit niet het geval was in het onderzoek van Nijman (2007). De grootte van de onderzoeksgroepen kan ook van invloed zijn. De groep van normaal begaafde kinderen bestond uit 70 proefpersonen, tegenover 32 in de groep van verstandelijk beperkten. Een dergelijke onderzoeksgroep brengt psychometrische nadelen met zich mee. Als laatste kan de aard van de onderzoeksgroep een verklaring bieden voor het leereffect. De verstandelijk beperkten zouden door een eerdere testafname bepaalde cognitieve strategieën kunnen ontdekken/leren, terwijl deze bij de normaal begaafde kinderen hoogstwaarschijnlijk reeds aanwezig zijn. Het gevonden Flynn-effect bij normaal begaafde kinderen is kleiner dan dat bij verstandelijk beperkten. Net zoals de sterkte van het Flynn-effect groter is bij lager begaafde kinderen (lagere IQ-niveaus) dan bij hoger begaafde kinderen (hogere IQ-niveaus). Aangezien er bij beide een Flynn-effect werd gevonden, kan gesteld worden dat dit effect waarschijnlijk het gevolg is van een ontwikkeling die zich in beide populaties heeft voorgedaan. Op het eerste gezicht zullen de verbeterde voeding, de toegenomen complexiteit van de omgeving, de komst van televisie en computer en genetische invloeden zich zowel bij normaal als laag begaafde mensen hebben afgespeeld. De educatie en de trend in de richting van kleinere gezinnen zullen zich wellicht bij mensen met een verstandelijke beperking anders ontwikkeld hebben dan bij normaal begaafde mensen. De educatie bij mensen met een verstandelijke beperking blijft immers vaak beperkt en aangezien de geteste personen in vrij grote leefgroepen wonen, lijkt de verklaring van het kleinere gezin bij deze mensen niet op te gaan. De verbeterde educatie en de trend in de richting van kleinere gezinnen bieden daarom wellicht geen goede verklaring voor het Flynn-effect. De verbeterde voeding, de toegenomen complexiteit van de omgeving, de komst van televisie en computer en genetische invloeden blijven ook met de aanwezigheid van het Flynn-effect bij verstandelijk beperkten bestaan als mogelijke verklaringen voor het Flynn-effect bij zowel normaal begaafden als mensen met een verstandelijke beperking.

Bovenstaande oorzaken dienen als algemene verklaringen voor het Flynn-effect. De mate van invloed van deze verklaringen is echter per subpopulatie verschillend. Sommige verklaringen lijken sterker van invloed te zijn op bepaalde groeperingen (hoog vs. laag IQ).

Het Flynn-effect blijkt dus sterker te zijn bij de lagere IQ-niveaus, zowel bij verstandelijk beperkten als normaal begaafden. De volgende verklaringen lijken hierbij de grootste invloed te hebben. Colom et al. (2004) vonden dat de toename van IQ-scores zich voornamelijk concentreerde rond de lage en middenlevels van de verdeling. Hun resultaten liggen hiermee meer in lijn met de voedingshypothese. Deze hypothese gaat er vanuit dat de verbetering van voedingsmiddelen een positief effect heeft op de lichamelijke ontwikkeling en gezondheid. Mensen worden langer en de hersenomvang neemt toe. Slechte voeding verkleint juist de hersenomvang. Onderzoeken vonden een positieve samenhang tussen intelligentie, lengte en hersenomvang. Dit leidt tot verbetering in de neurologische ontwikkeling en het functioneren van de hersenen (Lynn, 1990). Juist mensen met lagere IQ-scores profiteren volgens deze voedingshypothese het meest. Naast de voedingshypothese geven Colom et al. (2004) nog een andere verklaring voor de gevonden resultaten, namelijk de cognitieve-stimulatiehypothese. Deze hypothese gaat ervan uit dat de cognitieve vaardigheden toenemen door cognitieve stimulatie, zoals: beter geschoolde ouders, kleinere gezinnen, radio, tv, leerzaam speelgoed en betere educatie. De resultaten uit het onderzoek van Colom et al. (2004) zijn meer in lijn met de voedingshypothese, maar dit wil volgens hen niet zeggen dat de cognitieve hypothese vervalst. Deze dient ook als verklaring, maar is vanwege zijn karakter moeilijker te onderzoeken. De cognitieve hypothese is namelijk meer van toepassing op verbetering van verbale vaardigheden, terwijl de voedingshypothese meer gericht is op verbetering van visuo-spatieële vaardigheden. Naast deze twee hypothesen komt Lynn (1990) nog met een derde alternatief, die van genetische factoren. Er zijn hierbij twee mogelijkheden. De eerste heeft betrekking op de neiging van meer intelligente individuen om samen kinderen te krijgen. De tweede mogelijkheid heeft betrekking op hybride groeikracht (heterosis), zoals eerder besproken in de Inleiding. Teasdale en Owen (1989) kwamen met nog een andere alternatieve verklaring voor de grotere stijging in de lagere IQ-niveaus, namelijk veranderingen in het educatiesysteem. Het aantal jaren dat scholing wordt gevolgd is toegenomen en meer mensen behalen hun diploma. Tevens zijn er tegenwoordig meer educatieve bronnen en aanvullende scholing beschikbaar (remedial teaching, bijles en het testen van jonge kinderen).

Gezien de aard van het testinstrument (SON-tests) lijkt de voedingshypothese voor dit onderzoek en dat van Nijman (2007) een betere verklaring te geven. De voedingshypothese is namelijk meer van toepassing op een verbetering van de visuo-spatieële vaardigheden. Hiernaast profiteren zowel normaalbegaafden als verstandelijk beperkten van verbeterde voeding, hetgeen zich uit in de lichamelijke ontwikkeling. Aangezien het Flynn-effect sterker is bij verstandelijk beperkten, lijken zij meer voordeel te halen uit de verbeterde voeding.

Na deze uiteenzetting van verklaringen komt de eerder genoemde discussie opnieuw aan de orde. Is het Flynn-effect nu een reële stijging in de intelligentie of een artefact? Wanneer het Flynn-effect in de hogere IQ-niveaus aanwezig is, lijkt er sprake te zijn van een artefact. Flynn (1984) stelt dat een IQ-test voornamelijk aangeeft hoe goed iemand is in het oplossen van abstracte problemen. Mensen met een hoog IQ zijn hier van nature beter in dan mensen met een laag IQ. Er is dus sprake van een methodologisch artefact. Wanneer het Flynn-effect sterker is in de lagere IQ-niveaus, lijkt het te gaan om een reële IQ-stijging. Bovengenoemde verklaringen, oftewel verbeteringen in de omgeving, zorgen dat men zich mentaal en fysiek beter ontwikkelt. Juist mensen in de lagere IQ-niveaus profiteren hiervan, waardoor hun intelligentie echt stijgt.

Het vernieuwende aan het onderzoek van Nijman (2007) is het gebruik van een test die wat betreft moeilijkheidsniveau aansluit bij de mogelijkheden van mensen met een verstandelijke beperking. Hierdoor werden scores verkregen die vallen in het meest valide en betrouwbare gedeelte van de normaalverdeling van de scores die op de test behaald kunnen worden. Nijman (2007) heeft met haar onderzoek daarom met meer zekerheid aangetoond dat het Flynn-effect ook bij het intelligentieonderzoek bij mensen met een verstandelijke beperking een rol speelt. In de praktijk moet men bij het testen van de intelligentie van mensen met een verstandelijke beperking daarom ook rekening houden met het gevonden Flynn-effect. De aanwezigheid van het Flynn-effect bij mensen met een verstandelijke beperking heeft als gevolg dat een intelligentietest dient te worden gekozen waarvoor recente normen beschikbaar zijn. Bij het gebruik van een test met te oude normen wordt een te hoge IQ-schatting verkregen. Verstandelijk beperkten kunnen hierdoor ten onrechte als normaal begaafd worden aangemerkt. Bij herhaalde intelligentiemetingen met vernieuwde tests is het bovendien mogelijk dat een daling van het IQ wordt opgemerkt, terwijl het gevonden verschil in feite het gevolg is van de nieuwere testversie.

Het nadeel van het onderzoek van Nijman (2007) is dat er geen IQ-scores konden worden gebruikt. Er moest daarom gezocht worden naar een maat die wel bruikbaar is voor verstandelijk

beperkten. Er werd toen gekozen voor de referentieleeftijd. Aan deze maat kleven echter nadelen. Zo zijn de referentieleeftijden die op de SON 2,5 – 7 en de SON-R 2,5 – 7 te behalen zijn niet gelijk aan elkaar. Bij de SON 2,5 – 7 wordt gebruik gemaakt van referentieleeftijden van twee tot en met acht jaar, met intervallen van een kwart jaar tot de leeftijd van zes jaar en een half jaar tussen de leeftijden van zes en acht jaar. Bij de SON-R 2,5 – 7 wordt per ruwe score de referentieleeftijd in jaren en maanden gegeven en zijn de intervallen tussen de mogelijke referentieleeftijden niet gelijk. Als gevolg hiervan zal in het individuele geval vrijwel altijd een verschil gevonden worden tussen de referentieleeftijden op beide tests. Het gebruik van referentieleeftijden is hiermee ook een nadeel in het huidige onderzoek.

Een beperking van het huidige onderzoek is het gebrek aan kinderen in de leeftijd van 3,5 – 4 jaar. De groep van normaal begaafde kinderen omvatte daarmee niet de gehele mentale leeftijdsrange van de verstandelijk beperkten. Hiernaast is het vreemd dat er wel een significant Flynn-effect werd gevonden bij het gebruik van IQ-scores, maar niet bij referentieleeftijden. Dit kan komen doordat verstandelijk beperkten zich mentaal niet meer verder ‘ontwikkelen’, terwijl normale kinderen (in dezelfde mentale leeftijdsrange) zich bevinden in een levensfase waar de ontwikkeling snel verloopt. De referentieleeftijden van de verstandelijk beperkten waren gemiddeld dan ook veel lager dan die van de normaal begaafde kinderen. De gemiddelde referentieleeftijd van de verstandelijk beperkten op de SON 2,5 – 7 bedroeg 5.1 en die van de normaal begaafde kinderen 5.8. De gemiddelde referentieleeftijd van de verstandelijk beperkten op de SON-R 2,5 – 7 bedroeg 4.5 en die van de normaal begaafde kinderen 5.9. Vrijwel alle kinderen scoorden qua referentieleeftijd hoger dan hun kalenderleeftijd. Referentieleeftijden zijn dus beter toepasbaar bij mensen met een verstandelijk beperking dan bij normaal begaafde kinderen.

Toekomst

Ondanks dat een aantal onderzoeken een daling of een einde rapporteren van het Flynn-effect, werd in het huidige onderzoek samen met vele andere nog steeds een Flynn-effect gevonden. Dit impliceert de noodzaak tot het blijven hernormeren van intelligentietests. Men zou in ‘toekomstige’ onderzoeken niet alleen het Flynn-effect zelf, maar ook de sterkte ervan kunnen toetsen. Hierbij moet gekeken worden naar het feit of het Flynn-effect het grootst is in de lagere

of hogere IQ-levels of in beide extremen. Meer eenduidigheid over de sterkte van het Flynn-effect kan belangrijke implicaties met zich meebrengen. Indien de grootste stijging wordt gevonden in de lagere IQ-levels kan dit onder andere komen door bovengenoemde voedingshypothese. Verbeterde voeding heeft dan een positief effect op de lagere groepen, waardoor men rekening kan houden met de noodzaak tot goede voeding en dan voornamelijk bij deze 'lage' groeperingen.

De opzet en resultaten van het onderzoek van Nijman (2007) zijn vernieuwend. Door het gebruik van een test die aansluit bij de mogelijkheden van verstandelijk beperkten werd het Flynn-effect op een valide en betrouwbare manier aangetoond. Replicatie van dit onderzoek is dus uitermate interessant, mede doordat er maar een kleine onderzoeksgroep mogelijk was. Hiernaast kan een dergelijk onderzoek worden aangevuld met een controlegroep van normaal begaafde kinderen. Deze kan dan dienen als replicatie van de huidige onderzoeksopzet.

De voedingshypothese dient als belangrijkste verklaring in het huidige onderzoek en dat van Nijman (2007). Toekomstig onderzoek kan nagaan of deze verbeterde voeding daadwerkelijk een grote invloed heeft op een stijging van intelligentie. In de hedendaagse westerse wereld is er een overvloed aan voedsel in alle soorten en maten. Idealiter zou de invloed van de voedingshypothese dus onderzocht kunnen worden in een derde wereldland waar gebrek aan voedsel een blijvend probleem vormt.

Literatuur

- Blount, W.R. (1967). Language and the more severely retarded: a review. *American Journal of Mental Deficiency, 73*, 21-29.
- Bolen, L.M., Aichinger, K.S., Hall, C.W., & Webster, R.E. (1995). A comparison of the performance of cognitively disabled children on the WISC-R and the WISC-III. *Journal of Clinical Psychology, 51*, 89-94.
- Bradmetz, J., & Mathy, F. (2006). An estimate of the Flynn effect: Changes in IQ and subtest gains of 10-yr.-old French children between 1965 and 1988. *Psychological Reports, 99*, 743-746.
- Bray, N.W. (1987). A Symposium: Why are the mentally retarded strategically deficient? *Intelligence, 11*, 45-48.
- Colom, R., Lluís-Font, J.M., & Andrés-Pueyo, A. (2005). The generational intelligence gains are caused by decreasing variance in the lower half of the distribution: Supporting evidence for the nutrition hypothesis. *Intelligence, 33*, 83-91.
- Drunen, P. van (2003). *Tussen traditie en vernieuwing. Zestig jaar SON-tests*. Amsterdam: Boom test uitgevers.
- Evers, A., Vliet-Mulder, J.C. van, & Groot, C.J. (2000). *Documentatie van tests en testresearch in Nederland*. Assen: Van Gorcum.
- Flynn, J.R. (1984). The mean IQ of Americans: Massive gains 1932 to 1978. *Psychological Bulletin, 101*, 29-51.
- Flynn, J.R. (1987). Massive IQ gains in 14 nations. What IQ tests really measure. *Psychological Bulletin, 101*, 171-191.
- Flynn, J.R. (1998). WAIS-III and WISC-III IQ gains in the United States from 1972 to 1995. How to compensate for obsolete norms. *Perceptual and Motor Skills, 86*, 1231-1239.
- Flynn, J.R. (1999). Searching for justice – The discovery of IQ gains over time. *American Psychologist, 54*, 5-20.
- Flynn, J.R. (2000). IQ gains and fluid g. *American Psychologist, 55*, 543.
- Graf, M.H., & Hinton, R.N. (1994). A 3-year comparison study of WISC-R and WISC-III scores for a sample of special education students. *Educational and psychological measurement, 14*, 128-133.

- Greenfield, P.M. (1998). The cultural evolution of IQ. In U. Neisser (Ed.), *The rising curve* (pp. 25-66). Washington DC: American Psychological Association.
- Grossman, H.J. (1983). *Classification in mental retardation*. Washington: American Association on Mental Deficiency.
- Heiman, G.W. (1999). *Research methods in psychology*. New York: Houghton Mifflin Company.
- Ingalls, R.P. (1978). *Mental retardation: The changing outlook*. New York: John Wiley & Sons.
- Kanaya, T., Ceci, S.J., & Scullin, M.H. (2005). Age differences within secular IQ trends: An individual growth modeling approach. *Intelligence*, 33, 613-621.
- Kievit, Th., Tak, J.A., & Bosch, J.D. (2002). *Handboek psychodiagnostiek voor de hulpverlening aan kinderen*. Utrecht: De Tijdstroom.
- Kort, W., Schittekatte, M., Dekker, P.H., Verhaeghe, P., Compaan, E.L., Bosmans, M., & Vermeir, G. (2005). *WISC-III NL. Handleiding en Verantwoording*. London: Harcourt Assessment.
- Lynn, R. (1990). The role of nutrition in secular decreases in intelligence. *Personality and Individual Differences*, 11, 273-285.
- Lynn, R. (1998). In support of the nutrition theory. In U. Neisser (Ed.), *The rising curve* (pp. 207-218). Washington DC: American Psychological Association.
- Lynn, R., & Cooper, C. (1993). A secular decline in Spearman's G in France. *Learning and Individual Differences*, 5, 43-48.
- Lynn, R., & Hampson, S. (1986). The rise of national intelligence: Evidence from Britain, Japan and the U.S.A. *Personality and Individual Differences*, 7, 23-32.
- Mingroni, M.A. (2007). Resolving the IQ paradox: heterosis as a cause of the Flynn effect and other trends. *Psychological Review*, 114, 806-829.
- Must, O., Must, A., & Raudik, V. (2003). The secular rise in IQ's: In Estonia, the Flynn-effect is not a Jensen effect. *Intelligence*, 31, 461-471.
- Nettelbeck, T., & Wilson, C. (2004). The Flynn-effect: Smarter not faster. *Intelligence*, 32, 85-93.
- Nijman, E. (2007). Het Flynn-effect bij mensen met een verstandelijke beperking. Een vergelijking tussen de SON 2,5 – 7 en de SON-R 2,5 – 7. Masterthesis Universiteit van Tilburg.

- 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.
- Resing, W.C.M., & Blok, J.B. (2002). De classificatie van intelligentiescores: Een eerste stap. *De Psycholoog*, 37, 525-527.
- Resing, W.C.M., & Nijland, M.I. (2002). Een kwart eeuw onderzoek met de Leidse Diagnostische Test. *Kind en Adolescent*, 23, 42-59.
- Resing, W.C.M., & Tunteler, E. (2007). Children becoming more intelligent: Can the Flynn effect be generalized to other child intelligence tests? *International Journal of Testing*, 7, 192-208.
- Rosen, M., Stallings, L., Floor, L., & Nowakiwska, M. (1968). Reliability and stability of Wechsler IQ-scores for institutionalized mental subnormals. *American Journal of Mental Deficiency*, 73, 218-225.
- Sanborn, K.J., Truscott, S.D., Phelps, L., & McDougal, J.L. (2003). Does the Flynn-effect differ by IQ level in samples of students classified as learning disabled? *Intelligence*, 21, 145-159.
- Snijders, J.T. & Snijders-Oomen, N. (1975). *Snijders-Oomen niet-verbale intelligentieschaal, S.O.N. 2.5-7*. Groningen: Wolters-Noordhoff.
- Spitz, H.H. (1983). Intratest and intertest reliability and stability of the WISC, WISC-R and WAIS full scale IQ's in a mentally retarded population. *Journal of Special Education*, 17, 69-80.
- Spitz, H.H. (1989). Variations in Wechsler interscale IQ disparities at different levels of IQ. *Intelligence*, 13, 157-167.
- Sundet, J.M., Berlaug, D.G., & Torjussen, T.M. (2004). The end of the Flynn-effect? A study of secular trends in mean intelligence test scores of Norwegian conscripts during half a century. *Intelligence*, 32, 349-362.
- Teasdale, T.W., & Owen, D.R. (1989). Continuing secular increases in intelligence and a stable prevalence of high intelligence levels. *Intelligence*, 13, 255-262.
- Teasdale, T.W., & Owen, D.R. (2005). A long-term and recent decline in intelligence test performance: The Flynn-effect in reverse. *Personality and Individual Differences*, 39, 837-843.

- Teasdale, T.W., & Owen, D.R. (2008). Secular declines in cognitive test scores: A reversal of the Flynn Effect. *Intelligence*, 36, 121-126.
- Tellegen, P.J. (1993). A Nonverbal Alternative to the Wechsler Scales: The Snijders-Oomen Nonverbal Intelligence Tests. In *Compendium of Proceedings, First Annual South Padre Island Conference on Cognitive Assessment of Children and Youth in School and Clinical Settings* (pp. 35-49). Gevonden 25 september 2006, op www.testresearch.nl/sonro/altwchs.html.
- Tellegen, P.J. (2002a). De SON-R tests en normen voor personen met een verstandelijke handicap. Gevonden 25 september, 2006 op www.testresearch.nl/sonro/vrsthand.html.
- Tellegen, P., & Laros, J. (1993). The construction and validation of a nonverbal test of intelligence: the revision of the Snijders-Oomen tests. *European Journal of Psychological Assessment*, 9, 147-157.
- Tellegen, P.J., Winkel, M., Wijnberg-Williams, B.J., & Laros, J.A. (1998). *Handleiding en Verantwoording van de SON-R 2.5-7*. Amsterdam: Boom test uitgevers.
- Truscott, S.D., & Frank, A.J. (2001). Does the Flynn-effect affect IQ-scores of students classified as LD? *Journal of School Psychology*, 39, 319-334.
- Uterwijk, J. (2001). De Nederlandstalige bewerking van de WAIS-III. *De Psycholoog*, 36, 288-293.
- Wicherts, J.M., Dolan, C.V., Hessen, D.J., Oosterveld, P., Baal, G.C.M.van, Boomsma, D.I., & Span, M.M. (2004). Are intelligence test measurements invariant over time? Investigating the nature of the Flynn-effect. *Intelligence*, 32, 509-537.
- Wijnands, A. (1997). De SON-R tests: Verkennend onderzoek van de SON-R tests bij kinderen en volwassenen met een verstandelijke handicap. Afstudeeronderzoek Rijksuniversiteit van Groningen.
- Winkel, M. (1999). *SON-R 2,5-7: Constructie, plaatsbepaling en validering van deze algemene niet-verbale intelligentietest voor jonge kinderen*. Proefschrift Rijksuniversiteit Groningen.
- Zeeuw, J. de, Dekker, R., & Resing, W.C.M. (2004). *Algemene psychodiagnostiek 1: Testmethoden*. Leiden: Pits.

Bijlage 1

Een korte beschrijving van de subtests van de SON 2,5 – 7 en de SON-R 2,5 – 7 (Snijders & Snijders-Oomen, 1975; Tellegen, Winkel, Wijnberg-Williams, & Laros, 1998).

SON 2,5 – 7	SON-R 2,5 – 7
Sorteren:	Categorieën:
1. De participant sorteert materiaal op vorm, kleur en grootte.	1. De participant sorteert kaartjes op grond van de categorie waartoe ze behoren.
2. De participant sorteert kaartjes op grond van de categorie waartoe ze behoren.	2. De participant kiest uit vijf plaatjes op de linkerbladzijde twee plaatjes die tot dezelfde categorie behoren als de plaatjes op de rechterbladzijde.
Mozaïek:	Mozaïeken:
1. De participant legt het door de proefleider gemaakt mozaïekvoorbeeld met rode vierkantjes zelf na.	1. De participant legt het afgebeelde mozaïekvoorbeeld met rode vierkantjes zelf na.
2. De participant legt het afgebeelde mozaïekvoorbeeld met rode en gele vierkantjes zelf na.	2. De participant legt het afgebeelde mozaïekvoorbeeld met rode, gele en roodgele vierkantjes zelf na.
Combineren:	Puzzels:
1. De participant legt de ontbrekende helft bij de tekening.	1. De puzzels worden nagelegd aan de hand van een voorbeeldplaat.
2. De participant legt de puzzels met en zonder voorbeeldplaat.	2. De participant legt de puzzels zonder voorbeeldplaat.
Geheugen:	Situaties:
De participant wijst aan waar de poes door de proefleider in het poezenhuis verstopt is.	1. De participant legt de ontbrekende helft bij de tekening.
	2. De participant kiest uit vijf of zes alternatieven het ontbrekende stukje uit de tekening.
Natekenen:	Patronen:
De participant tekent door de proefleider voorgetekende figuren en tekenvoorbeelden uit het boekjes na.	De participant tekent afgebeelde figuren na, door leidstukken met elkaar te verbinden.
	Analogieën:
	1. De participant sorteert materiaal op vorm, kleur en grootte.
	2. Er staat een veranderingsvoorbeeld aangegeven van A naar B. Bij toepassing van dezelfde verandering ontstaat een van de antwoordalternatieven, die door de participant gekozen moet worden.