

NEDERLANDS INSTITUUT VAN PSYCHOLOGEN
Commissie Testaangelegenheden Nederland

Beoordeling van de SON-R 6-40, 2011

Beoordeling 2012

1. Uitgangspunten bij de testconstructie:	Goed
2. Kwaliteit van het testmateriaal:	Goed
3. Kwaliteit van de handleiding:	Goed
4. Normen:	Goed
5. Betrouwbaarheid:	Goed
6. Begripsvaliditeit:	Goed/Voldoende
7. Criteriumvaliditeit:	Goed

1.	
2.	
3.	
4.	
5.	
6.	De 'voldoende' voor Begripsvaliditeit betreft de leeftijdsgroep van 18-40 jaar, waarvoor minder validiteitsonderzoek beschikbaar is dan voor de groep van 6-17 jaar.
7.	

SON-R 6-40

Theoretische Uitgangspunten

De theoretische uitgangspunten zijn allemaal zeer helder en duidelijk uitgewerkt. De meetpretentie is niet-verbale intelligentie waarbij ook redeneertaken zijn opgenomen. De doelgroep wordt gevormd door kinderen en volwassenen in de leeftijd van 6 tot 40 jaar. De test is volgens de auteurs in het bijzonder geschikt voor personen met problemen en handicaps op spraak-taal gebied en in communicatie. Hoofdstuk 3 geeft een goede beschrijving van de constructen en de testinhoud.

Kwaliteit van het Testmateriaal

Papier-en-potloodversie

De SON-R 6-40 bevat geen open of essayvragen waarbij geen objectieve beoordeling mogelijk zou zijn. Er is een apart instructieboekje met scoringsaanwijzingen. De (nonverbale) aanwijzingen voor de geteste zijn helder. De adaptieve testprocedure wordt ook goed omschreven. Instructies goed leesbaar; prettig lettertype, goede accentuering van belangrijke aspecten in afname. Duidelijk onderscheid verbale en non-verbale instructie. Er is duurzaam materiaal gebruikt dat bij kinderen ook tegen een stootje kan.

De scoringsprocedure is duidelijk omschreven. Voor de subtest Patronen is er een correctiesleutel, die wel enig gepuzzel vraagt voordat die vlot kan worden gebruikt. De scoring van de observaties over motivatie, concentratie, samenwerking en begrip instructies (op een scoringsschaal van 1 tot 4) is niet helder en suggereert een nauwkeurigheid die er niet is. Bovendien worden matig/wisselend in tabel 4.3 samengenomen en wordt het oordeel "slecht" maar één keer gegeven (paragraaf 6.8).

Al met al uitstekend testmateriaal. Goed hanteerbare koffer; instructieboekjes van subtests in verschillende kleuren.

Computerversie

N.v.t.

Kwaliteit van de Handleiding

Er is een aparte instructiehandleiding met duidelijk onderscheid tussen verbale en non-verbale instructies waarbij de testleider precies wordt voorgeschreven wat hij/zij mag zeggen en doen. In de handleiding wordt in paragraaf 11.4 het gebruik van de test bij kinderen met handicaps beschreven en hoe daarmee om te gaan. Er worden uitgebreide samenvattingen van onderzoek gegeven.

Er worden wel voorbeelden gegeven, maar een volledig voor een casus uitgewerkte berekening en ingevuld scoreformulier ontbreekt. Interpretatie van scores wordt wel uitgewerkt maar niet aan de hand van een casus.

Een expliciete beschrijving van de vereiste deskundigheid van de testgebruiker ontbreekt evenals een verwijzing naar de kwalificatieregeling NIP. Dit onderdeel zou zeker gezien de bijzondere aard van afname (verbaal en nonverbaal) een aparte paragraaf rechtvaardigen in de handleiding. In de instructie handleiding staat in de inleiding wel kort iets vermeld: "de afname van de test is niet voorbehouden aan psychologen en (ortho)pedagogen." Uit deze omschrijving zou men kunnen afleiden dat een fysiotherapeut of andere (para)medicus met "ervaring in het testen en omgang met de te testen personen", deze test ook zou mogen afnemen.

Samenvattend: Zeer zorgvuldige handleiding met uitstekende inhoud, layout en opbouw. Het ontbreken van een volledig uitgewerkt voorbeeld (zowel kind als volwassene) en een omschrijving van de minimaal benodigde kennis en ervaring om de test betrouwbaar af te kunnen nemen worden als gemis gezien.

Over de bediening en mogelijkheden van de software wordt informatie gegeven zowel in de handleiding als op de site (www.testresearch.nl) die ook vanuit het programma gemakkelijk te vinden is. Op de site wordt sowieso veel additionele informatie gegeven over de test (en verwante tests).

NB: Twee kleine dingen die opvielen tijdens het lezen: * p. 80, paragraaf 5.5, 5 typo in "individuele"; * p. 136, regel 9, spatie tussen de en relevante ontbreekt.

Normen

De grootte van de steekproef is een wat lastig te beoordelen kwestie, omdat het continue normeringsmodel van de SON-R 6-40 niet meer met leeftijdsgroepen werkt. De auteurs bespreken dit op p. 25. Post hoc worden de gegevens in 8 groepen van gelijke omvang met $n = 241$ verdeeld. Hierbij zij opgemerkt dat het wel om de gecombineerde Nederlandse en Duitse steekproef gaat. Overigens zijn de argumenten van de auteurs om deze gecombineerde steekproef te gebruiken, overtuigend. Voor continue normering met 8 groepen is deze omvang 'goed'.

Expliciet wordt gemeld dat de pretentie is met de normen uitspraken te kunnen doen over alle in Nederland woonachtige mensen in de leeftijdsrange van 6 tot en met 40 jaar, met uitzondering van mensen met zware geestelijke en/of lichamelijke handicaps. De onderzoekers hebben veel aandacht besteed aan het samenstellen van een representatieve normgroep. De auteurs staan uitgebreid stil bij de geplande manier van gegevensverzameling: beoogde aantallen per leeftijdscategorie, selectie van onderwijsgebieden, selectie van gemeenten (voor de selectie van basisscholen en volwassenen) en VO scholen binnen die gebieden. Deze beschrijvingen zijn helder. Uit de tekst wordt echter ook duidelijk dat de selectie door omstandigheden (bijvoorbeeld door een gebrek aan animo om deel te nemen of door een gebrek aan kinderen die aan de juiste criteria voldeden) niet altijd volgens dit plan kon verlopen. Het blijft in die gevallen wat onduidelijk wat er dan wèl is gedaan. Als er bijvoorbeeld alternatieve scholen zijn gevonden, hoe zijn deze dan geselecteerd en hoe goed passen die scholen dan binnen het beoogde profiel? Over het geheel genomen lijkt de representativiteit echter in orde. De kenmerken van de uiteindelijke Nederlandse steekproef zoals leeftijd, geslacht, opleidingsniveau en autochtoon/allochtoon, staan helder beschreven en weergegeven in tabellen. De vergelijking met CBS-gegevens is goed. Over zulke kenmerken van de Duitse steekproef worden amper gegevens verstrekt. De auteurs volstaan met een verwijzing naar de handleiding van de Duitse versie van de test. Dat is jammer, want voor de normering zijn de Nederlandse en Duitse steekproeven gecombineerd. De lezer krijgt dus maar half inzicht in de kenmerken van de steekproef die is gebruikt voor de normering.

Er wordt gebruik gemaakt van een SON IQ-score waarvan de verdeling een gemiddelde heeft van 100 en een standaarddeviatie van 15. Dit is gebruikelijk voor intelligentietests. In de bijlage wordt helder beschreven welke informatie staat gepresenteerd in de normtabellen en hoe deze geïnterpreteerd moet worden.

Voor iedere leeftijdsgroep zijn gegevens verstrekt over gemiddelde, SD, min, max, scheefheid en kurtosis, zowel voor de totale test als de afzonderlijke subtests en zowel voor de ruwe scores als voor de genormeerde scores.

Zowel de verschillen in genormeerde scores voor het SON-IQ als voor de afzonderlijke subtests zijn onderzocht voor de Nederlandse versus de Duitse steekproef, mannen versus vrouwen en autochtone versus allochtone deelnemers. Er bleken alleen verschillen te bestaan op de totaalscore en op alle subtests tussen autochtone en allochtone deelnemers. De implicatie hiervan wordt niet besproken.

Er worden voldoende gegevens over de nauwkeurigheid van de genormeerde scores gegeven (standaardmeetfout en betrouwbaarheidsintervallen). Ook wordt een uitleg gegeven over hoe deze gegevens te gebruiken en wanneer men bijvoorbeeld mag spreken van significante verschillen tussen subtestscores.

Algemeen commentaar: Zeer prettig leesbaar hoofdstuk (ook over het toch complexere continue normeringsmodel) dat logisch qua opbouw is. Zeer zorgvuldige en uitgebreide normstudie (Nederlandse en Duitse steekproef). De testgebruiker wordt systematisch en grondig geïnformeerd over de normeringsstudie ($n = 1933$). Door standaard gebruik te maken van het bijgeleverd computerprogramma (wat de beoordelaars overigens niet hebben kunnen beoordelen) wordt vertekening van uitkomsten voorkomen en dit resulteert in de meest nauwkeurige uitkomsten.

Betrouwbaarheid

De betrouwbaarheid van de totaal scores is voor alle leeftijden goed (generaliseerbaarheid SON IQ > .90, zie Tabel 5.8). De betrouwbaarheid van de subtests is voor alle leeftijdsgroepen goed (voor de subtests ligt

de betrouwbaarheid tussen de .80 en .90), alleen de score op de subtest mozaïeken voor 6-jarigen is iets lager.

De test-hertestbetrouwbaarheid van de SON-IQ scores is goed, $r = .917$ ($n = 116$, gemiddeld tijdsinterval 4.2 maanden): (Goed). Deze stabiliteit is ook voor 3 leeftijdsgroepen bekeken: bij kinderen van 6 tot 11 jaar is die het laagst maar nog steeds .90. De waarden voor de subtestscores zijn iets lager, maar voldoende.

Vanwege de adaptieve afnameprocedure rijst wel een vraag over herhaalde afname. In de klinische praktijk wordt vaak na twee jaar (zeker bij kinderen) een herhaalde afname gedaan. Over zo'n herhaalde afname worden geen uitspraken gedaan. Het is dus aan de gebruiker zelf om hierin keuzes te maken.

Er zijn testleider effecten onderzocht (paragraaf 6.3), die een verschil kunnen uitmaken van 1 tot 2 IQ punten. De betrouwbaarheidsanalyses zijn (correct) verricht over gegevens van de normeringsgroep

Het betreft hier een test voor belangrijke beslissingen op individueel niveau. De auteurs wijzen hier zelf ook op in paragraaf 11.5 van de verantwoording. De betrouwbaarheid van de totaalscores (het SON-IQ) is daarom beoordeeld volgens de criteria voor belangrijke beslissingen op individueel niveau. De auteurs geven ook aan dat het verstandig is geen vergaande conclusies te verbinden aan uitkomsten op subtestniveau. Vanwege deze nuancering is de betrouwbaarheid van de subtestscores beoordeeld met de criteria voor minder belangrijke beslissingen op individueel niveau.

Begripsvaliditeit

Paragraaf 5.3. beschrijft correlaties tussen de subtest scores en de samenhang met de totaalscore. Hoofdstuk 7, 8 en 10 beschrijven de externe structuur van de SON-R 6-40. De analyses zijn uitsluitend gebaseerd op de uitkomsten van het normeringsonderzoek dat in Nederland is verricht. Decorrelaties tussen de subtests wijzen in de verwachten richting. De waarden van item-restcorrelaties worden niet gerapporteerd. De analyse van de structuur werd geëxploreerd via PCA: er is één factor die 94% van de variantie verklaart. Dit strookt met het idee dat de totaalscore de belangrijkste en meest betekenisvolle uitkomstmaat is voor deze test.

De correlaties tussen de subtest scores zijn onderzocht voor verschillende subgroepen (land, leeftijd, sekse, herkomst) en het patroon van deze correlaties is goeddeels vergelijkbaar voor de verschillende groepen.

De scores op de SON-R 6-40 zijn vergeleken met 4 veelgebruikte andere intelligentietests. Er worden betrouwbare en valide soortgenoot instrumenten gebruikt. De correlaties met soortgenoot tests zijn hoog: gemiddeld .69 met verbale schalen en .79 met performale schalen.

Verschillen in testprestaties zijn onderzocht voor verschillende relevante groepen (verschillen tussen leeftijdsgroepen, tussen mannen en vrouwen en tussen allochtone en autochtonen). De verschillen tussen leeftijdsgroepen zijn zoals verwacht mag worden. Zoals bij de meeste intelligentie tests het geval is, scoren de allochtone groepen over het algemeen wat lager dan de autochtone groep. Dit verschil valt voor een deel weg wanneer rekening gehouden wordt met het opleidingsniveau van de ouders. De prestaties van mannen en vrouwen verschillen weinig. Ook verschillen in scores tussen leerlingen op het regulier en enkele groepen in het speciaal onderwijs werden onderzocht (onder andere 57 ZMLK kinderen, 82 cluster-2 kinderen, 13 ADHD kinderen). Hypothesen omtrent specifieke verwachtingen voor deze doelgroepen zijn niet geformuleerd. Bijvoorbeeld: ADHD kinderen scoren 15 IQ punten hoger met ritalin dan zonder ritalin. Dit suggereert dat aandacht een belangrijke factor is in de testscore. Dit wordt theoretisch onvoldoende uitgewerkt. Ook bij de groep kinderen van cluster 2 zou het zinvol zijn geweest een onderscheid te maken tussen kinderen met een receptieve en expressieve taalstoornis; men zou immers verwachten dat kinderen met een receptieve taalstoornis lager scoren dan kinderen met een expressieve taalstoornis.

Gezien de leeftijdsrange van de test is het wel jammer dat er vooral onderzoek is gedaan bij groepen kinderen en veel minder bij groepen volwassenen. De onderzoeksgroepen worden summier beschreven qua eventuele demografische en psychologische variabelen.

Samenvattend: De begripsvaliditeit is uitgebreid onderzocht voor kinderen/jongeren tot en met 17 jaar met positieve resultaten. Voor deze groep komt het oordeel 'goed' in aanmerking. Voor volwassenen (18-40) is veel minder onderzoek gedaan, echter ook met positieve resultaten. Voor deze groep wordt het oordeel

Documentatie van Tests en Testresearch in Nederland
TOELICHTING BIJ DE BEOORDELING

‘voldoende’, er is meer onderzoek nodig voordat ook voor deze groep het oordeel ‘goed’ kan worden gegeven.

Criteriumvaliditeit

In hoofdstuk 7 worden relaties met onderwijsniveau en schoolprestaties besproken. De criteriumvaliditeit wordt ondersteund door de gevonden samenhang tussen het SON-IQ met verschillende indicatoren van schoolniveau en prestaties (o.a. rapportcijfer, schoolloopbaan, CITO scores). Het is jammer dat de relaties steeds alleen voor de hele groep worden gerapporteerd en er niet gekeken wordt of de samenhang tussen test en criterium systematisch anders is voor bijvoorbeeld autochtone en allochtone deelnemers. Ook hierbij is een opvallend gemis dat er geen relaties zijn onderzocht die specifiek zijn voor volwassenen (18-40 jaar), behalve het behaalde onderwijstype.

In Tabel 7.2 blijken VWO-leerlingen nog binnen de “normale range” te scoren,. Men kan zich daarbij afvragen of de SON-R 6-40 voor de kinderen met hoge begaafdheid niet een onderschatting vormt. Het aspect van kinderen met hoge begaafdheid had explicieter besproken dienen te worden, zoals ook ZML-kinderen besproken worden. Wellicht had men kunnen overwegen een groep kinderen met een aantoonbare hoge intelligentie op andere tests specifiek met de SON- R 6-40 te testen.

Het onderzoek naar de criteriumvaliditeit van de SON-R 4-60 betreft gelijktijdigheidsonderzoek (concurrent validity) en niet toekomstig gedrag.