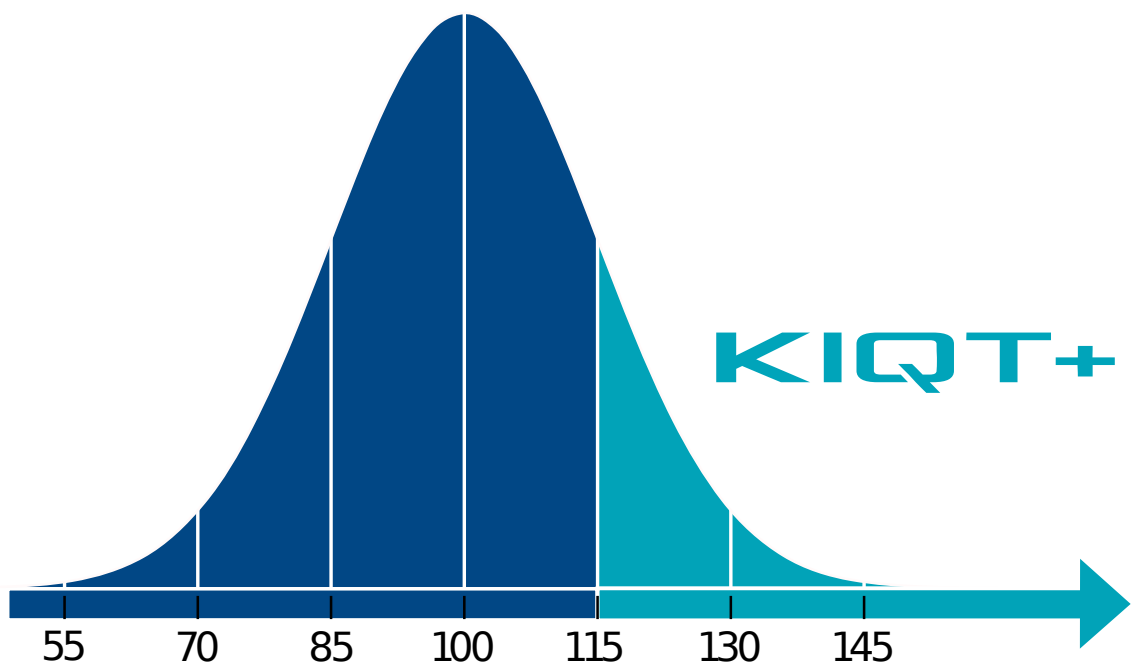


KIQQT+



Wat is de KIQT+

De KIQT+ (Kinder IQ-Test Plus) is een intelligentietest speciaal ontwikkeld voor meer- en hoogbegaafde kinderen. De test kan worden ingezet wanneer er een vermoeden van (hoog)begaafdheid is. De KIQT+ meet het meest accuraat van IQ 115 (ongeveer 159 op de 1000, of 1 op de 6 kinderen) tot IQ 170 (ongeveer 1 op de 650.000 kinderen).



De KIQT+ is gemaakt vóór en dóór bovengemiddeld intelligente mensen. De drijfveer daarvoor is dat er nog geen IQ-test beschikbaar is speciaal voor slimme kinderen, waardoor zij soms niet tot hun recht komen. Met de KIQT+ is dat verleden tijd! Hoogbegaafde en dubbel bijzondere kinderen zijn van harte welkom.

Uitgangspunten van de KIQT+

De KIQT+ heeft vier belangrijke uitgangspunten: betrouwbaar, hb-proof, eerlijk & efficiënt. Hieronder wordt per punt besproken wat we hieronder verstaan en hoe dit in de test terug te zien is.

Betrouwbaar

- De KIQT+ is gemaakt voor kinderen tussen de vijf en tien jaar oud met een (vermoedelijk) IQ tussen de 115 en 170. De KIQT+ is gemaakt om dit gebied nauwkeuriger te meten dan andere testen en heeft daarom geen last van het plafondeffect. Zie bijvoorbeeld het onderzoek bij noot [17]
- De KIQT+ maakt gebruik van een continue leeftijdsnorm welke wordt berekend door de geboortedatum en de testdatum in te geven. SCALIQ stelt centraal dat kinderen zich continu ontwikkelen, niet in drie- of viermaandelijke intervallen. Deze invalshoek voorkomt dat een kind net een dag te oud is voor een interval en het IQ daardoor een paar punten lager uitvalt. KIQT+ rekent de norm tot op de dag nauwkeurig uit.
- De KIQT+ geeft informatie over hoe goed de test het kind heeft kunnen meten, ofwel hoe goed het kind in de test past. De testdeskundige geeft het ook aan als het kind niet in de test past. Bij een dergelijk inconsistent antwoordpatroon wordt er geen IQ-score afgegeven, omdat deze niet betrouwbaar zou zijn. Gezien het uitgebreide onderzoek dat ten grondslag ligt aan de ontwikkeling van de KIQT+ , verwachten de ontwikkelaars zelden met een dergelijk antwoordpatroon in aanraking te komen.
- De KIQT+ is een objectieve test. Dit komt doordat antwoorden niet door de testafnemer hoeven te worden geïnterpreteerd om aan eventuele afbreekregels te voldoen. Dit voorkomt ook vertekeningen vanuit de (bewuste of onbewuste) verwachting van de tester over de intelligentie van het kind.
- De eerste testrondes (N=54) laten zien dat dat KIQT+ een zeer hoge interne consistentie heeft, welke een voorwaarde is voor een betrouwbare meting (Cronbach's $\alpha = 0.952$, Guttman's $\lambda-2 = 0.956$, bron: SCALIQ - intern onderzoek). Dit is ruim boven de grens van 0.9 die de COTAN (Commissie Testaangelegenheden Nederland) hanteert voor testen geschikt voor belangrijke beslissingen op individueel niveau. Volgens de EFPA (European Federation of Psychologists' Associations) richtlijnen krijgen dergelijke betrouwbaarheidscoëfficiënten het predicaat 'excellent'.

HB-proof

- De KIQT+ vermijdt ambiguïteit en heeft geen vragen met meerdere goede antwoorden. Met name open vragen kunnen er in andere intelligentietesten toe leiden dat (hoog)begaafde kinderen te moeilijk gaan denken, of langer dan nodig over een opgave doen omdat zij denken: zo makkelijk kan het antwoord toch niet zijn.
- De KIQT+ is speciaal ontworpen voor kinderen met een hoge intelligentie. Door het hoge bereik van de KIQT+ worden plafondeffecten vermeden: slecht 1 op de 650.000 kinderen zal het plafond weten te bereiken. In Nederland wonen ongeveer 1 miljoen

kinderen tussen de 5 en 10 jaar oud, slecht 1 à 2 hiervan zullen het plafond van de test dus bereiken.

- De KIQT+ kan, doordat zij specifiek is ontwikkeld voor (hoog)begaafde kinderen, een waardevol instrument zijn om te onderzoeken hoe een kind omgaat met uitdagingen die op (of boven) het niveau van het kind liggen. Dit kan inzicht bieden in werkhouding, motivatie en mogelijke faalangst.
- De KIQT+ kan omgaan met 'slordigheidsfoutjes' van hoogbegaafden. Bij een hoogbegaafd kind is er voldoende cognitieve capaciteit om de makkelijke opgaven goed te maken. De verwachting is dan ook dat er pas bij de moeilijke opgaven fouten ontstaan. Soms zie je echter dat ineens een zeer makkelijke opgave fout wordt gemaakt, bijvoorbeeld door een slordigheidsfout. Als het IQ zou worden berekend aan de hand van het totaal aantal juiste antwoorden, heeft een makkelijke of moeilijke vraag net zoveel invloed op de totale score.

Doordat de KIQT+ niet met een somscoremodel maar een Item Respons Theorie model werkt (zie voor verdere uitleg pagina 9), is de test veel minder gevoelig voor dergelijke inconsistenties in antwoordpatronen. Daardoor zal de KIQT+ het kind met slordigheidsfouten in het begin van de test een (ongeveer) gelijke score toekennen als het kind dat alle makkelijke vragen goed had beantwoord. Dit omdat uit het antwoordpatroon blijkt dat het denkniveau van het kind boven het niveau van de fout gemaakte vragen moet liggen.

Eerlijk

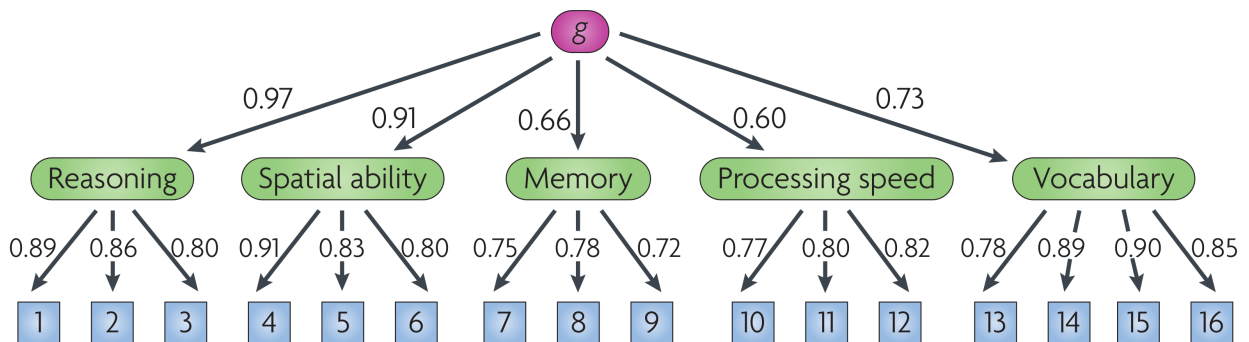
- Bij de afname van de KIQT+ wordt tijdsdruk zo veel mogelijk vermeden. Kinderen die last hebben van faalangst, prestatiedruk, concentratieproblemen of gewoon minder goed functioneren onder tijdsdruk worden zo niet onnodig benadeeld. Ook kinderen met leerstoornissen functioneren minder goed op intelligentietesten met tijdsdruk [5] en komen op de KIQT+ dus beter tot hun recht. Ook in andere landen wordt inmiddels geadviseerd om tijdsdruk te beperken bij het identificeren van hoogbegaafde kinderen [22] omdat blijkt dat dit belemmerend kan werken [29].
- De KIQT+ is niet/minder biased naar kinderen met autisme, omdat er geen gebruik wordt gemaakt van verbale items [13]. Uit onderzoek blijkt dat de prestatie van kinderen zonder autisme vergelijkbaar is voor intelligentietesten met en zonder een verbaal deel [6, 30].
- De KIQT+ is niet/minder biased naar kinderen uit lagere sociaaleconomische milieus, kinderen die minder/niet-optimaal onderwijs hebben gevolgd en kinderen met een taalachterstand of migratieachtergrond. Dit komt doordat de KIQT+ geen veronderstellingen doet over aanwezige/aangeleerde kennis van het kind.

Efficiënt

- De (sub)scores die een kind behaalt op de KIQT+ worden volledig door het SCALIQ algoritme uitgerekend. Nadat zowel de ruwe data als de uitgerekende scores door een psychometricus van SCALIQ zijn gecontroleerd, ontvangt de testafnemer een digitale rapportage. In deze rapportage staan alle benodigde gegevens voor het maken van een uitgebreid onderzoeksverslag. Dit scheelt de testafnemer veel waardevolle tijd.
- De testprocedure van de KIQT+ is eenvoudig en kan zich snel eigen gemaakt worden door de testafnemer.

Een intelligentietest die écht intelligentie meet!

Elke intelligentietest maakt gebruik van één of meer achterliggende theorieën van wat intelligentie is en hoe ze gemeten kan worden. Vandaag de dag wordt het CHC (Cattell-Horn-Carroll) model van cognitieve capaciteiten gezien als het meest omvattende en empirisch ondersteunde model voor intelligentie [19, 14]. Het model is een samensmelting van de werken van Raymond Cattell, John Horn en John Carroll [1, 16, 19, 28]. Vanwege de indrukwekkende hoeveelheid empirisch bewijs voor dit model in verschillende onderzoeksdomeinen (ontwikkelingspsychologie, neuropsychologie, etc.) wordt het model uitvoering gebruikt voor het construeren, interpreteren en categoriseren van intelligentietesten. De meeste nieuwe en gereviseerde intelligentietesten worden op het CHC model gebaseerd [12].



Figuur 1: Factoranalyse van 16 verschillende cognitieve testen afgenomen bij bijna 7000 mensen tussen de 18 en 95 jaar oud [8]. De 16 testen vallen samen in 5 'brede factoren': Redeneren, Ruimtelijk, Geheugen, Verwerkingssnelheid en Vocabulaire. Alle 5 de 'brede factoren' hebben een positieve correlatie met de overlappende *g*-factor. Figuur gebaseerd op origineel werk van Salthouse [27].

Met enige regelmaat worden factoranalyses van bekende intelligentietesten gepubliceerd die de structuur van het CHC-model ondersteunen (zie bijvoorbeeld [11]). Ook als een veelheid aan testen wordt gebruikt blijft een factoranalyse het CHC-model ondersteunen [8]. Opvallend aan de factoranalyses van testgegevens volgens het CHC-model is dat steeds de fluïde redeneerfactor de hoogste correlatie vertoont met de achterliggende ("stratum 3") *g*-factor, gevolgd door de visueel/ruimtelijke factor, de verbale factor, het werkgeheugen en de verwerkingssnelheid. Deze bevinding sluit dan ook aan bij het feit dat visuele redeneertaken (zoals de Raven Progressive Matrices) vaak worden gezien als intelligentietesten met een hoge *g*-lading [27, 14].

SCALIQ vindt dat een goede intelligentietest bovenal goed moet zijn in het meten van daadwerkelijke intelligentie. Dat betekent dat we geen eerder opgedane kennis (zoals woordenschat) toetsen met de KIQT+. Deze keuze hebben we gemaakt omdat het kennisniveau per kind verschilt. Hoewel het opdoen van kennis wel gecorreleerd is met intelligentie, is het ook afhankelijk van de omgeving en mogelijkheden. Ondanks het feit dat kinderen in Nederland vrijwel allemaal naar school gaan, zijn er grote verschillen in de mate, soort en hoeveelheid kennis en informatie die aan een kind wordt aangeboden in de jeugd. Ook bij kinderen met

een migratieachtergrond, kinderen die opgroeien in armoede en kinderen in andere nadelige situaties kunnen we er niet van uit gaan dat zij dezelfde kansen op informatie hebben gehad als gemiddelde kinderen.

Eerder opgedane kennis zegt alleen iets over het verleden, namelijk hoe goed een kind de hem/haar aangeboden kennis heeft opgenomen. Dit is ook een weergave van het onderwijsaanbod wat hij/zij tot nu toe geniet. In het bijzonder bij (hoog)begaafde kinderen is het onderwijsaanbod niet altijd passend. Een meting van de algemene intelligentie (g) heeft zeker voor deze groep kinderen een grotere voorspellende waarde voor het opnemen van kennis en vaardigheden in de toekomst. Wanneer onderwijskundige beslissingen (bijvoorbeeld plaatsing in een plusklas / op een HB-school) mede worden genomen op basis van een IQ-score, is het belangrijk dat het intellectuele potentieel van het kind goed in kaart is gebracht - onafhankelijk van eerder opgedane kennis en eerder genoten onderwijs.

Naast dat eerder opgedane kennis niet als gelijk kan worden verondersteld bij alle kinderen, is het gebruik hiervan niet nodig is voor een betrouwbare intelligentiemeting. Zoals te zien is in het hiërarchische model van intelligentie, is de algemene intelligentie (g) het doel van de intelligentiemeting. Om deze redenen kiest SCALIQ ervoor om enkel subtesten aan te bieden die een hoge g -lading bezitten en niet afhankelijk zijn van eerder opgedane kennis.

Keuze voor een intelligentietest zonder verbale items

Hoewel verbale items in de vorige eeuw nog het hoofdonderdeel waren van intelligentietests, heeft voortschrijdend inzicht uit onderzoek ervoor gezorgd dat veel intelligentietesten minder talige items bevatten. Bij SCALIQ kiezen we nadrukkelijk voor een intelligentietest met een verminderde nadruk op taal, waarbij een verbaal antwoord niet noodzakelijk is. Dit doen we om de volgende redenen:

- Verbale items maken per definitie gebruik van door het kind opgedane taalkundige kennis. De verwerving van taal heeft - net zoals de verwerving van andere kennis en vaardigheden - te maken met zowel de aanleg als de omgeving van een kind.
- De KIQT+ moet inclusief zijn voor de hele groep meer- en hoogbegaafden. Dit betekent dat deze ook aansluit bij kinderen met een migratieachtergrond, lage sociaal-economische status en leer- en ontwikkelingsstoornissen zoals dyslexie en autisme. Voor deze groepen betekent het testen van de intelligentie middels talige/verbale items vaak een grote onderschatting van hun ontwikkelingspotentieel. Onderzoek naar kinderen met autisme laat zien dat intelligentietesten met een verbaal deel deze kinderen flink kunnen onderschatten [21, 30, 20, 6]: tot wel 14 IQ-punten voor jongere kinderen [13]. Ook is bekend dat hoogbegaafde kinderen uit voorgenoemde groepen minder snel (of niet) worden geïdentificeerd.
- Taal wordt regelmatig gebruikt om een schatting te verkrijgen van de hoeveelheid kennis die een kind uit zijn omgeving heeft opgedaan. Een voorbeeld hiervan is het vragen naar woordkennis. Zoals hiervoor beschreven zijn deze opgaven erg afhankelijk van eerder opgedane kennis.
- Taal kan worden gebruikt om abstracte concepten te beschrijven, waarbij vervolgens logische redenering, classificatie of deductie op het abstracte concept tot het juiste antwoord leidt. Voorbeelden hiervan zijn het verbaal aanbieden van overeenkomsten (wat is de overeenkomst tussen ... en ...?), verbale analogieën (arm staat tot hand als been

staat tot ...?) Bij opgaven van dit type geldt vaak dat de meetpretentie wel redentatie, classificatie of deductie betreft, maar dat het succesvol maken van een dergelijke opgave toch afhankelijk is van woordkennis of taalniveau.

Uit het voorgaande volgt dat talige / verbale opgaven niet noodzakelijk zijn om de algemene intelligentie te meten. Net zoals bij andere intelligentietesten staat het redeneren, kwalificeren, sorteren en kwantificeren van abstracte concepten in de KIQT+ centraal. SCALIQ kiest er echter voor om de onderdelen/concepten aan te bieden op een manier die geen gebruik maakt van taalkundige vaardigheid of woordkennis.

De KIQT+ is echter geen test voor non-verbale intelligentie. Ook kan de KIQT+ niet gezien worden als een volledig non-verbale intelligentietest, omdat onderdelen zoals het begrijpen van de instructie nog altijd (deels) een beroep doet op taal. De KIQT+ is in dit licht het beste te zien als een algemene intelligentietest met een verminderde nadruk op eerder opgedane kennis en taalvaardigheid.

De KIQT+ past hiermee in een bredere beweging van intelligentietesten die ervoor kiezen om minder de nadruk te leggen op verbale inhoud en eerder opgedane kennis [10], zoals de NNAT (Naglieri Nonverbal Ability Test), de (C)TONI ((Comprehensive) Test of Nonverbal Intelligence) en de Raven CPM/SPM/APM (Coloured, Standard en Advanced Progressive Matrices). Ook bij het identificeren van hoogbegaafde kinderen wordt steeds vaker geadviseerd om te kijken naar die onderdelen van een intelligentietest die sterk laden op het abstracte redeneervermogen en minder beroep doen op werkgeheugen en verwerkingssnelheid. [22]

Keuze voor een intelligentietest zonder ambiguïteit

Welk dier hoort niet in onderstaande rijtje thuis:

Koe Haan Varken Schaap

Is het Varken? Omdat deze niet halal is?

Is het Koe? Omdat deze als enige een oneven aantal letters heeft?

Is het Haan? Omdat dit een vogel is?

Is het Koe? Omdat mensen allergisch kunnen zijn voor koemelk?

Of is het toch Varken? Omdat bij dit dier de huid volledig zichtbaar is?

Of is het toch Koe? Omdat deze als enige niet bij de kinderboerderij bij mij om de hoek woont?

Of is het toch Haan? Omdat die maar twee poten heeft?

Of is het toch Koe? Omdat deze vier magen heeft?

Of is het toch Schaap? Omdat deze op geen van voorgaande manieren niet in het rijtje thuis-hoort?

SCALIQ vindt dat ambiguïteit niet thuishoort in een intelligentietest. Hoewel veel (hoog)begaafde kinderen waarschijnlijk wel een idee hebben wat het goede antwoord zou kunnen zijn, vinden velen het ook moeilijk om een gokje te wagen [29]. SCALIQ maakt bij de ontwikkeling van de KIQT+ gebruik van wiskundige modellen om mogelijke relaties tussen stam (de opgave / vraag), het correcte antwoord en de incorrecte antwoorden (afleiders) inzichtelijk te maken en waar nodig items aan te passen of te verwijderen. Dit om ervoor te zorgen dat ambiguïteit, onduidelijkheid en verwarring zo min mogelijk voorkomt.

Vormgeving

Bij de ontwikkeling van de KIQT+ is specifiek aandacht besteed aan de vormgeving. De structuur en vormgeving van de test is ontwikkeld met de gedachte dat scorings- en slordigheidsfouten zo veel mogelijk vermeden kunnen worden. Daarnaast is er uitgebreid aandacht besteed aan het voorkomen van visuele overprikkeling of onduidelijkheid, door geen onnodige vorm- en/of kleurvariaties aan te brengen.

Voorbeelditem (zie volgende pagina)

- Door te allen tijde slechts 1 item tegelijk aan te bieden wordt de veronderstelling van logische verbanden over items heen vermeden. Ook kan de aanwezigheid van andere items in het zicht op deze manier geen afleiding vormen.
 - Elk item bevat zo min mogelijk overbodige informatie. Dit om visuele overprikkeling en afleiding te voorkomen.
1. Een herkenbaar itemnummer op een vaste positie zorgt voor duidelijkheid en vermindert het aantal scoringsfouten.
 2. De KIQT+ maakt gebruik van een vastgesteld, eenvoudig kleurenschema dat geschikt is voor kleurenblinden. Op deze manier zijn onderdelen van het item die alleen op het kleuraspect verschillen duidelijk differentieerbaar. Nieuwe en overbodige kleuren worden vermeden.
 3. Door het zwarte kader is direct duidelijk welke onderdelen bij de stam van het item horen.
 4. Afleiders staan op ruime afstand van de stam gepositioneerd, zodat visuele interferentie wordt geminimaliseerd. De concentratie op de stam van het item wordt bevorderd en responseliminatie wordt ontmoedigd. Onderzoek laat zien dat responseliminatie niet alleen de g -loading van een intelligentietest verlaagt, maar ook de constructvaliditeit in gevaar kan brengen. [2] [4]
 5. Afleiders staan afzonderlijk van elkaar gepositioneerd, zodat geen reeksvorming in de afleiders wordt geïmpliceerd.
 6. Een gebalanceerd aantal afleiders zorgt voor een juiste verhouding tussen verkleining van de gokkans en het risico op informatieverstrekking in de afleiders over het correcte antwoord.
 7. Het gebruik van meerdere hoogwaardige, volgens vastgestelde regels gevormde afleiders bevordert het gebruik van 'constructive matching', wat de g -loading van items verhoogt [2].
 8. Zowel het responsvak als de antwoordopties zijn asymmetrisch vormgegeven door middel van een visueel anker aan de linker onderzijde. Op deze manier wordt de suggestie dat antwoordopties geroteerd zouden kunnen worden voorkomen. Bijna een eeuw geleden wordt al in de literatuur genoemd dat hier rekening mee gehouden dient te worden [23].

9. De aanduiding van antwoordopties is op visueel niveau geïntegreerd. Dit voorkomt fouten ten opzichte van het plaatsen van de antwoordnummering boven, onder of zijdelings van de antwoordopties.

27

The puzzle consists of a 3x3 grid of stars. The colors of the stars in each row are as follows:

- Row 1: Blue, Yellow, Red
- Row 2: Yellow, Red, Blue
- Row 3: Red, Blue, and an empty box with a question mark.

Below the grid are eight options, each in a box and labeled with a letter:

- A: Blue star (rotated 45 degrees)
- B: Yellow star (rotated 45 degrees)
- C: Red star (rotated 45 degrees)
- D: Blue star (upright)
- E: Yellow star (rotated 45 degrees)
- F: Red star (upright)
- G: Blue star (rotated 45 degrees)
- H: Yellow star (upright)

Option D is the correct answer, as it maintains the alternating color pattern (Blue, Yellow, Red) in each row.

Item Respons Theorie (IRT)

Iedere IQ-test heeft een onderliggend theoretisch model om een meting (onderbouwde schatting) te doen van de intelligentie. De verwachting in de klassieke testtheorie is dat intelligentere kinderen meer vragen goed zullen beantwoorden. In de traditionele testtheorie wordt voor het score van testen gebruik gemaakt van somscores. Het aantal goed gemaakte opgaven wordt opgeteld, dit getal is dan de ruwe score van het kind. Item Respons Theorie (IRT) werkt op een fundamenteel andere manier.

Voorbeeld

De eenvoudigste manier om het verschil tussen klassieke testtheorie en Item Respons Theorie uit te leggen is aan de hand van een voorbeeld. In onderstaande tabel zien we een test van tien multiple choice vragen met oplopende moeilijkheid. De test is volgens de Klassieke Test Theorie (KTT) en volgens de Item Respons Theorie (IRT) gescoord.

	← makkelijke vragen					moeilijke vragen →					KTT score	IRT score	IRT betrouwbaarheid
	1	2	3	4	5	6	7	8	9	10			
Anne	✓	✓	✓	✓	✓	✓	✓	x	x	x	7	7	Uitstekend
Bart	✓	x	✓	✓	✓	✓	✓	x	x	x	6	6.9	Goed
Claartje	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	10	10+	Uitstekend
Dirk	-	-	-	✓	✓	✓	✓	✓	✓	✓	6	10+	Goed
Eline	✓	✓	x	✓	✓	x	x	✓	x	x	5	5	Matig
Fred	-	x	✓	x	x	✓	x	✓	✓	✓	5	-	Slecht

Anne heeft 7 van de 10 vragen goed beantwoord en krijgt daarom een score van 7 binnen de klassieke testtheorie. Als we kijken naar welke vragen zij juist heeft beantwoord, dan zien we dat zij de 7 makkelijkste vragen juist heeft beantwoord, en de 3 moeilijkste vragen niet. De (oplopende) moeilijkheid van de vragen lijkt te kloppen bij de intellectuele vaardigheden van Anne, kort gezegd lijkt het erop dat Anne goed gemeten is door deze test. De IRT-score is in dit geval hetzelfde als de KTT-score. Verder geeft de kolom 'IRT betrouwbaarheid' aan dat het antwoordpatroon van Anne uitstekend past bij de verwachting van het IRT model.

Bart heeft 6 van de 10 vragen goed beantwoord en krijgt dus een KTT score van 6. Opvallend is dat Bart één van de makkelijkste vragen, vraag 2, niet juist had. Is hij slordig geweest? Het lijkt niet realistisch dat Bart minder slim is dan Anne. Op de moeilijkere vragen, vraag 3 t/m 10, had hij immers dezelfde antwoorden als Anne! IRT houdt rekening met het patroon van de antwoorden in combinatie met de moeilijkheid van de vragen en geeft Bart een score van 6.9. De betrouwbaarheid van de IRT-score wordt nog steeds als goed bestempeld, omdat 8 van de 10 antwoorden van Bart wel voldoen aan het verwachte antwoordpatroon van een kind met een echte score van 6.9.

Claartje heeft alle vragen goed beantwoord, haar KTT-score is dus 10. Als er nog moeilijkere vragen waren geweest, had zij die dan ook goed kunnen beantwoorden? Dat weten we niet, omdat de test duidelijk te makkelijk is voor Claartje. Het IRT-model geeft in dit geval aan dat de werkelijke score waarschijnlijk boven de 10 ligt. In de kolom 'IRT betrouwbaarheid' geeft het IRT-model ook aan hier vrij zeker van te zijn, omdat Claartje alle vragen goed beantwoord heeft.

Dirk heeft de eerste drie vragen niet ingevuld. Zijn KTT-score is daarom 7. Het IRT-model rekent oningevulde vragen echter niet direct als fout. Omdat de wel ingevulde vragen allemaal goed zijn beantwoord door Dirk - en dit ook nog eens de moeilijkste vragen waren - geeft het IRT-model Dirk een score van 10+. Het IRT-model is hier wel iets minder zeker over, omdat niet duidelijk is waarom de eerste 3 vragen niet zijn ingevuld.

Het antwoordpatroon van Eline is wat inconsistent. Hoewel zij vraag 8 goed heeft beantwoord, kan dit bij een multiple choice test natuurlijk toeval zijn. Het IRT-model geeft aan dat het inconsistente antwoordpatroon reden geeft om te twijfelen aan de score van Eline.

Fred heeft de drie moeilijkste vragen goed beantwoord. De kans dat hij deze allemaal goed gegokt heeft is zeer klein. Naast dat Fred vraag 1 niet heeft ingevuld, heeft hij vraag 2, 4, 5 en 7 fout. De kans dat dit allemaal slordigheidsfoutjes zijn, is ook erg klein. Het IRT model geeft hier aan dat het antwoordpatroon van Fred niet goed genoeg te interpreteren is om een betrouwbare score te geven.



Normering

Traditioneel gezien worden intelligentietesten genormeerd door gebruik te maken van een grote aselechte normeringsgroep die over enkele variabelen zoals geslacht, leeftijd, opleidingsniveau, urbanisatiegraad en geografische locatie worden gestratificeerd. Het is belangrijk om in te zien dat deze procedure niet zomaar is ontstaan en dat elk aspect ervan zorgvuldig gevolgd dient te worden om uiteindelijk een representatieve steekproef te bemachtigen waaruit een zorgvuldige normering kan worden geconstrueerd. Bij het normeringsplan voor de KIQT+ is zorgvuldig overwogen op welke manier een correcte normering verkregen kan worden. Omdat de KIQT+ geen test is voor gemiddelde kinderen, maar nadrukkelijk bedoeld is voor kinderen vanaf 1 standaarddeviatie boven het gemiddelde, zou het strikt volgen van een traditioneel normeringsplan enkele onoverkomelijke problemen opleveren.

Bij het afnemen van de KIQT+ binnen een aselechte steekproef met kinderen zal ongeveer 16 % van de kinderen binnen de doelgroep voor de KIQT+ vallen. Dit betekent ook dat voor 84 % van de kinderen de KIQT+ (veel) te moeilijk is, en hun testgegevens dus niet in de normeringsgroep gebruikt kunnen worden. Een mogelijke oplossing voor dit probleem zou inzetten van een selectiemethode zijn, bijvoorbeeld een eerdere IQ-test. Het nadeel van het inzetten van een selectieprocedure is echter dat de normeringsgroep direct afhankelijk wordt van de gebruikte selectieprocedure en op die manier al niet meer aselechte te noemen is. Bijkomend nadeel is dat, uitgaande van ongeveer 1000 benodigde kinderen in de normeringsgroep, dit betekent dat ruim 6000 kinderen aan een dergelijke selectietest onderworpen zouden moeten worden!

Naast problemen in het aselechte samenstellen van de normeringsgroep, is op grond van de normale verdeling van IQ-scores te verwachten dat er in een aselechte samengestelde normeringsgroep vele malen meer kinderen met een IQ rond de 120 zullen voorkomen dan kinderen met een IQ van rond de 160. Dit zou ervoor zorgen dat het normeren van de test erg onevenwichtig wordt benut en er minder informatie beschikbaar komt over hoe juist de moeilijkere items tussen kinderen met een hoger IQ discrimineren. Een oplossing voor dit probleem zou natuurlijk zijn om ook over IQ-score (zoals dan gemeten door de selectietest) te stratificeren en op die manier bewust meer kinderen met een hogere IQ-score te werven om een evenwichtige normering van ook de moeilijkere testitems te garanderen. Helaas zou dit betekenen dat, zeker om genoeg kinderen met een IQ van meer dan 3 standaarddeviaties boven het gemiddelde te werven, er nog vele malen meer kinderen dan de eerder genoemde 6000 aan een selectieprocedure zouden moeten worden onderworpen. Daarnaast zijn er tot op heden weinig tot geen testen beschikbaar die accuraat de intelligentie van kinderen tot ongeveer 5 standaarddeviaties boven het gemiddelde kunnen vaststellen.

Mede om de voorgaande redenen is besloten om de KIQT+ op een andere en nieuwe manier te normeren. De normering van de KIQT+ bestaat uit een samensmelting van traditionele normeringstechnieken, wiskundige eigenschappen van de Item Respons Theorie en nieuw onderzoek naar de voorspelbaarheid van de score-distributie van cognitieve taken over leeftijd heen. Dit alles om, ondanks voor de hand liggende uitdagingen in het normeren van de KIQT+ , toch tot een accurate en betrouwbare normering te komen.

Invariantie van item parameters in Item Respons Theorie

En van de voordelen van Item Respons Theorie is dat, wanneer het gekozen model past bij de data, de item parameters invariant zijn ten opzichte van populatieverschillen. Deze eigenschap van IRT wordt ook wel 'parameter invariance', 'population invariance', 'group invariance' of 'sample invariance' genoemd [18, 9, 7, 15, 3]. Naast dat deze invariantie-eigenschap een gevolg is van de theoretische ondergrond van IRT, laat recenter onderzoek zien dat ook in de praktijk deze eigenschap stand houdt als aan enkele zwakke eisen is voldaan [26, 25, 24]. Consequentie van de invariantie eigenschap van IRT voor de normering van de KIQT+ is dat nog slechts enkele constanten empirisch bepaald dienen te worden om tot de correcte item-parameters voor de doelpopulatie te komen.

Ankeringspunten

Het bepalen van de nodige constanten ('ankeringspunten') is een onderdeel van het normeringsproces van de KIQT+ waarbij traditionele normeringstechnieken hun betrouwbaarheid en waarde laten zien. Vanzelfsprekend zullen deze steekproeven aselekt worden samengesteld en worden gestratificeerd over enkele variabelen zoals geslacht, leeftijd, opleidingsniveau (van ouders), urbanisatiegraad en geografische locatie.

Voorspelbaarheid van de score-distributie voor cognitieve taken over leeftijd heen

Eerder onderzoek van SCALIQ (bron: SCALIQ - intern onderzoek) laat zien dat de score-distributie voor cognitieve taken over leeftijd heen zich uitermate voorspelbaar gedraagt. In dit onderzoek werden normeringsgegevens van verschillende intelligentietesten, zoals de Raven Matrices (CPM, SPM, SPM+ en APM), enkele subtesten van de WISC-V-NL en de RAKIT-2 en de NNAT, van verschillende jaartallen tussen 1982-2018 en uit verschillende landen waaronder Nederland, de Verenigde Staten, Polen, China, het Verenigd Koninkrijk, Slovenië, Duitsland, Taiwan, Frankrijk, Roemenië, Argentinië, Nieuw Zeeland, Australië, India, Zwitserland en Qatar samengevoegd. In de normeringsgegevens werd hierbij gekeken naar zogenaamde 'equivalentiepunten': stel een kind van 6.5 jaar uit Engeland haalt een score van 25 op de CPM, en behaalt daarmee het 90-ste percentiel binnen de normering van 1984. Binnen diezelfde normering uit 1984 zou het kind echter slechts het 75-ste percentiel behalen indien hij of zij destijds 8 jaar oud was geweest. Het equivalentiepunt bestaat op deze manier uit 4 datapunten, twee setjes van een leeftijd en een z -score die op deze manier 'bij elkaar horen'. Nadat in het regressiemodel werd geforceerd dat als de leeftijd gelijk is de bijbehorende z -score ook gelijk moet zijn en vice versa bleek dat met 3 van de 4 datapunten de variantie in het 4e datapunt voor 97.5% verklaarbaar is (correlatie tussen voorspelling en daadwerkelijke datapunten 0.987). Omdat normeringstabellen vaak ruwe (sub)test scores bevatten en/of onderhevig zijn aan afrondingsfouten is dit onderzoeksresultaat sterk overtuigend te noemen voor de voorspelbaarheid van de score-distributie van (sub)testen voor intelligentie over leeftijd heen.

Voor de normering van de KIQT+ zullen we deze onderzoeksresultaten verder ontwikkelen en uitbreiden. Zo garanderen we met de invariantie-eigenschap van de IRT en de ankeringspunten die via de traditionele normeringstechnieken zijn verkregen een accurate en betrouwbare normering voor de KIQT+ .



Contact

Nog vragen over de KIQT+ ? Wil je meewerken aan de normering (als ouder of professional) of wil je de test wellicht gaan gebruiken? Neem contact op met ons via info@scaliq.com

Via www.scaliq.com lees je meer én kun je je inschrijven voor de nieuwsbrief. Zo blijf je op de hoogte van verdere ontwikkelingen.



Bibliografie

- [1] V. C. Alfonso, D. P. Flanagan, and S. Radwan. The Impact of the Cattell-Horn-Carroll Theory on Test Development and Interpretation of Cognitive and Academic Abilities. In *Contemporary Intellectual Assessment: Theories, Tests, and Issues*, pages 185–202. The Guilford Press, New York, NY, US, 2005.
- [2] M. E. Arendasy and M. Sommer. Reducing response elimination strategies enhances the construct validity of figural matrices. *Intelligence*, 41(4):234–243, July 2013.
- [3] F. B. Baker. *The basics of item response theory*. ERIC Clearinghouse on Assessment and Evaluation, College Park, Md., 2nd ed edition, 2001.
- [4] N. Becker, F. Schmitz, A. Falk, J. Feldbrgge, D. Recktenwald, O. Wilhelm, F. Preckel, and F. Spinath. Preventing Response Elimination Strategies Improves the Convergent Validity of Figural Matrices. *Journal of Intelligence*, 4(1):2, Feb. 2016.
- [5] C. Cornoldi, D. Giofr, A. Orsini, and L. Pezzuti. Differences in the intellectual profile of children with intellectual vs. learning disability. *Research in Developmental Disabilities*, 35(9):2224–2230, Sept. 2014.
- [6] M. Dawson, I. Soulires, M. Ann Gernsbacher, and L. Mottron. The Level and Nature of Autistic Intelligence. *Psychological science*, 18(8):657–662, Aug. 2007.
- [7] R. J. De Ayala. *The theory and practice of item response theory*. Methodology in the social sciences. Guilford Press, New York, 2009. OCLC: ocn270230503.
- [8] I. J. Deary, L. Penke, and W. Johnson. The neuroscience of human intelligence differences. *Nature Reviews Neuroscience*, 11(3):201–211, Mar. 2010.
- [9] C. DeMars. *Item response theory*. Series in understanding statistics. Measurement. Oxford University Press, Oxford ; New York, 2010. OCLC: ocn429474027.
- [10] L. S. DeThorne and B. A. Schaefer. A Guide to Child Nonverbal IQ Measures. *American Journal of Speech-Language Pathology*, 13(4):275, Nov. 2004.
- [11] S. C. Dombrowski, G. L. Canivez, and M. W. Watkins. Factor Structure of the 10 WISC-V Primary Subtests Across Four Standardization Age Groups. *Contemporary School Psychology*, 22(1):90–104, Mar. 2018.
- [12] D. P. Flanagan and P. L. Harrison, editors. *Contemporary intellectual assessment: theories, tests, and issues*. Guilford Press, New York, 2nd ed edition, 2005.
- [13] S. N. Grondhuis, L. Lecavalier, L. E. Arnold, B. L. Handen, L. Scahill, C. J. McDougle, and M. G. Aman. Differences in verbal and nonverbal IQ test scores in children with autism spectrum disorder. *Research in Autism Spectrum Disorders*, 49:47–55, May 2018.
- [14] R. J. Haier. *The neuroscience of intelligence*. Cambridge fundamentals of neuroscience in psychology. Cambridge University Press, New York, NY, 2017.
- [15] R. K. Hambleton, H. Swaminathan, and H. J. Rogers. *Fundamentals of item response theory*. Number 2 in Measurement methods for the social sciences series. Sage Publications, Newbury Park, Calif, 1991.

- [16] J. L. Horn and N. Blankson. Foundations for Better Understanding of Cognitive Abilities. In *Contemporary Intellectual Assessment: Theories, Tests, and Issues*, pages 41–68. The Guilford Press, New York, NY, US, 2005.
- [17] C. Kaplan. Ceiling Effects in Assessing High-IQ Children With the WPPSI-R. *Journal of Clinical Child Psychology*, 21(4):403–406, Dec. 1992.
- [18] W. J. v. d. Linden, editor. *Handbook of item response theory*. Chapman & Hall/CRC Statistics in the Social and Behavioral Sciences Series. CRC Press, Taylor & Francis Group, Boca Raton, 2015.
- [19] K. S. McGrew. The Cattell-Horn-Carroll (CHC) theory of cognitive abilities: Past, present and future. In D. Flanagan, & Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues-Second Edition*, pages 136–202. Guilford Press, New York, 2005.
- [20] S. D. Na and T. G. Burns. Wechsler Intelligence Scale for Children-V: Test Review. *Applied Neuropsychology: Child*, 5(2):156–160, Apr. 2016.
- [21] A.-M. Nader, V. Courchesne, M. Dawson, and I. Soulires. Does WISC-IV Underestimate the Intelligence of Autistic Children? *Journal of Autism and Developmental Disorders*, 46(5):1582–1589, May 2016.
- [22] NAGC. Use of the WISC-V for Gifted and Twice Exceptional Identification - Position Statement. Technical report, National Association for Gifted Children, Aug. 2018.
- [23] L. S. Penrose and J. C. Raven. A NEW SERIES OF PERCEPTUAL TESTS: PRELIMINARY COMMUNICATION. *British Journal of Medical Psychology*, 16(2):97–104, Nov. 1936.
- [24] A. A. Rupp. Quantifying Subpopulation Differences for a Lack of Invariance Using Complex Examinee Profiles: An Exploratory Multigroup Approach Using Functional Data Analysis. *Educational Research and Evaluation*, 11(1):71–97, Feb. 2005.
- [25] A. A. Rupp and B. D. Zumbo. A Note on How to Quantify and Report Whether Irt Parameter Invariance Holds: When Pearson Correlations are Not Enough. *Educational and Psychological Measurement*, 64(4):588–599, Aug. 2004.
- [26] A. A. Rupp and B. D. Zumbo. Understanding Parameter Invariance in Unidimensional IRT Models. *Educational and Psychological Measurement*, 66(1):63–84, Feb. 2006.
- [27] T. Salthouse. Localizing age-related individual differences in a hierarchical structure. *Intelligence*, 32(6):541–561, Dec. 2004.
- [28] W. Schneider and K. S. McGrew. The Cattell-Horn-Carroll Model of Intelligence. In *Contemporary Intellectual Assessment: Theories, Tests, and Issues*, pages 41–68. The Guilford Press, New York, NY, US, 2005.
- [29] L. K. Silverman. Assesment of giftedness. In *Handbook of giftedness in children: psychoeducational theory, research, and best practices*. Springer Science+Business Media, New York, NY, 2018.
- [30] I. Soulires, M. Dawson, M. A. Gernsbacher, and L. Mottron. The level and nature of autistic intelligence II: what about Asperger syndrome? *PLoS One*, 6(9):e25372, 2011.



KIQT+ Whitepaper

Versie 11- 3- 2020

Copyright ©2020 SCALIQ

Niets uit deze uitgave mag worden verveelvoudigd of verspreid,
door middel van druk, fotokopieën, geautomatiseerde
gegevensbestanden of op welke andere wijze ook zonder
voorafgaande schriftelijke toestemming van SCALIQ .