

Kwaliteitsbepaling  
Competentie Belevingsschaal  
voor Kinderen  
Gebruikmakend van  
Item Respons Theorie

I. J. L. Egberink (s0037826)

Begeleiders: dr. R. R. Meijer  
dr. ir. B. P. Veldkamp

Enschede, augustus 2005

Universiteit Twente  
Faculteit Gedragwetenschappen



**Universiteit Twente**  
*de ondernemende universiteit*

# Kwaliteitsbepaling Competentie Belevingschaal voor Kinderen Gebruikmakend van Item Respons Theorie

Iris J. L. Egberink  
Universiteit Twente

De kwaliteit van de subschalen Schoolvaardigheden, Sportieve vaardigheden en Gevoel van Eigenwaarde van de Competentie Belevingschaal voor Kinderen (CBSK) zijn onderzocht met Item Response Theorie (IRT); dat wil zeggen met het parametrische Graded Respons Model (GRM) en het nonparametrische Model van Monotone Homogeniteit (MMH). Wanneer het GRM wordt toegepast blijkt dat het model fit bij de data, hoewel de resultaten zijn gebaseerd op een kleine steekproef en dus voorzichtig moeten worden geïnterpreteerd. Uit de nonparametrische IRT analyse blijkt dat 'Gevoel van Eigenwaarde' de beste schaal is en dat er vraagtekens zijn bij de monotonie en unidimensionaliteit van 'Sportieve vaardigheden'. Het onderzoek naar de samenhang tussen persoonlijkheid en zelfbeeld van kinderen tussen de 8 en 12 jaar liet zien dat met name de persoonlijkheidstrekken Inadequatie en Volharding een grote invloed hebben op het zelfbeeld.

Persoonlijkheid is een centraal begrip binnen de psychologie. Dit is alleen al te zien aan het grote aantal internationale tijdschriften dat persoonlijkheid als onderwerp heeft, bijvoorbeeld "Journal of Personality", "European Journal of Personality" en "Journal of Research in Personality". Onderzoek richt zich veelal op persoonlijkheid bij volwassenen door middel van de Big Five of het ontstaan van persoonlijkheid bij baby's in de vorm van temperament (bijv. Judge, Martocchio, & Thoresen, 1997; Srivastava, John, Gosling, & Potter, 2003; Graziano, Jensen-Campbell, & Sullivan-Logan, 1998; Rothbart, Ahadi, & Evans, 2000).

In deze studie gaat de aandacht uit naar persoonlijkheid bij kinderen tussen 8 en 12 jaar. Om deze in kaart te brengen zal gebruik gemaakt worden van de Junior Nederlandse Persoonlijkheds Vragenlijst (NPV-J; Luteijn, Van Dijk, & Van der Ploeg, 1981). In literatuur over persoonlijkheid bij kinderen tussen 8 en 12 jaar lijkt de nadruk in eerste instantie niet zozeer te liggen op persoonlijkheidsontwikkeling zelf, maar meer op de ontwikkeling van de 'zelf' en het zelfconcept (Sigelman & Rider, 2003). Dit omdat het ontwikkelen van de 'zelf' gezien wordt als een belangrijk aspect bij de ontwikkeling van persoonlijkheid (Shiner & Caspi, 2003). Zelfconcept van kinderen tussen de 8 en 12 jaar wordt in deze studie in kaart gebracht door de Competentiebelevingschaal voor Kinderen (CBSK; Veerman, Straathof & Treffers, 1994).

Van oudsher wordt er bij de kwaliteitsbepaling van tests en/of vragenlijsten, zoals de NPV-J en de CBSK, gebruik gemaakt van de Klassieke Test Theorie (KTT). Daarbij worden maten als item-test correlatie en betrouwbaarheidscoëfficiënt  $\alpha$  gebruikt om de kwaliteit van een item en de gehele test te bepalen. Een nieuwe en betere methode voor het bepalen van de kwaliteit van een test of vragenlijst, die bovendien nieuwe toepassingen mogelijk maakt is Item Respons Theorie (IRT). IRT wordt vooral gebruikt in het onderwijskundig meten en wordt nu ook steeds meer toegepast in de psychologie. IRT kent verschillende voordelen ten opzichte van KTT, onder andere dat bij IRT op itemniveau gedetailleerdere informatie verkregen kan worden. Terwijl de KTT vooral uitspraken doet over de test in zijn geheel. Bovendien is IRT beter theoretisch onderbouwd en wordt er van tevoren een model

gespecificeerd om daarna te controleren of het model ook past bij de data; er wordt gekeken naar de zogeheten ‘fit’ van het model. KTT daarentegen specificeert geen model (Embretson & Reise, 2000).

IRT is dan ook bijzonder geschikt voor de kwaliteitsbepaling van psychologische tests en/of vragenlijsten. In deze studie zal IRT ter illustratie toegepast worden op de CBSK, de onderzoeksvraag daarbij is:

“Wat kan er gezegd worden over de kwaliteit van de Competentiebelevingsschaal voor Kinderen, gebruikmakend van IRT?”

Naast deze kwaliteitsbepaling zal nagegaan worden hoe de scores op de NPV-J samenhangen met de scores op de CBSK. De onderzoeksvraag luidt:

“Hoe hangt persoonlijkheid van een kind tussen 8 en 12 jaar samen met zijn/haar zelfconcept, zoals gemeten door respectievelijk de NPV-J en de CBSK?”

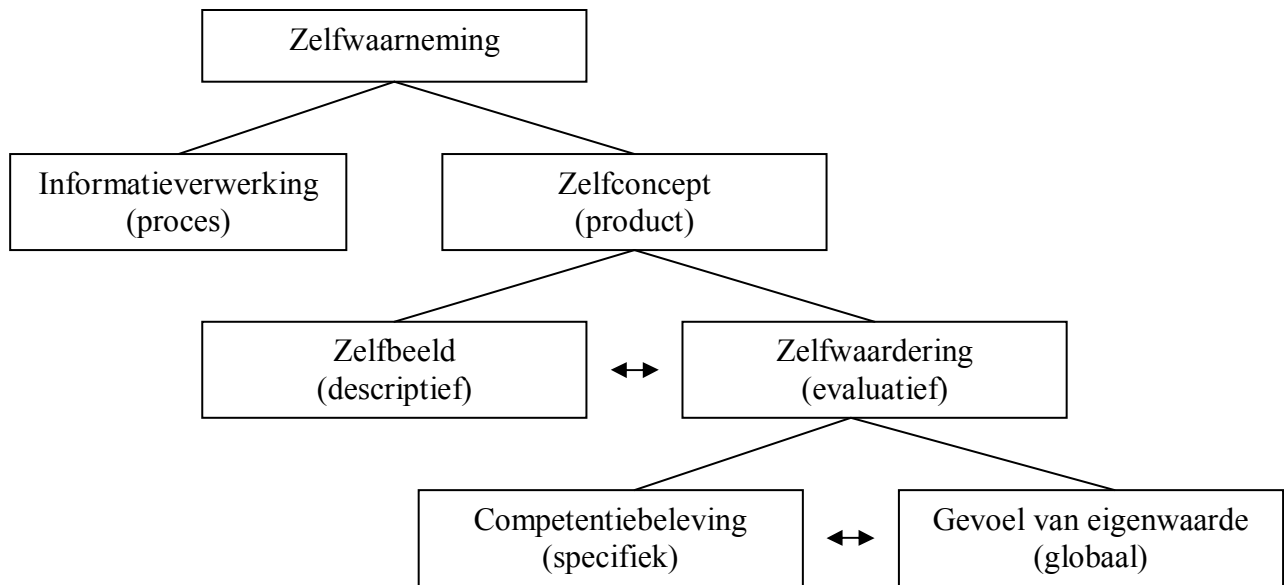
Om de vraag over de samenhang tussen persoonlijkheid en zelfbeeld bij kinderen tussen de 8 en 12 jaar beter te kunnen begrijpen, zal in dit artikel eerst ingegaan worden op hetgeen bekend is over persoonlijkheid en zelfconcept bij kinderen. Vervolgens zal de werking van IRT modellen beschreven worden, zodat de IRT analyses bij de kwaliteitsbepaling van de CBSK later in dit artikel beter begrepen worden. Na deze uitleg zullen de resultaten van de uitgevoerde analyses aan bod komen en besproken worden.

### Zelfconcept (bij kinderen)

Zelfconcept is een belangrijk begrip binnen de ontwikkelingspsychologie, met name bij kinderen en adolescenten. Een positief zelfconcept wordt gezien als een belangrijk aspect bij de ontwikkeling van kinderen (o.a. Houck & Spegman, 1999; Muris, Meesteres, & Fijen, 2003; Veerman, Ten Brink, Straathof, & Treffers, 1996; Trzesniewski, Donnellan & Robins, 2003). In dit artikel wordt met zelfconcept ‘het geheel aan denkbeelden dat een persoon ten aanzien van zichzelf heeft’ (Veerman, Straathof, Treffers, Van den Bergh, & Ten Brink, 1997, 2004) bedoeld.

Binnen de literatuur over zelfconcept en verwante begrippen is echter veel verwarring over de definities van deze begrippen (Jacobs, Bleeker, & Constantino, 2003; Veerman et al., 1997, 2004); naast een groot aantal omschrijvingen, worden dezelfde begrippen door verschillende auteurs vaak verschillend gedefinieerd. Daarom volgt nu in Figuur 1 een overzicht van de begrippen rond zelfconcept om zodoende verwarring weg te nemen en om aan te geven waar in dit artikel vanuit wordt gegaan (Veerman et al., 1997, 2004).

*Figuur 1. Overzicht begrippen rond zelfconcept*



In dit overzicht is te zien dat zelfconcept en verwante begrippen vallen onder het overkoepelende begrip ‘zelfwaarneming’. Het begrip zelfwaarneming heeft een proceskant en een productkant. De proceskant bestaat uit informatieverwerking; enerzijds het richten van de aandacht, het selecteren en verwerken van informatie (de zogenaamde basale cognitieve processen) en anderzijds bewustzijnsprocessen waardoor men ervaringen van continuïteit, distinctie en oorzakelijkheid ervaart.

Het product van zelfwaarneming is ‘zelfconcept’. Zelfconcept op zichzelf bestaat ook weer uit twee componenten, namelijk een descriptieve en een evaluatieve component. De descriptieve component is het zelfbeeld, dat wil zeggen relatief objectieve beschrijvingen van eigenschappen en prestaties. De evaluatieve component geeft vervolgens aan hoe ‘goed’ of ‘slecht’ deze eigenschappen of prestaties gevonden worden; dat wordt verstaan onder zelfwaardering.

Zelfwaardering is verder onder te verdelen in twee vormen; namelijk specifieke en globale zelfwaardering. Met specifieke zelfwaardering wordt competentiebeleving bedoeld en met globale zelfwaardering het gevoel van eigenwaarde. Competentiebeleving is ‘de evaluatie van het zelf ten aanzien van bepaalde domeinen van eigen kunnen’. Het gevoel van eigen waarde is de evaluatie van de zelf in meer algemeen opzicht als ‘persoon’.

Zelfwaardering is door deze tweedeling geen ééndimensioneel begrip, maar een multidimensioneel begrip. Hierdoor zijn de domeinen uit de specifieke zelfwaardering op te vatten als facetten van zelfwaardering en in ruimere zin als facetten van het zelfconcept (Veerman et al., 1997, 2004).

### *Constructie CBSK*

De Amerikaanse Ontwikkelingspsychologe Susan Harter ontwikkelde in 1979 op basis van de begrippen specifieke en globale zelfwaardering (competentiebeleving en gevoel van eigen waarde) de ‘Perceived Competence Scale for Children’. Zij reviseerde deze schaal in 1985 en zo ontstond het ‘Self-Perception Profile for Children’. De Nederlandse versie van deze vragenlijst is de ‘Competentiebelevingsschaal voor Kinderen’ (Veerman et al., 1997, 2004).

In deze studie zal gebruik worden gemaakt van de CBSK om vast te stellen hoe kinderen van 8 tot 12 jaar zichzelf op een aantal specifieke gebieden evalueren en hoe zij hun globaal gevoel van eigenwaarde evalueren.

### Persoonlijkheid (bij kinderen)

Een begrip dat sterk verband houdt met het zelfconcept is persoonlijkheid. Dit geldt met name voor kinderen; in de literatuur die zich richt op persoonlijkheid bij kinderen (tussen de 8 en 12 jaar) lijkt de nadruk namelijk in eerste instantie niet te liggen op persoonlijkheidsontwikkeling, maar op de ontwikkeling van de 'zelf' en het zelfconcept (Sigelman & Rider, 2003). Dit omdat het ontwikkelen van de 'zelf' gezien wordt als een belangrijk aspect van persoonlijkheid (Shiner & Caspi, 2003).

#### *Persoonlijkheidstheorieën*

In de loop van de jaren zijn er verschillende theorieën ontwikkeld met betrekking tot persoonlijkheid, zoals de psychoanalytische theorie van Freud, de neopsychoanalytische theorie van onder andere Jung en Murray, de humanistische theorie van Maslow en Rogers, de 'behavioral approach' van Skinner en de 'social learning approach' van Bandura (Schultz & Schultz, 2001; Funder, 2004). Een andere theorie over persoonlijkheid is de 'trait approach'. Deze theorie wordt steeds vaker gezien als dé theorie over persoonlijkheid (Funder, 2004).

De 'trait approach' richt zich op individuele verschillen in persoonlijkheid en gedrag én de psychologische processen erachter. Die individuele verschillen in persoonlijkheid en gedrag zijn volgens de 'trait approach' terug te vinden in kenmerkende patronen van gedachten, gevoelens en gedrag van het individu. Die kenmerkende patronen worden 'traits' / trekken genoemd.

Binnen de 'trait approach' zijn er vier basismethoden die op basis van deze trekken gedrag helpen verklaren:

1. 'many-trait approach'; de methode richt zich op een bepaald gedrag en onderzoekt zijn verbanden met zoveel mogelijk verschillende persoonlijkheidstrekken
2. 'single-trait approach'; de methode richt zich op één bepaalde trek en probeert zoveel mogelijk te weten te komen over zijn verbanden met gedrag, ontwikkelingsantecedenten en levensconsequenties
3. 'essential trait approach'; de methode die de lijst van duizenden trekken probeert terug te dringen tot trekken die er echt toe doen
4. 'typological approach'; de methode probeert typen individuen te ontdekken. Elk type wordt gekenmerkt door een bepaald patroon van trekken (Funder, 2004).

In dit artikel wordt uitgegaan van de 'essential trait approach'. Dit omdat uit deze methode de meest geaccepteerde en gebruikte oplossing voor het probleem van het reduceren van persoonlijkheidstrekken is gekomen, namelijk de Big Five factoren. Deze factoren zijn Neuroticisme, Openheid, Extraversie, Altruïsme en Consciëntieusheid. De Big Five maakt gebruik van factor analyse; een data reductie techniek (Kaplan & Saccuzzo, 2001). Factor analyse is een methode voor het identificeren van het aantal verschillende dimensies (eigenschappen, karaktertrekken), factoren genoemd, die gemeten worden door de items van een test. De correlaties tussen alle items op een test worden gebruikt om de mate te bepalen waarin elk item gerelateerd is aan elke factor (dit wordt ook wel factorlading genoemd). Op

die manier worden groepen items gevormd die hoger correleren met elkaar dan met items uit andere groepen. Zo'n groep items heet dan een factor (Dooley, 2001; Kaplan & Saccuzzo, 2001).

Factor analyse is een methode die veel wordt toegepast bij de testconstructie van persoonlijkheidsvragenlijsten. De Junior Nederlandse Persoonlijkheids Vragenlijst (NPV-J) is zo'n vragenlijst (Luteijn, Van Dijk, & Van der Ploeg, 1989).

### *Constructie NPV-J*

De NPV-J is afgeleid van de Nederlandse Persoonlijkheids Vragenlijst (NPV; Luteijn, Starren, & Van Dijk, 1979, 1985, 2001) (Luteijn et al., 1989), die op zijn beurt gebaseerd is op de California Psychological Inventory (CPI; Gough, 1957, 1987, 1996) (Luteijn, Starren, & Van Dijk, 2001). De CPI wordt gezien als een brede persoonlijkheidsvragenlijst met zijn 434 items. De vragenlijst heeft echter geen theoretische achtergrond met betrekking tot persoonlijkheid; de vragenlijst is namelijk bedoeld om zogenaamde 'folk concepts' te meten (McCrae, Costa, & Piedmont, 1993; Hakstian & Farrell, 2001). Gough verstond daaronder "variabelen die mensen in het dagelijks leven gebruiken om hun eigen gedrag en dat van anderen te begrijpen, te classificeren en te voorspellen" (Gough, 1987, p. 1).

Een ingekorte en vertaalde versie van die 434 items van de CPI werd de basis voor een nieuwe persoonlijkheidsvragenlijst, die sinds 1974 de NPV is gaan heten. Het doel van deze vragenlijst zou het meten van enige hoofddimensies van persoonlijkheid worden.

De items van de CPI werden vertaald en kritisch bekeken door verschillende beoordelaars, waarbij vooral gelet is op de helderheid van formuleren en de psychologische relevantie. Op deze wijze bleven 184 items over, waarvan 176 items verdeeld waren over 8 a priori schalen. Een a priori schaal houdt in dat de items van tevoren ingedeeld zijn in groepen op basis van logisch redeneren.

Met deze versie van de NPV is vervolgens een aantal jaren onderzoek gedaan en op basis van die gegevens is verder gekeken naar de schaalindeling. Over alle gegevens is een multiële groepsfactoranalyse uitgevoerd. Hieruit kwam naar voren dat er een aantal schalen samengevoegd moesten worden en dat een aantal items niet in de vragenlijst pasten. Op deze manier ontstonden de volgende zeven schalen van de NPV:

1. Inadequatie,
2. Sociale Inadequatie,
3. Rigiditeit,
4. Verongelijkheid,
5. Zelfgenoegzaamheid,
6. Dominantie en
7. Zelfwaardering (Luteijn, et al., 2001).

De 184 items oorspronkelijke items van de NPV vormden de basis voor de NPV-J. Allereerst werden de items omgezet in de ik-jij vorm en vervolgens voorgelegd aan kinderen van 9-14 jaar om de begrijpelijkheid van de items te beoordelen. Een aantal items kwamen hierdoor te vervallen, werden herschreven of aangevuld. De overgebleven items werden kritisch bekeken door psychologen en pedagogen en vervolgens aan kinderen voorgelegd van 9-12 jaar. Hieruit ontstonden 125 items die verdeeld waren over 7 a priori schalen.

Met deze items werd verder onderzoek gedaan en uit factoranalyse van die onderzoeksgegevens bleek dat de vragenlijst 5 onderliggende factoren meet en dat 20 items daar niet in pasten. Op deze manier ontstond de huidige versie van de NPV-J met 105 items en 5 onderliggende schalen (Luteijn et al., 1989).

## Item Response Theorie

### *De basisprincipes van IRT (modellen)*

Item Response Theorie wordt ook wel de latente trek theorie genoemd. Een fundamenteel idee is dat psychologische constructen latent, dat wil zeggen niet direct waarneembaar, zijn en dat kennis over deze constructen slechts verkregen kan worden door de zichtbare responsen van personen op een set items. IRT verklaart de responsen door het bestaan van een latente trek (ook wel theta,  $\theta$ ) aan te nemen waarop personen en items een positie hebben (Embretson & Reise, 2000; Meijer & Baneke, 2004).

IRT is een modelgebaseerde meting waarbij het trek niveau geschat wordt op basis van zowel de responsen van de persoon als de eigenschappen van de test-items. Een IRT model beschrijft hoe het trek niveau en de eigenschappen van de test-items gerelateerd zijn aan de item responsen van een persoon. Er zijn IRT modellen voor zowel dichotome als polytome data (Embretson & Reise, 2000).

IRT modellen zijn gebaseerd op de volgende assumpties; 1. unidimensionaliteit, 2. lokale onafhankelijkheid en 3. item respons functies (IRF) hebben een gespecificeerde vorm. Deze drie assumpties zullen hieronder toegelicht worden.

De meeste IRT modellen nemen een unidimensionele  $\theta$  aan. Dit betekent dat alle items in de test dezelfde latente trek meten, waardoor personen langs een lineaire schaal geordend kunnen worden. De aanname van het begrip ‘lokale onafhankelijkheid’ hangt hiermee samen; lokale onafhankelijkheid houdt in dat de responsen in een test statistisch onafhankelijk zijn conditioneel op  $\theta$ . Dat wil zeggen dat de respons op het ene item onafhankelijk is van de respons op het andere item. Lokale onafhankelijkheid is daarmee evidentie voor unidimensionaliteit.

Daarnaast is er de aanname dat IRF's een gespecificeerde vorm hebben. De vorm van IRF's beschrijft hoe veranderingen in trek niveau veranderingen te weeg brengen in de waarschijnlijkheid van een gespecificeerde respons. De IRF relateert de geobserveerde responsen aan de onderliggende  $\theta$ ; hoe hoger de  $\theta$  waarde, des te hoger is de waarschijnlijkheid dat een item dat  $\theta$  meet ‘goed’ beantwoord wordt (Embretson & Reise, 2000; Meijer & Baneke, 2004; Sijtsma & Molenaar, 2002). Een veel gebruikt model dat de vorm van de IRF specificeert is het ‘two-parameter logistic model (2PL-model)’. Dit model zal ter illustratie beschreven worden zodat meer duidelijkheid verkregen wordt over een IRF en zijn vorm.

Het 2PL-model gaat uit van twee parameters om itemeigenschappen weer te geven; namelijk het discriminerend vermogen van het item  $\alpha_i$  en de moeilijkheid van het item  $\beta_i$ . De grootte van  $\alpha_i$  geeft de mate weer waarin het item gerelateerd is aan de onderliggende latente trek. Een hoge waarde van  $\alpha$  betekent dat de antwoordcategorieën nauwkeurig discrimineren tussen trek niveaus. De waarde van  $\beta_i$  geeft het trek niveau weer dat noodzakelijk is om boven die categorie te antwoorden met 50% kans. Dus respondenten met een trek niveau hoger dan de waarde van  $\beta_i$  hebben meer dan 50% kans om in die categorie of een hogere categorie te antwoorden (Embretson & Reise, 2000; Emons, Meijer, & Denollet, 2005). Deze twee parameters worden dan ook beide meegenomen in de vergelijking van de waarschijnlijkheid van een respons:

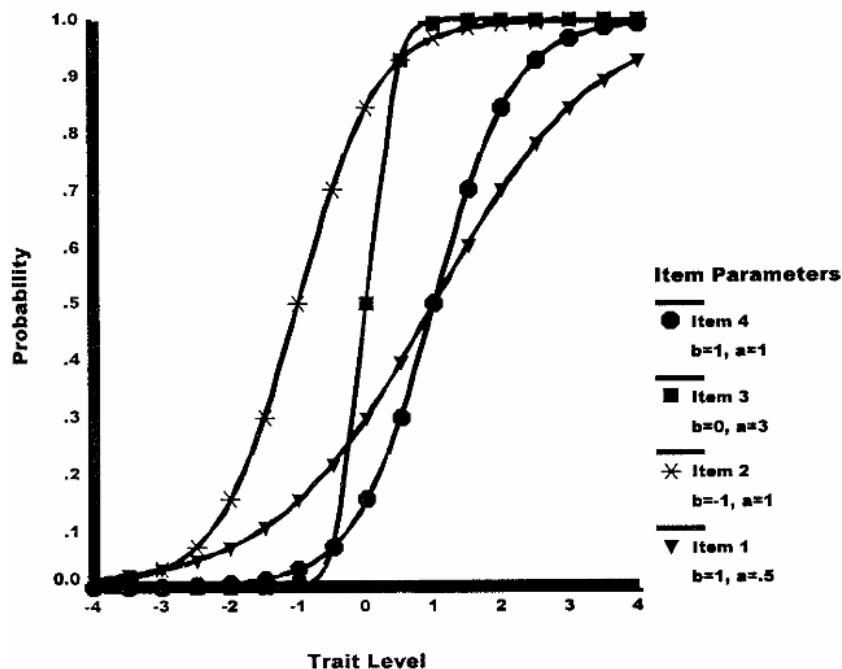
$$P(X_{is} = 1 | \theta_s, \beta_i, \alpha_i) = \frac{\exp[\alpha_i(\theta_s - \beta_i)]}{1 + \exp[\alpha_i(\theta_s - \beta_i)]}$$

Deze vergelijking laat zien dat de invloed van het verschil tussen trek niveau ( $\theta_s$ ) en de moeilijkheid van het item ( $\beta_i$ ) afhangt van het discriminerend vermogen van het item ( $\alpha_i$ ). Dit

houdt verder in dat het verschil tussen trekniveau en item moeilijkheid een grotere invloed heeft op de waarschijnlijkheden van items met een hoog discriminerend vermogen.

Het discriminerend vermogen van het item  $\alpha_i$  en de moeilijkheid van het item  $\beta_i$  bepalen de logistische vorm van de IRF. In Figuur 2 zijn de IRFs weergegeven voor vier verschillende items met respectievelijk  $\alpha_i$  is .5, 1, 3 en 1 én  $\beta_i$  1, -1, 0 en 1.

Figuur 2. IRFs voor vier items met verschillende  $\alpha_i$  en  $\beta_i$



De waarde van  $\alpha_1$  is relatief laag en van  $\alpha_3$  relatief hoog, dit verschil is te zien in de steilheid van de curve; de curve van item 1 is veel vlakker dan die van item 3. Bij item 1 hebben veranderingen in het trekniveau veel minder invloed op de waarschijnlijkheden en bij item 3 is dat juist meer invloed. Item 3 biedt daardoor meer informatie over de latente trek omdat het trekniveau meer betrekking heeft op veranderingen in het wel of niet goed beantwoorden van het item.

Ondanks het feit dat item 2 en 4 hetzelfde discriminerend vermogen ( $\alpha = 1$ ) hebben en daardoor dezelfde helling, kruisen de twee curves elkaar nooit. Dit heeft te maken met de verschillende  $\beta$ -waarden van de items. De waarde van  $\beta$  komt overeen met het trekniveau bij een waarschijnlijkheid van .5. De  $\beta$ -waarde bepaalt dus waar de curve zich bevindt langs het latente trek continuüm. Omdat de  $\beta$ -waarde van item 2 lager ( $\beta_2 = -1$ ) is dan die van item 4 ( $\beta_4 = 1$ ) ligt de curve van item 2 verder naar links ten opzichte van de curve van item 4 (Embretson & Reise, 2000).

### Parametrische en nonparametrische IRT

Er zijn twee soorten IRT modellen: parametrische en nonparametrische modellen. Beide modellen gaan uit van lokale onafhankelijkheid en unidimensionaliteit. Het verschil zit in de vorm van de IRF's. Bij parametrische modellen, de naam zegt het al, worden de IRF's volledig bepaald door parameters, zoals bijvoorbeeld het discriminerend vermogen van het item,  $\alpha$ , of de moeilijkheid van het item,  $\delta$  bij het 2PL-model. De IRF's hebben altijd de vorm van een logistische curve. Bij nonparametrische IRT zijn er geen parameters die de vorm van



de IRF beschrijven, de enige beperking die opgelegd wordt aan de IRF's is dat ze stijgend moeten zijn.

Dit maakt dat nonparametrische IRT modellen soms beter passen bij empirische data sets doordat er minder beperkingen zijn opgelegd aan de vorm van de IRF's (Sijtsma & Molenaar, 2002). Bovendien worden de data niet gedwongen een structuur aan te nemen die ze soms niet hebben en daarnaast zijn de software programma's voor nonparametrische IRT modellen vaak gebruikersvriendelijker en zijn de resultaten gemakkelijker te interpreteren (Meijer & Baneke, 2004).

Toch heeft het gebruik van nonparametrische IRT ook nadelen ten opzichte van parametrische IRT, zo kunnen alleen personen geordend worden langs het latente trek continuüm en kunnen items alleen geordend worden naar hun moeilijkheid bij de strengere modellen. Bovendien zijn geavanceerde toepassingen zoals gecomputeriseerd testen alleen mogelijk met parametrische IRT modellen (Embretson & Reise, 2000, p. 263-271).

In deze studie zijn ter illustratie beide vormen van IRT modellen toegepast op de data voor de kwaliteitsbepaling van de CBSK. Daarom zal het gebruikte parametrische IRT model het Graded Response Model (GRM; Samejima, 1969;1996) eerst verder toegelicht worden, alvorens het nonparametrische Mokken's Model van Monotone Homogeniteit (MMH; Mokken, 1971) beschreven zal worden.

#### *Parametrisch IRT model: het Graded-Response Model (GRM; Samejima, 1969;1996)*

Het GRM is een veel gebruikt parametrisch IRT model. Er is in deze studie gekozen voor het GRM, omdat dit model geschikt is bij tests waar gebruik gemaakt wordt van geordende antwoordcategorieën, zoals bijvoorbeeld Likertschalen (Embretson & Reise, 2000), zoals het geval is bij de CBSK.

GRM is een uitbreiding van het eerder beschreven 2PL-model. Het GRM kan namelijk gebruikt worden bij polytome items in plaats van dichotome items waarbij het 2PL-model gebruikt wordt. Het GRM werkt ook met twee parameters om iteigenschaften weer te geven. Bij het GRM worden de items gedefinieerd door een parameter die de helling aangeeft ( $\alpha$ ) en door twee of meer locatieparameters ( $\beta$ ). (Het aantal locatieparameters is gelijk aan het aantal antwoordcategorieën min één.) De grootte van de hellingparameter geeft de mate weer waarin het item gerelateerd is aan de onderliggende latente trek. Een hoge waarde van  $\alpha$  betekent dat de antwoordcategorieën nauwkeurig discrimineren tussen trekniveaus. De locatieparameters geven de tussenruimtes aan van de geordende antwoordcategorieën langs de  $\theta$ -schaal. De waarde van de locatieparameters geeft het trekniveau weer dat noodzakelijk is om boven die categorie te antwoorden met 50% kans. Dus respondenten met een trekniveau hoger dan de locatieparameter hebben meer dan 50% kans om in die categorie of een hogere categorie te antwoorden (Embretson & Reise, 2000; Emons et al., 2005). Deze twee itemparameterschattingen kunnen berekend worden met het programma MULTILOG (Thissen, 1991).

Ook hier worden deze twee parameters meegenomen in de vergelijking van de waarschijnlijkheid om in een bepaalde categorie te antwoorden. Bij het GRM zijn er echter twee stappen nodig om die waarschijnlijkheden te kunnen berekenen. De eerste stap is de berekening van de IRFs voor elk item. Elke curve geeft de waarschijnlijkheid weer dat het gegeven antwoord in of boven de gegeven 'categorietussenruimte' valt voorwaardelijk op  $\theta$ . De vergelijking ziet er dan als volgt uit:

$$P_{ix}^*(\theta_s) = \frac{\exp[\alpha_i(\theta_s - \beta_{ij})]}{1 + \exp[\alpha_i(\theta_s - \beta_{ij})]} \quad \text{waarbij geldt } x = j = \text{'categorietussenruimte'}$$

Als deze waarschijnlijkheden berekend zijn, kan de tweede stap uitgevoerd worden, namelijk het berekenen van de werkelijke waarschijnlijkheden om in een bepaalde categorie te antwoorden. Dit kan als volgt:

$$P_{ix}(\theta_s) = P_{ix}^*(\theta_s) + P_{i(x+1)}^*(\theta_s)$$

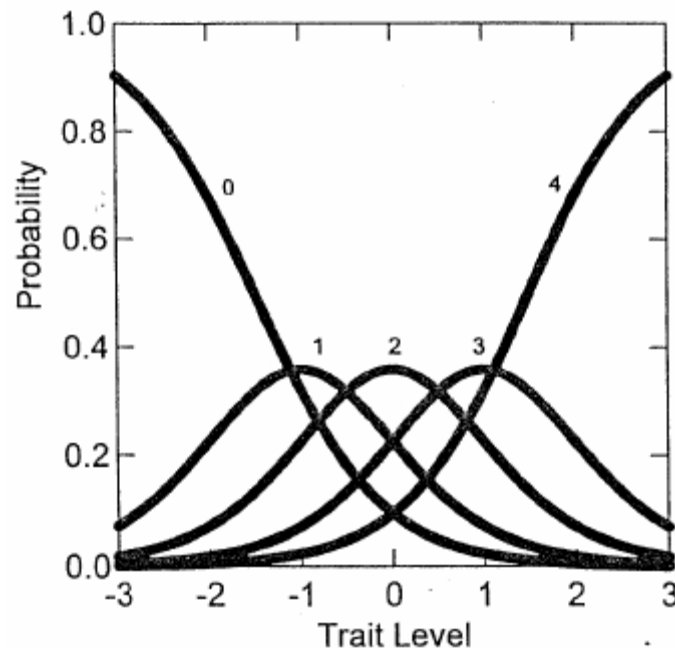
De waarschijnlijkheid om in één van de bijvoorbeeld vijf antwoordcategorieën te antwoorden, kan dan als volgt berekend worden:

$$\begin{aligned} P_{i0}(\theta_s) &= P_{i0}^*(\theta_s) - P_{i1}^*(\theta_s) \\ P_{i1}(\theta_s) &= P_{i1}^*(\theta_s) - P_{i2}^*(\theta_s) \\ P_{i2}(\theta_s) &= P_{i2}^*(\theta_s) - P_{i3}^*(\theta_s) \\ P_{i3}(\theta_s) &= P_{i3}^*(\theta_s) - P_{i4}^*(\theta_s) \\ P_{i4}(\theta_s) &= P_{i4}^*(\theta_s) - P_{i5}^*(\theta_s) \end{aligned}$$

Hierbij geldt dat  $P_{i0}^*(\theta_s) = 1.0$  omdat de waarschijnlijkheid om in of boven de laagste antwoordcategorie te antwoorden gelijk is aan 1.0 en dat  $P_{i5}^*(\theta_s) = .0$  omdat de waarschijnlijkheid om boven de hoogste antwoordcategorie te antwoorden .0 is.

Met deze waarschijnlijkheden kunnen zogeheten ‘categorie respons curves’ berekend worden, ze worden ook wel ‘item characteristic curves’ (ICC) genoemd. Deze curves zien er als volgt uit:

*Figuur 3.* Voorbeeld van een Item Characteristic Curve



Deze curves geven de waarschijnlijkheid weer om in een bepaalde antwoordcategorie te antwoorden gegeven het trekniveau. In het algemeen geldt bij deze curves dat hoe hoger de hellingparameters ( $\alpha$ ) des te smaller en ‘gepiekter’ de curves zijn, wat aangeeft dat de antwoordcategorieën goed onderscheid maken tussen de trekniveaus. De locatieparameters ( $\beta$ ) zeggen iets over de breedte van de curves en hoe ver de categorietussenruimten van elkaar af liggen (Embretson & Reise, 2000).

*Item-fit* Met het eerder genoemde programma MULTILOG kan ook de mate van ‘item-fit’ bepaald worden door middel van  $\chi^2$ . ‘Item-fit’ geeft de mate weer hoe goed het gekozen IRT model de responsen op een bepaald item verklaart of voorspelt. De  $\chi^2$  wordt als volgt berekend:

$$\chi^2 = \sum_{g=1}^G ((\text{obs}-\text{exp})^2/\text{exp})$$

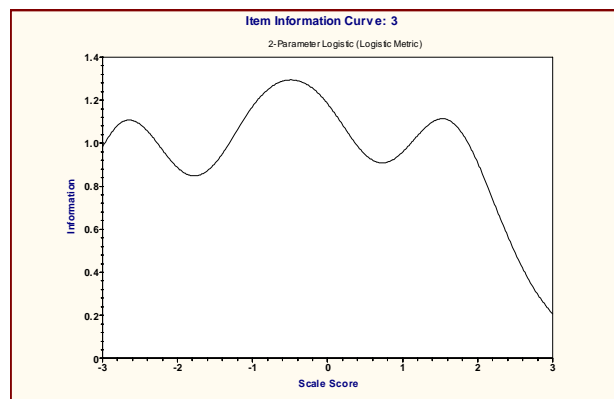
met  $\text{exp} = \text{exp. prop} * N$   
en waarbij geldt:

obs	= observed frequency
exp	= expected frequency
exp. prop	= expected proportion
g	= aantal antwoordcategorieën
N	= aantal deelnemers

De kritieke waarde van  $\chi^2$  wordt bepaald door het aantal vrijheidsgraden en de staartkans. De staartkans wordt gekozen; meestal wordt gekozen voor een staartkans van .05. Het aantal vrijheidsgraden is het aantal antwoordcategorieën min één. Als de  $\chi^2$ -waarde van een item kleiner is dan de kritieke waarde dan ‘fit’ het item, is de  $\chi^2$ -waarde van een item echter groter dan ‘fit’ het item niet met de data binnen het gekozen model. Dat houdt dus in dat het gekozen model de responsen op dat item niet goed verklaart of voorspelt (Embretson & Reise, 2000).

*Item Informatie Curves* Daarnaast geeft het programma MULTILOG zogeheten “Item Informatie Curves” (IICs). Een IIC geeft de hoeveelheid psychometrische informatie van een item weer op alle punten langs het latente trek continuüm (Embretson & Reise, 2000). In Figuur 4 is een voorbeeld van een IIC te zien.

*Figuur 4.* Voorbeeld van een Item Informatie Curve (IIC)



Deze curves zijn bij IRT een maat voor de meetnauwkeurigheid, waar KTT slechts één maat gebruikt, betrouwbaarheidscoëfficiënt  $\alpha$ , om de meetnauwkeurigheid van de gehele test weer te geven. Bovendien laten deze curves zien hoe het item zich verhoudt ten opzichte van de latente trek; doordat de KTT slechts één maat gebruikt, kan het geen onderscheid maken tussen verschillende  $\theta$  niveaus. IRT biedt met deze curves dus meer informatie over het item dan KTT kan geven.

*Nonparametrisch IRT model: Mokken's model van Monotone Homogeniteit (MMH; Mokken, 1971)*

Een nonparametrisch IRT model is het Mokken's model van Monotone Homogeniteit (MMH). Het MMH is gekozen voor deze studie omdat het een veel gebruikt nonparametrisch model is. Dit model gaat uit van de eerder genoemde aannames van nonparametrische IRT; 1. unidimensionaliteit, 2. lokale onafhankelijkheid en 3. monotone stijgende IRFs.

Het MMH zou gezien kunnen worden als een stochastische/probabilistische versie van het deterministische Guttman (1950) model. Het Guttman model neemt aan dat elk item en elke persoon een locatie heeft op het  $\theta$  continuüm én dat een 'goed' antwoord gegeven wordt wanneer een persoon bekwaam is dan het item moeilijk is.

Stel een test bestaat uit 6 opgaven die oplopen in moeilijkheid, het Guttman model stelt dan dat het minder slimme kind alleen het volgende antwoordpatroon kan hebben 100000 (1 = 'goed', 0 = 'fout') en dat het antwoordpatroon van een slimmer kind er bijvoorbeeld alleen zo uit kan zien, 111000. Op deze manier bepaalt het aantal 'goede' antwoorden (= totaalscore) welke items 'goed' zijn. Dit is echter niet realistisch.

In andere tests waar de items bijvoorbeeld minder in moeilijkheid verschillen, kan het best zo zijn dat een kind een makkelijker item 'fout' heeft en een moeilijker item 'goed'. Volgens het Guttman model is dit niet mogelijk, het model spreekt dan ook van Guttman fouten als dat het geval is.

Het MMH gaat er echter vanuit dat dit wel kan gebeuren en dat de kans op een 1 of een 0 niet gelijk is aan één. Het hoeft dus niet zo te zijn dat twee personen met dezelfde totaalscore ook hetzelfde antwoordpatroon hebben.

De zogeheten Guttman fouten komen echter wel terug in het MMH, namelijk bij het onderzoeken van de monotonie.

*Het onderzoeken van monotonie* Mokken (1971, 1997) stelde voor om bij het onderzoeken van de monotonie gebruik te maken van de schalingscoëfficiënt  $H_{gh}$  voor itemparen (g, h), de schalingscoëfficiënt  $H_g$  voor een item met betrekking op andere items in de test en de schalingscoëfficiënt  $H$  voor totale itemset in de test.

De  $H_{gh}$  coëfficiënt geeft de relatie weer tussen twee items; hoe hoger de coëfficiënt des te beter passen de items bij elkaar.

De  $H_g$  coëfficiënten worden bij nonparametrische IRT gebruikt om de item fit te bepalen (Sijtsma & Molenaar, 2002). Daarnaast kunnen de  $H_g$  coëfficiënten gezien worden als de  $\alpha$  parameters, zoals die gebruikt wordt bij parametrische IRT modellen als het 2PL model; ze duiden het discriminerend vermogen van de items aan (Meijer & Baneke, 2004).

De  $H$  coëfficiënt duidt de sterkte cq zwakte van de gehele test aan. Mokken beveelt de volgende indeling aan voor het gebruik van  $H$  coëfficiënten bij praktische testconstructiedoelen:  $.3 \leq H < .4$  is een zwakke schaal,  $.4 \leq H < .5$  is een gemiddelde schaal en  $H \geq .5$  is een sterke schaal (Meijer & Baneke, 2004; Sijtsma & Molenaar, 2002).

Deze coëfficiënten kunnen berekend worden met behulp van gemaakte Guttman fouten, het gaat echter te ver om deze berekening hier uit te leggen. Met het programma MSP5 (Molenaar, Schuur, Sijtsma & Molenaar, 2002) kunnen deze coëfficiënten namelijk gemakkelijk berekend worden.

*Het onderzoeken van unidimensionaliteit* Om de unidimensionaliteit van een schaal te onderzoeken zijn er verschillende nonparametrische procedures voorgesteld. Voor deze studie is gekozen voor de relatief simpele procedure in het programma MSP5. MSP5 bevat een geautomatiseerde itemselectie procedure, die gebaseerd is op  $H_g$  coëfficiënten voor een item met betrekking tot andere items in de test.

De geautomatiseerde itemselectie procedure in MSP5 is een zogeheten “bottom-up” procedure; waarbij het begint met een itempaar en er steeds meer items geselecteerd worden die samen een cluster vormen. Voor het eerste itempaar waar mee begonnen wordt geldt dat:

1.  $H_{gh}$  significant groter is dan 0 en
2.  $H_{gh}$  de grootste coëfficiënt is van alle mogelijke itemparen.

Als het eerste itempaar geselecteerd is, kan er een derde item  $j$  geselecteerd worden om de schaal uit te breiden. Voor dat item geldt dat:

1. het hoog correleert met de al geselecteerde items,
2. het een  $H_j$  coëfficiënt heeft die groter is dan 0 en
3. het een  $H_j$  coëfficiënt heeft die groter is dan een waarde  $C$  die door de gebruiker is ingesteld (een hogere  $C$  ondergrens voor  $H_j$  zorgt ervoor dat items met een hoger discriminerend vermogen geselecteerd worden en items met een lager discriminerend vermogen uitgesloten worden.  $C = .3$  wordt aanbevolen als minimum ondergrens).

Het programma gaat door met het selecteren van items zolang er items beschikbaar zijn die aan de gestelde voorwaarden voldoen. Op deze manier kunnen items geïdentificeerd worden die samen een cluster vormen om dezelfde latente trek te meten.

## Onderzoeksvragen

In deze studie zal getracht worden de volgende twee hoofdvragen te beantwoorden:

- Wat kan er gezegd worden over de kwaliteit van de CBSK, gebruikmakend van IRT?
- Hoe hangt de persoonlijkheid van een kind tussen 8 en 12 jaar samen met zijn/haar zelfconcept, zoals gemeten door respectievelijk de NPV-J en de CBSK?

## Methode

### *Respondenten*

Er zijn gegevens verzameld bij 108 basisschoolleerlingen tussen de 7 en 13 jaar. De groep bestaat uit 52 jongens (gemiddelde leeftijd = 10,1 jaar) en 56 meisjes (gemiddelde leeftijd = 10,0 jaar). Deze kinderen zitten op basisschool de Borgh in Hengelo (O).

### *Metingen*

*Junior Nederlandse Persoonlijkheidsvragenlijst (NPV-J)* De NPV-J is een persoonlijkheidsvragenlijst voor kinderen van 9 tot en met 15 jaar. De vragenlijst meet de volgende vijf subschalen (voor de inhoud van de subschalen zie Bijlage 1):

1. Inadequatie (IN) (items 1, 4, 6, 8, 13, 14, 19, 28, 32, 34, 36, 38, 48, 50, 52, 54, 57, 59, 66, 70, 72, 75, 91, 93, 96, 98, 100, 102)
2. Volharding (VO) (items 2, 10, 12, 30, 31, 33, 39, 41, 43, 45, 53, 63, 68, 69, 71, 73, 77, 78, 84, 88, 94, 95, 101, 103, 104)
3. Sociale Inadequatie (SI) (items 21, 22, 23, 25, 26, 44, 51, 62, 79, 80, 85, 89, 105)
4. Recalcitrantie (RE) (items 5, 11, 15, 16, 18, 20, 24, 29, 35, 37, 40, 42, 46, 47, 49, 55, 61, 65, 74, 82, 83, 86, 87, 92)
5. Dominantie (DO) (items 3, 7, 9, 17, 27, 56, 58, 60, 64, 67, 76, 81, 90, 97, 99)

De vragenlijst bestaat uit 105 stellingen, waar het kind kan kiezen uit de antwoordcategorieën ‘ja’ en ‘nee’, en ‘?’ in het geval ze echt niet kunnen kiezen (Luteijn, et al., 1989).

*Competentiebelevingsschaal voor Kinderen* De Competentiebelevingsschaal voor Kinderen (CBSK) is een zelfrapportage vragenlijst die gebruikt wordt om vast te stellen hoe kinderen van 8 tot 12 jaar hun eigen functioneren op een aantal specifieke gebieden ervaren en hoe zij hun globale gevoel van eigenwaarde beoordelen. De CBSK kent de volgende zes subschalen (voor de inhoud van de subschalen zie Bijlage 2):

1. Schoolvaardigheden (SV) (items 1, 7, 13, 19, 25, 31)
2. Sociale acceptatie (SA) (items 2, 8, 14, 20, 26, 32)
3. Sportieve vaardigheden (SP) (items 3, 9, 15, 21, 27, 33)
4. Fysieke verschijning (FV) (items 4, 10, 16, 22, 28, 34)
5. Gedragshouding (GH) (items 5, 11, 17, 23, 29, 35)
6. Gevoel van Eigenwaarde (GE) (items 6, 12, 18, 24, 30, 36)

Elke subschaal wordt gemeten door 6 items, in totaal bestaat de vragenlijst uit 36 items. Bij het beantwoorden van een item moet het kind eerst een keuze maken tussen twee stellingen en kiezen voor die stelling die voor hem/haar van toepassing is. Nadat deze keuze gemaakt is, moet het kind aangeven of de gekozen stelling ‘helemaal waar voor mij’ of ‘een beetje waar voor mij’ is (Veerman et al., 1997, 2004).

### *Analyses*

Data-analyse werd gedaan in twee fases. De eerste fase werd gebruikt om de kwaliteit van de CBSK te onderzoeken. De CBSK is gebruikt ter illustratie zodat duidelijk gemaakt kan worden hoe IRT bij dergelijke subschalen, zoals die van de CBSK, werkt. Hiervoor is allereerst een analyse uitgevoerd volgens de Klassieke Test Theorie middels het programma SPSS om de standaard deviatie en gemiddelde per subschaal te bepalen en om de item-test correlaties en coëfficiënt  $\alpha$  te berekenen. Verder werd een analyse uitgevoerd volgens het parametrische IRT model, het GRM, met het programma MULTILOG om de itemparameterschattingen te laten berekenen, de item-fit te bepalen en om de item-informatiecurves weer te geven. Als laatste werd in deze fase een analyse uitgevoerd volgens het nonparametrische IRT model, het MMH, met het programma MSP5 om de monotonie en de dimensionaliteit te onderzoeken.

De tweede fase werd gebruikt om de samenhang tussen persoonlijkheid en zelfconcept te onderzoeken, hiertoe werd een correlatieanalyse uitgevoerd tussen de scores op de vijf subschalen van de NPV-J en de scores op de zes subschalen van de CBSK.

## Resultaten

### Fase 1: Kwaliteitsbepaling CBSK

#### *KTT*

*Standaard deviaties en gemiddelden* In Tabel 1 op de volgende pagina zijn de gemiddelde scores en de standaarddeviaties weergegeven per subschaal van de CBSK en uitgesplitst naar geslacht.

Tabel 1. *Gemiddelden en standaarddeviaties per geslacht*

	Jongens (N = 52)		Meisjes (N = 56)	
	Mean	Std. dev.	Mean	Std. dev.
SV	17.73	4.04	17.36	3.63
SA	18.98	3.28	19.34	3.52
SP	19.33	3.15	17.93	3.69
FV	20,44	3.42	19.73	3.54
GH	17.65	3.46	18.71	3.00
GE	20,81	3.12	20,45	3.49

*Item-test correlatie en coëfficiënt  $\alpha$*  In Tabel 2 staan de item-test correlaties en coëfficiënt  $\alpha$  voor de zes subschalen van de CBSK weergegeven.

Tabel 2. *Item-test correlaties en coëfficiënt  $\alpha$  voor subschalen CBSK*

	SV	SA	SP	FV	GH	GE
<b>Item 1</b>	.4547	.4339	.4602	.4251	.2369	.5866
<b>Item 2</b>	.3932	.5567	.3413	.3700	.5406	.5661
<b>Item 3</b>	.4525	.4551	.3505	.5514	.2747	.4253
<b>Item 4</b>	.5446	.2963	.4402	.4785	.4049	.5877
<b>Item 5</b>	.5682	.4154	.3267	.3565	.6062	.4906
<b>Item 6</b>	.6712	.4425	.4069	.5659	.5292	.5040
<b><math>\alpha</math></b>	.7645	.7010	.6501	.7171	.6874	.7797

Alle subschalen van de CBSK hebben een  $\alpha$  tussen .6501 en .7797, ondanks het feit dat een aantal afzonderlijke items een redelijk lage item-testcorrelatie heeft. Item 3 ('Sommige kinderen doen meestal wat van hen wordt verwacht' vs 'Andere kinderen doen vaak niet wat van hen wordt verwacht') bijvoorbeeld van de subschaal Gedragshouding (GH) heeft een item-test correlatie van .2747 en item 4 ('Sommige kinderen doen altijd allerlei dingen met een heleboel andere kinderen' vs 'Andere kinderen doen de dingen meestal in hun eentje) van de subschaal Sociale acceptatie een item-test correlatie van .2963.

### *Parametrische IRT*

In deze studie dient de parametrische IRT analyse vooral ter illustratie, daarom is er voor gekozen niet alle subschalen van de CBSK te analyseren, maar alleen de subschalen Schoolvaardigheden, Sportieve Vaardigheden en Gevoel van Eigenwaarde. Er is gekozen voor de subschaal Gevoel van Eigenwaarde, omdat deze subschaal de hoogste betrouwbaarheidscoëfficiënt  $\alpha$  heeft en het enige globale aspect meet in deze vragenlijst. Dit in tegenstelling tot de subschalen Schoolvaardigheden en Sportieve vaardigheden, die specifieke competenties meten. Daarnaast is gekozen voor deze twee subschalen omdat Sportieve vaardigheden de laagste coëfficiënt  $\alpha$  heeft en de subschaal Schoolvaardigheden de hoogste coëfficiënt  $\alpha$  van de subschalen die specifieke competenties meten.

Vooraf dient vermeld te worden dat de resultaten van deze parametrische analyse met enige voorzichtigheid behandeld moeten worden. Dit omdat het gebruikte programma MULTILOG bij weinig proefpersonen vaak instabiele resultaten geeft, het is echter goed om eens te kijken hoe parametrische IRT in zijn werk gaat.

*Item-fit* Door middel van item-fit kan worden nagegaan hoe goed het IRT model de responsen op een bepaald item verklaart of voorspelt (Embretson & Reise, 2000). Hierbij is in deze studie gebruik gemaakt van de  $\chi^2$  als maat voor item fit.

De kritieke waarde van  $\chi^2$  wordt bepaald door het aantal vrijheidsgraden en de staartkans. Bij deze studie is gekozen voor een staartkans die het meest gebruikt wordt, namelijk .05. Het aantal vrijheidsgraden is het aantal antwoordcategorieën min één, dus  $4-1=3$  vrijheidsgraden. Dit aantal vrijheidsgraden geldt voor alle items, aangezien alle items 4 antwoordmogelijkheden hebben. Hierdoor wordt de kritieke waarde:  $\chi^2 = 9.49$  (Moore & McCabe, 2001).

Als de  $\chi^2$  waarde van een item kleiner is dan deze kritieke waarde dan ‘fit’ het item, is de  $\chi^2$  waarde echter groter dan ‘fit’ het item niet met de data binnen het gekozen model. Het model kan dan de responsen op dat item niet goed verklaren of voorspellen (Embretson & Reise, 2000).

In Tabel 3 zijn de  $\chi^2$  waarden van de items van de subschalen Schoolvaardigheden, Sportieve vaardigheden en Gevoel van eigenwaarde overzichtelijk weergegeven. Daarin is te

Tabel 3.  $\chi^2$  van de items van de subschalen Schoolvaardigheden, Sportieve vaardigheden en Gevoel van eigenwaarde

	School- vaardigheden	Sportieve vaardigheden	Gevoel van eigenwaarde
	$\chi^2$	$\chi^2$	$\chi^2$
Item 1	.05645	.05141	.01446
Item 2	.05654	.21847	.00660
Item 3	.16433	.01463	.04940
Item 4	.26416	.00612	.03253
Item 5	.22142	.04546	.02836
Item 6	.10248	.11216	.02385

zien dat de  $\chi^2$  waarden van alle items kleiner zijn dan de kritieke waarde van 9.49. Dit betekent dat het gekozen model past bij de data.

*Itemparameterschattingen* In tabel 4 op de volgende pagina staan de itemparameterschattingen overzichtelijk weergegeven voor de subschalen Schoolvaardigheden, Sportieve vaardigheden en Gevoel van Eigenwaarde uit de CBSK. De  $\alpha$ -waarden kunnen zowel als getal geïnterpreteerd worden als in de vorm (van de steilheid) van een “Item-Informatiecurve” (IIC). De IICs worden in hetvolgende onderdeel *Item-Informatiecurves* besproken, daar zal kort aandacht geschonken worden aan weergave van de  $\alpha$ -waarden in de curves. Hier bij het onderdeel *Itemparameterschattingen* zal gekeken worden naar de  $\alpha$ -waarden als getallen.

De  $\beta$  waarden kunnen door MULTILOG weergegeven worden in de vorm van “Item Characteristic Curves” (ICCs), dat is overzichtelijker dan een tabel met getallen. Alle ICCs voor de subschalen Schoolvaardigheden, Sportieve vaardigheden en Gevoel van Eigenwaarde zijn opgenomen in respectievelijk de bijlagen 3, 4 en 5. Een goede en slechte curve zullen kort besproken worden om een indruk te geven hoe de waarden geïnterpreteerd dienen te worden.



Tabel 4. *Itemparameterschattingen voor de items van de subschalen Schoolvaardigheden, Sportieve vaardigheden en Gevoel van eigenwaarde*

	SCHOOLVAARDIGHEDEN				SPORTIEVE VAARDIGHEDEN				GEVOEL VAN EIGENWAARDE			
	$\alpha$ (S.E.)	$\beta_1$ (S.E.)	$\beta_2$ (S.E.)	$\beta_3$ (S.E.)	$\alpha$ (S.E.)	$\beta_1$ (S.E.)	$\beta_2$ (S.E.)	$\beta_3$ (S.E.)	$\alpha$ (S.E.)	$\beta_1$ (S.E.)	$\beta_2$ (S.E.)	$\beta_3$ (S.E.)
Item 1	1.12 (0.28)	-2.28 (0.56)	-0.34 (0.26)	1.41 (0.43)	2.12 (0.39)	-2.32 (0.49)	-1.22 (0.22)	-0.30 (0.16)	2.51 (0.45)	-2.04 (0.41)	-1.17 (0.21)	-0.03 (0.15)
Item 2	1.00 (0.26)	-2.31 (0.67)	-0.35 (0.29)	0.99 (0.41)	0.99 (0.28)	-1.64 (0.56)	-0.85 (0.36)	0.05 (0.31)	2.06 (0.50)	-2.23 (0.47)	-1.54 (0.31)	-0.64 (0.18)
Item 3	1.50 (0.31)	-1.66 (0.37)	-0.94 (0.25)	0.52 (0.26)	1.11 (0.27)	-2.81 (0.76)	-1.67 (0.44)	0.34 (0.28)	1.46 (0.37)	-3.79 (1.24)	-2.01 (0.45)	-0.79 (0.23)
Item 4	2.10 (0.41)	-2.15 (0.40)	-0.74 (0.18)	0.49 (0.19)	1.67 (0.33)	-2.31 (0.49)	-0.58 (0.20)	1.19 (0.26)	2.38 (0.54)	-2.10 (0.42)	-1.53 (0.27)	-0.65 (0.17)
Item 5	2.22 (0.41)	-2.40 (0.45)	-1.26 (0.23)	0.14 (0.16)	1.02 (0.32)	-2.63 (0.89)	-1.64 (0.51)	-0.53 (0.30)	1.97 (0.44)	-2.56 (0.64)	-1.61 (0.32)	-0.44 (0.18)
Item 6	2.82 (0.48)	-1.47 (0.22)	-0.49 (0.14)	0.54 (0.15)	1.65 (0.36)	-1.86 (0.40)	-1.02 (0.24)	0.40 (0.22)	1.40 (0.33)	-2.70 (0.65)	-1.43 (0.34)	0.42 (0.22)

Opvallend aan de  $\alpha$  waarden van de subschaal Schoolvaardigheden in Tabel 4 is dat de  $\alpha$  waarden van de laatste drie items een stuk hoger zijn dan die van de eerste drie items. Dit betekent dat de laatste drie items in hogere mate gerelateerd aan schoolvaardigheden zijn, met name item 6 ‘sommige kinderen vinden het moeilijk op school het antwoord op een vraag te bedenken’ vs ‘andere kinderen weten bijna altijd wel een antwoord’ is sterk gerelateerd aan de subschaal. Item 1 ‘sommige kinderen vinden, dat ze erg goed zijn in hun schoolwerk’ vs ‘andere kinderen maken zich er soms zorgen over of ze hun schoolwerk wel goed doen’ is in deze subschaal het minst gerelateerd aan ‘schoolvaardigheden’.

Opvallend aan de  $\beta$ -waarden van de items van de drie subschalen is dat bijna alle items de linkerkant van het latente trek continuüm (dat wil zeggen  $-3 < \theta < 0$ ) meten. Dit is te zien aan de  $\beta_3$  waarden die dicht bij nul liggen. Dit is ook duidelijk te zien aan de ICCs in de bijlagen 3, 4 en 5; de lijnen zijn vooral in het linkerdeel van de afbeelding te zien.

De ICC van item 4 ‘sommige kinderen vinden dat ze beter zijn in sport en gymnastiek dan andere kinderen’ vs ‘andere kinderen vinden dat ze minder goed zijn in sport en gymnastiek’ van de subschaal Sportieve vaardigheden is een redelijk goed voorbeeld van een goede ICC; namelijk weinig overlap tussen de verschillende lijnen en hoge en minder brede pieken.

De ICC van item 3 ‘sommige kinderen zijn tevreden met zichzelf als persoon’ vs ‘andere kinderen zijn vaak niet tevreden met zichzelf als persoon’ van de subschaal Gevoel van eigenwaarde is een hele slechte; lijn 2 (blauw) is hoger en ligt boven lijn 1 (zwart). Bovendien meet dit item erg sterk de linkerkant van het trek continuüm en kan het weinig zeggen over de rechterkant. Daarnaast kan er weinig onderscheid gemaakt worden tussen de verschillende antwoordcategorieën.

*Item-informatiecurves* In de Bijlagen 6, 7 en 8 zijn de IICs weergegeven voor respectievelijk de subschalen Schoolvaardigheden, Sportieve Vaardigheden en Gevoel van Eigenwaarde. Een IIC laat zien hoe het item zich verhoudt ten opzichte van de latente trek. Er zal een goede en een slechte IIC ter illustratie besproken worden. Ook zal er eerst nog kort aandacht geschonken worden aan de interpretatie van de  $\alpha$ -waarden uit tabel 4 in de vorm van de IIC.

Het beste is de interpretatie van de  $\alpha$ -waarden uit te leggen aan de hand van de  $\alpha$ -waarden van de items 2 ( $\alpha = 1.00$ ) en 6 ( $\alpha = 2.82$ ) uit de subschaal Schoolvaardigheden. De waarde van item 2 is relatief veel lager dan de waarde van item 6, dit is goed te zien aan de steilheid van de IICs van de items; de IIC van item 2 is veel vlakker dan de IIC van item 6 die veel steiler is. Hoe hoger de  $\alpha$ -waarde, des te meer is het item aan de latente trek gerelateerd en des te steiler is de helling van de IIC.

Een voorbeeld van een slechte IIC is de IIC van item 2 uit de subschaal Sportieve vaardigheden. Dit item biedt de minste informatie over de latente trek die er gemeten wordt, dit is af te leiden uit de vlakheid van de curve, bovendien heeft de curve geen enkel ‘reliëf’. De IIC van item 1 uit de subschaal Gevoel van Eigenwaarde daarentegen heeft meer reliëf en is ook veel hoger, waardoor item 1 meer informatie biedt over de latente trek. Daarmee is de IIC van item 1 een relatief goede curve. Er moet echter wel vermeld worden dat item 1 met name informatie biedt over de linkerkant van het latente trek continuüm (dat wil zeggen  $-3 < \theta < 0$ ) en daarmee niets zegt over de rechterkant.

De ‘linksscheefheid’ van veel items is ook goed terug te vinden in de Test Informatie Curves (TICs), die tevens door het programma MULTILOG worden gegeven. De TICs van de subschalen Schoolvaardigheden, Sportieve vaardigheden en Gevoel van Eigenwaarde zijn opgenomen in respectievelijk de bijlagen 9, 10 en 11. Een TIC geeft de hoeveelheid psychometrische informatie van de gehele test weer op alle punten langs het latente trek

continuüm en niet van één item, zoals een IIC . In de bijlagen 9, 10 en 11 is de TIC weergegeven met een blauwe ononderbroken lijn. Daarnaast is er ook een rode gestippelde lijn te zien, deze geeft de standaard fout aan. Naar aanleiding van de TICs van de drie verschillende subschalen kan gezegd worden dat de subschalen nauwkeurig meten aan de linkerkant van het latente trek continuüm (dit houdt een positief zelfbeeld in) en minder nauwkeurig aan de rechterkant (dit houdt een positief zelfbeeld in), terwijl juist de mensen met een negatief zelfbeeld de mensen zijn die interessant zijn om verder te onderzoeken. De meetnauwkeurigheid is af te leiden aan de curve die de standaard fouten aangeeft, die is namelijk aan de linkerkant veel lager dan aan de rechterkant van het continuüm.

Nu de resultaten van de parametrische analyse behandeld zijn, zal hierna de nonparametrische IRT analyse aan bod komen. Nonparametrische IRT kan in tegenstelling tot parametrische IRT met relatief weinig proefpersonen worden uitgevoerd, bovendien worden vrij eenvoudige analyses gebruikt. De nonparametrische IRT analyse zal uitgevoerd worden op dezelfde drie subschalen die ook gebruikt zijn bij de parametrische IRT analyse.

### *Nonparametrische IRT*

*Monotonie onderzoeken* Om de monotonie te onderzoeken wordt er gekeken naar de  $H_g$  coëfficiënten van de items om iets te kunnen zeggen over het discriminerend vermogen van de items en naar de  $H$  coëfficiënten om iets te kunnen zeggen over de schaalbaarheid van de items. In Tabel 5 op de volgende pagina zijn deze coëfficiënten weergegeven voor respectievelijk de subschalen Schoolvaardigheden, Sportieve Vaardigheden en Gevoel van Eigenwaarde.

Item 6 van de subschaal Schoolvaardigheden (‘sommige kinderen vinden het moeilijk op school het antwoord op een vraag te bedenken’ vs ‘andere kinderen weten bijna altijd wel een antwoord’) heeft het hoogste discriminerend vermogen van de drie onderzochte subschalen en item 5 van de subschaal Sportieve vaardigheden (‘sommige kinderen staan bij sport en spel vaak te kijken in plaats van dat ze mee doen’ vs ‘andere kinderen spelen eerder mee dan dat ze alleen maar staan te kijken’) heeft het laagste discriminerend vermogen. Dat betekent dat item 6 beter onderscheid kan maken tussen de verschillende trek niveaus dan item 5.

Tabel 5.  $H_g$  coëfficiënten en  $H$  coëfficiënt voor de subschalen Schoolvaardigheden, Sportieve vaardigheden en Gevoel van Eigenwaarde

	School- vaardigheden	Sportieve vaardigheden	Gevoel van Eigenwaarde
	$H_g$	$H_g$	$H_g$
Item 1	.36	.32	.46
Item 2	.31	.24	.43
Item 3	.35	.25	.34
Item 4	.41	.33	.45
Item 5	.45	.23	.39
Item 6	.48	.29	.42
$H$ coëfficiënt	.39	.27	.42

De subschaal Gevoel van Eigenwaarde is de beste subschaal van de drie onderzochte subschalen, dat is te zien aan de hoogste  $H$  coëfficiënt. Volgens Mokken duidt dit op een gemiddelde schaal. De subschaal Sportieve vaardigheden is volgens Mokken een zwakke schaal, aangezien de  $H$  coëfficiënt rond de door Mokken aanbevolen ondergrens van .30 ligt.

*Unidimensionaliteit onderzoeken* De uitvoer van het programma MSP5 laat zien dat wanneer  $C = .10$  gebruikt wordt als ondergrens, alle items per subschaal opgenomen worden in die subschaal. Daarom is de ondergrens verhoogd naar  $C = .20$  en ook hier geldt dat alle items van de originele subschaal opgenomen worden in de door MSP5 gevormde schaal. Als reactie wordt de ondergrens opnieuw verhoogd, nu naar  $C = .30$ . Voor de subschalen Schoolvaardigheden en Gevoel van Eigenwaarde geldt dat alle items uit de originele schaal ook door MSP5 nu weer opgenomen worden in de gevormde schaal. Voor de subschaal Sportieve vaardigheden geldt echter dat nu de items 2, 3 en 5 weggelaten worden uit de door MSP5 gevormde schaal.

Met betrekking tot de unidimensionaliteit van de schalen zou voorzichtig geconcludeerd kunnen worden dat de subschalen Schoolvaardigheden en Gevoel van Eigenwaarde wel unidimensioneel zijn, maar dat er grote vraagtekens zijn bij Sportieve vaardigheden wat betreft de unidimensionaliteit.

## Fase 2 Correlatieanalyse subschalen NPV-J en CBSK

In Tabel 6 staat de correlatiematrix van de NPV-J met de CBSK. Daarin valt direct op dat de subschaal Inadequatie (IN) van de NPV-J negatief correleert met alle subschalen van de CBSK. Dit betekent dat als men veel gevoelens van inadequatie kent, men een laag algemeen zelfconcept zal hebben.

Daarnaast is te zien dat de subschaal Volharding (VO) van de NPV-J met alle subschalen van de CBSK positief correleert, behalve met de subschaal Sportieve vaardigheden. De correlatie tussen Sportieve vaardigheden en Volharding is namelijk niet significant verschillend van nul. Wat echter wel gevonden is, is dat wanneer kinderen beschikken over de persoonlijkheidstrekk 'volharding' zij een positiever beeld hebben van hun schoolvaardigheden, sociale acceptatie, fysieke verschijning, gedragshouding en gevoel van eigenwaarde.

In Tabel 6 is ook te zien dat de persoonlijkheidstrekk 'Sociale inadequatie' negatief correleert met de subschalen Sociale acceptatie, Fysieke verschijning en Gevoel van Eigenwaarde van de CBSK. Dit is enigszins logisch, want een kind dat sociale contacten liever vermijdt en/of zich niet gelukkig voelt in sociale contacten zal zich vaak ook niet sociaal geaccepteerd voelen. Hoe een kind denkt over zijn fysieke verschijning speelt daarbij ook een grote rol; als een kind zichzelf heel erg mooi vindt zal het niet snel sociale contacten willen vermijden. En dat kan weer een consequentie zijn voor het gevoel van eigenwaarde van het kind.

Tabel 6. *Correlatiematrix NPV-J en CBSK*

		Subschalen NPV-J					
		IN	VO	SI	RE	DO	
Subschalen CBSK	SV	Pearson Correlatie	<b>-.262</b>	<b>.251</b>	-.124	-.045	<b>.338</b>
		Sig. (2-zijdig)	.006	.009	.202	.646	0
		N	107	107	107	107	107
	SA	Pearson Correlatie	<b>-.227</b>	<b>.198</b>	<b>-.248</b>	-.154	.147
		Sig. (2-zijdig)	.019	.041	.01	.114	.13
		N	107	107	107	107	107
	SP	Pearson Correlatie	<b>-.216</b>	.086	-.155	-.073	.231
		Sig. (2-zijdig)	.026	.377	.11	.455	.017
		N	107	107	107	107	107
	FV	Pearson Correlatie	<b>-.445</b>	<b>.223</b>	<b>-.25</b>	<b>-.197</b>	.184
		Sig. (2-zijdig)	0	.021	.01	.042	.058
		N	107	107	107	107	107
	GH	Pearson Correlatie	<b>-.337</b>	<b>.407</b>	-.078	<b>-.224</b>	-.08
		Sig. (2-zijdig)	0	0	.423	.02	.415
		N	107	107	107	107	107
	GE	Pearson Correlatie	<b>-.514</b>	<b>.366</b>	<b>-.322</b>	-.176	.153
		Sig. (2-zijdig)	0	0	.001	.071	.116
		N	107	107	107	107	107

**VET** Correlatie is significant op het 0.01 niveau (2-zijdig).

**Schuin** Correlatie is significant op het 0.05 niveau (2-zijdig).

Verder is te zien dat de persoonlijkheidstrek 'Recalcitrantie' negatief correleert met de subschalen Fysieke verschijning en Gedragshouding. Dit betekent dat wanneer een kind 'hoog' scoort op Recalcitrantie het zijn/haar fysieke verschijning 'laag' inschat; een kind dat wantrouwender naar andere mensen toe is zal zichzelf minder mooi vinden. De correlatie tussen Recalcitrantie en Gedragshouding is vrij logisch; een kind dat minder wantrouwend en zich minder wil afzetten tegen anderen, zal zich ook beter (willen) gedragen.

Als laatste is er een vrij opvallende positieve correlatie gevonden tussen de persoonlijkheidstrek Dominantie en de subschaal Schoolvaardigheden van de CBSK; dit betekent dat kinderen die dominant zijn vinden dat ze het goed doen op school en andersom dat kinderen die juist niet dominant zijn vinden dat ze het niet goed doen op school.

## Discussie

In deze studie is gebruikt gemaakt van de CBSK om de kwaliteit daarvan te bepalen met behulp van IRT, dat wil zeggen met het parametrische GRM en het nonparametrische MMH. Verder is de NPV-J gebruikt om de samenhang tussen persoonlijkheid en zelfbeeld te onderzoeken.

Uit deze studie is gebleken dat het nonparametrische MMH beter geschikt is om toe te passen op de data verkregen met het afnemen van de CBSK dan het parametrische GRM. Dit omdat de parametrische IRT analyse minder valide resultaten opleverde vanwege het relatief kleine aantal proefpersonen. De resultaten van de item-fit zouden eventueel als een indicatie hiervoor kunnen dienen, aangezien alle items een  $\chi^2$  hebben die nagenoeg nul is. Deze resultaten dienen vanwege het kleine aantal proefpersonen voorzichtig te worden behandeld.

Uit de nonparametrische IRT analyse met betrekking tot het onderzoeken van de monotonie blijkt dat de subschaal Gevoel van Eigenwaarde volgens de door Mokken aanbevolen grenzen een gemiddelde schaal is, 'Schoolvaardigheden' een redelijke schaal en 'Sportieve vaardigheden' een zwakke schaal. Met betrekking tot de unidimensionaliteit van de schalen zou voorzichtig geconcludeerd kunnen worden dat de subschalen Schoolvaardigheden en Gevoel van Eigenwaarde unidimensioneel zijn, maar dat er grote vraagtekens zijn bij Sportieve vaardigheden wat betreft de unidimensionaliteit.

Wat betreft de samenhang tussen persoonlijkheid van een kind tussen de 8 en 12 jaar en zijn/haar zelfbeeld kan gezegd worden dat daar veel significante verbanden gevonden zijn. Met name de persoonlijkheidstrekken Inadequatie en Volharding spelen een belangrijke rol bij dit verband. Inadequatie speelt in negatieve zin een rol, namelijk kinderen die veel inadequate gevoelens ervaren zullen een slechter zelfbeeld hebben dan kinderen die weinig inadequate gevoelens kennen. Volharding daarentegen heeft een positief effect op het zelfbeeld; als men volhardend is, zal men ook een beter zelfbeeld hebben.

Toch is het verband tussen persoonlijkheid van een kind tussen de 8 en 12 jaar en het zelfbeeld een verband dat mijns inziens beter onderzocht zou moeten worden. Ten eerste alleen al omdat er in de literatuur nauwelijks een duidelijk onderscheid te vinden is tussen deze twee begrippen en dan vooral in deze leeftijdscategorie. En ten tweede vanwege het 'person-situation debate' wat in 1968 door Walter Mischel werd aangewakkerd naar aanleiding van zijn boek genaamd *Personality and Assessment*. Het debat komt in het kort op de volgende vraag neer: Wat is belangrijker bij de bepaling van wat mensen doen, de persoon of de situatie? (Funder, 2004). Het is namelijk opvallend dat persoonlijkheid, waarvan verondersteld wordt dat het constant is, vaststaat en situatieafhankelijk is (Caspi, 2000; Caspi & Roberts, 2001), zoveel significante correlaties oplevert (zie *Resultaten, Correlatieanalyse subschalen NPV-J en CBSK*) met zelfbeeld, waarvan gezegd wordt dat het situatieafhankelijk en veranderlijk is. Hoe kan persoonlijkheid, dat situatieafhankelijk is, zelfbeeld, dat situatieafhankelijk is, bepalen? Dat is een vraag die verband houdt met het oude 'person-situation debate' en die nader onderzoek behoeft.

Een andere vraag die ook voor verder onderzoek gesteld zou kunnen worden en die meer gericht is op de inhoud van de vragenlijst zelf is de vraag hoe vaak een vragenlijst kinderen eenzelfde soort vraag kan laten beantwoorden voordat ze het vervelend en storend gaan vinden en of hier verschil in is tussen verschillende leeftijdsgroepen. Opvallend tijdens de groepsgewijze afname van de CBSK was dat ik kinderen uit groep 8 terug zag bladeren. Wanneer ik hen er keer op keer op wees dat ze niet moesten terugbladeren, zeiden ze telkens dat ze dat deden omdat ze het idee hadden dat ze die vraag al eerder beantwoord hadden. In de groepen 4 tot en met 7 is dit niet voorgekomen.

Wat er verder in groep 8 en ook in groep 7 en in mindere mate in groep 6 gebeurde was dat een (groot) aantal kinderen veel moeite had met het beantwoorden van vragen uit de

CBSK zoals item 4 uit de subschaal Sportieve vaardigheden ('Sommige kinderen vinden dat ze beter zijn in sport en gymnastiek dan andere kinderen' vs 'Andere kinderen vinden dat ze minder goed zijn in sport en gymnastiek'). De kinderen vonden eigenlijk niet dat ze die vraag konden beantwoorden want dat zou toch "dik arrogant" zijn als je zou invullen dat je jezelf beter vindt dan anderen. De vraag van sociaal wenselijkheid komt hieruit naar boven. Er zou verder onderzoek gedaan kunnen worden naar het sociaal wenselijk invullen van kinderen op zelfrapportage vragenlijsten.

Al dit soort onderzoek kan ons helpen bij het begrijpen van het verband tussen persoonlijkheid en zelfbeeld bij kinderen tussen 8 en 12 jaar en kan ons tevens meer inzicht verschaffen in de accuraatheid van zelfrapportage vragenlijsten bij kinderen in deze leeftijdscategorie mbt het stellen van hetzelfde soort vragen en sociaal wenselijkheid.

## Literatuur

Caspi, A. (2000). The child is father of the man: Personality continuities from childhood to adulthood. *Journal of Personality and Social Psychology*, 78 (1), 158-172.

Caspi, A., & Roberts, B. W. (2001). Personality development across the life course: The argument for change and continuity. *Psychological Inquiry*, 12 (2), 49-66.

Dooley, D. (2001). *Social Research Methods*. New Jersey, Prentice Hall.

Embretson, S. E., & Reise, S. P. (2000). *Item Response Theory for Psychologists*. Mahwah, New Jersey: Lawrence Erlbaum Associates, Publishers.

Emons, W. H. M., Meijer, R. R., & Denollet, J. (2005). Construct Validity of Type D Personality. *Submitted for publication*.

Funder, D. C. (2004). *The Personality Puzzle*. New York, W. W. Norton & Company.

Gough, H. G. (1987). *California Psychological Inventory administrator's guide*. Palo Alto, CA: Consulting Psychologists Press.

Graziano, W. G., Jensen-Campbell, L. A., & Sullivan-Logan, G. M. (1998). Temperament, Activity, and Expectations for Later Personality Development. *Journal of Personality and Social Psychology*, 74, 1266-1277.

Hakstian, A. R., & Farrell, S. (2001). An Openness scale for the California Psychological Inventory. *Journal of Personality Assessment*, 76, 107-134.

Houck, G.M., & Spegman, A. M. (1999). The development of self: Theoretical understandings and conceptual underpinnings. *Infants and young children*, 12 (1), 1-16.

Jacobs, J. E., Bleeker, M. M., & Constantino, M. J. (2003). The self-system during childhood and adolescence: development, influences, and implications. *Journal of Psychotherapy Integration*, 13 (1), 33-65.

- Judge, T. A., Martocchio, J. J., & Thoresen, C. J. (1997). Five-Factor Model of Personality and Employee Absence. *Journal of Applied Psychology*, 82 (5), 745-755.
- Kaplan, R. M., & Saccuzzo, D. P. (2001). *Psychological testing. Principles, Applications, and Issues*. Belmont, Wadsworth/Thomson Learning.
- Luteijn, F., Dijk, H. van, & Ploeg, F. A. E. van der (1989). *Junior Nederlandse Persoonlijkheidsvragenlijst: handleiding*. Amsterdam, Harcourt Assessment BV.
- Luteijn, F., Starren, J., & Dijk, H. van (2001) *Nederlandse Persoonlijkheidsvragenlijst: handleiding*. Lisse, Swets Test Publishers.
- McCrae, R. R., Costa, P. T., & Piedmont, R. L. (1993). Folk concepts, natural language, and psychological constructs: The California Psychological Inventory and the Five-Factor Model. *Journal of Personality*, 61, 1-26.
- Meijer, R. R., & Baneke, J. J. (2004). Analyzing psychopathology items: A case for nonparametric item response theory modeling. *Psychological Methods*, 9 (3), 354-368.
- Mokken, R. J. (1971). *A theory and procedure of scale analysis*. Den Haag, Mouton.
- Mokken, R. J. (1997). Nonparametric models for dichotomous responses. In W. J. van der Linden & R. K. Hambleton (Eds.), *Handbook of modern item response theory* (pp. 351-367). New York, Springer-Verlag.
- Moore, D. S., & McCabe, G. P. (2001). *Statistiek in de Praktijk: Theorieboek*. Schoonhoven, Academic Service.
- Muris, P., Meesters, C., & Fijen, P. (2003). The Self-Perception Profile for Children: further evidence for its factor structure, reliability and validity. *Personality and Individual Differences*, 35, 1791-1802.
- Rothbart, M. K., Ahadi, S. A., & Evans, D. E. (2000). Temperament and Personality: Origins and Outcomes. *Journal of Personality and Social Psychology*, 78 (1), 122-135.
- Schultz, D. P., & Schultz, S. E. (2001). *Theories of Personality*. Belmont, Wadsworth/Thomson Learning.
- Shiner, R., & Caspi, A. (2003). Personality differences in childhood and adolescence: measurement, development, and consequences. *Journal of Child Psychology and Psychiatry*, 44, 2-32.
- Sigelman, C. K., & Rider, E. A. (2003). *Life-Span Human Development*. Belmont, Wadsworth/Thomson Learning.
- Sijtsma, K., & Molenaar, I. W. (2002). *Introduction to nonparametric item response theory*. Thousand Oaks, CA: Sage.



- Srivastava, S., John, O. P., Gosling, S. D., & Potter, J. (2003). Development of Personality in Early and Middle Adulthood: Set Like Plaster or Persistent Change? *Journal of Personality and Social Psychology*, 84 (5), 1041–1053.
- Trzesniewski, K. H., Donnellan, M. B., & Robins, R. W. (2003). Stability of self-esteem across the life span. *Journal of Personality and Social Psychology*, 84 (1), 205-220.
- Veerman, J. W., Brink, L. T. ten, Straathof, M. A. E., & Treffers, Ph. D. A. (1996). Measuring children's self-concept with a Dutch version of the "Self-Perception Profile for Children". Factorial validity and invariance across a nonclinic and a clinic group. *Journal of Personality Assessment*, 67, 142-154.
- Veerman, J. W., Straathof, M. A. E., Treffers, Ph. D. A., Bergh, B. R. H. van den, & Brink, L. T. ten (1997, 2004). *Competentiebelevingsschaal voor Kinderen: handleiding*. Lisse, Harcourt Assessment BV.

## Bijlage 1

### Inhoud subschalen NPV-J

1. Inadequatie (IN) de stellingen vragen naar vage lichamelijke klachten, een gedrukte stemming, vage angsten en insufficiëntie-gevoelens
2. Volharding (VO) de stellingen vragen naar een positieve taakopvatting, het goed aangepast zijn aan de eisen van het (school)werk, het willen beantwoorden van hoge verwachtingen, het zich aan afspraken willen houden en het ordelijk laten verlopen van de dingen
3. Sociale inadequatie (SI) de stellingen vragen naar het vermijden van of het zich ongelukkig voelen in sociale contacten
4. Recalcitrantie (RE) de stellingen vragen naar het zich willen afzetten tegen anderen, het wantrouwen van anderen en het alleen zijn zaken willen oplossen
5. Dominantie (DO) de stellingen vragen naar de baas willen spelen en vertrouwen hebben in het eigen kunnen

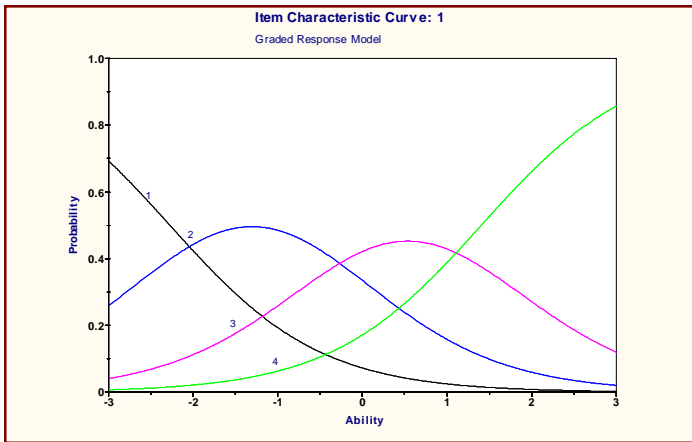
## Bijlage 2

### Inhoud subschalen CBSK

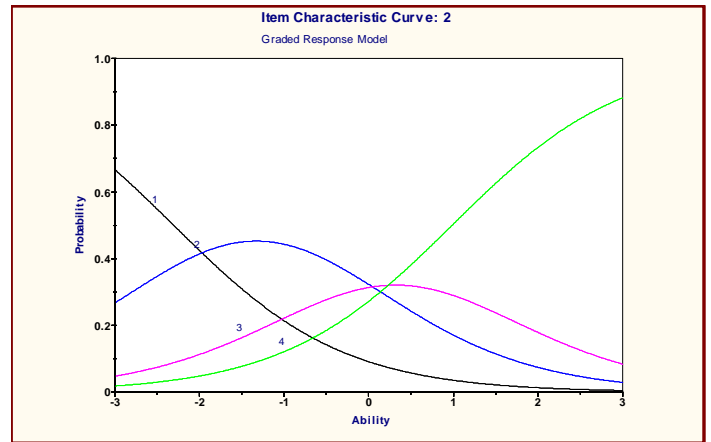
- |                                   |   |
|-----------------------------------|---|
| 1. Schoolvaardigheden<br>(SV)     | Hoe goed (slim, knap) vindt het kind zichzelf op school, hoe kijkt het kind tegen zijn schoolprestaties aan?  |
| 2. Sociale acceptatie<br>(SA)     | Vindt het kind van zichzelf dat het ‘erbij hoort’? Meent hij/zij genoeg vrienden te hebben, kan hij/zij makkelijk vrienden maken? Denkt het kind geliefd te zijn?       |
| 3. Sportieve vaardigheden<br>(SP) | Meent het kind goed mee te kunnen komen in sport, gymnastiek en buitenspelletjes?   |
| 4. Fysieke verschijning<br>(FV)   | Hoe beoordeelt het kind zijn uiterlijk? Hoe vindt hij/zij zichzelf eruit zien?  |
| 5. Gedragshouding<br>(GH)         | Vindt het kind dat het zich behoorlijk, ‘netjes’ gedraagt, geen dingen doet die (moreel) niet mogen? Denkt het kind zich te gedragen zoals van hem/haar wordt verwacht? |
| 6. Gevoel van Eigenwaarde<br>(GE) | Hoe beoordeelt het kind zichzelf in het algemeen, als persoon? Hoe is zijn/haar algemeen gevoel van eigenwaarde?  |

# Bijlage 3

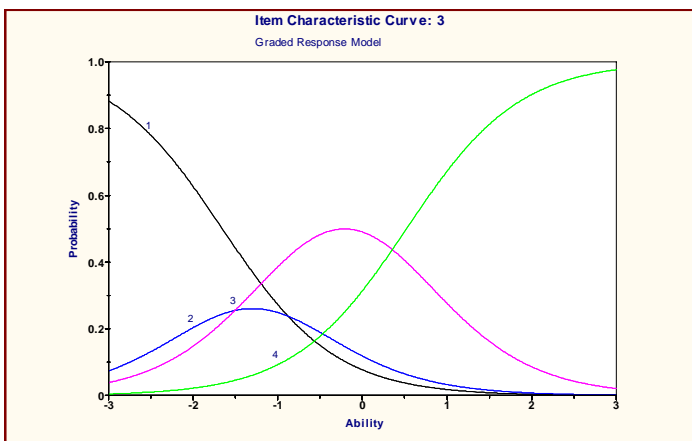
## Item Characteristic Curves voor de subschaal Schoolvaardigheden



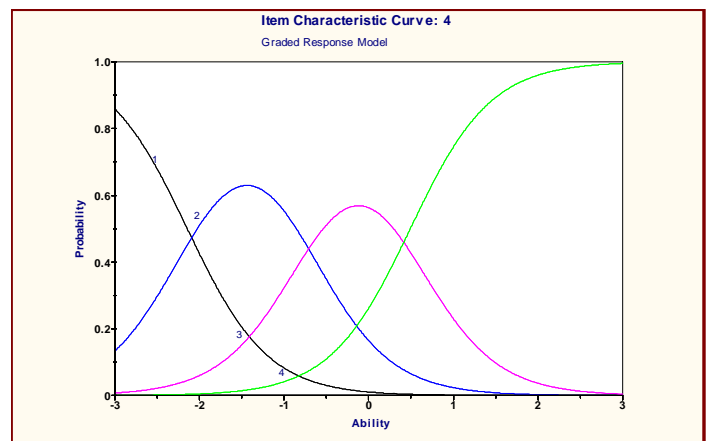
Item 1



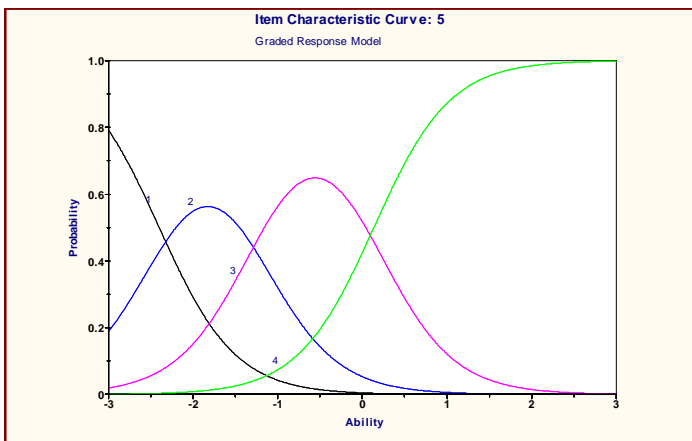
Item 2



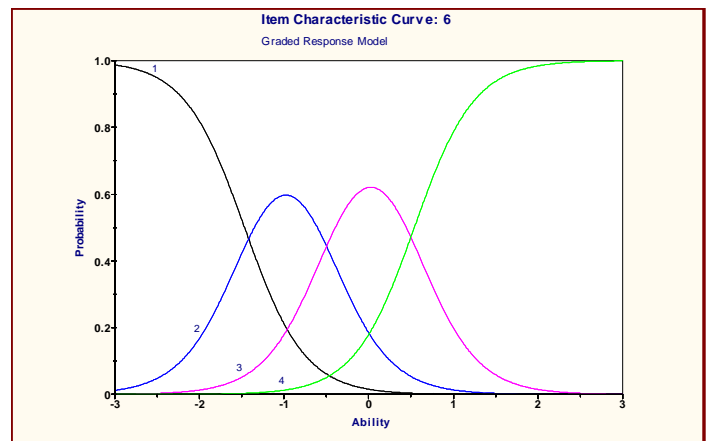
Item 3



Item 4



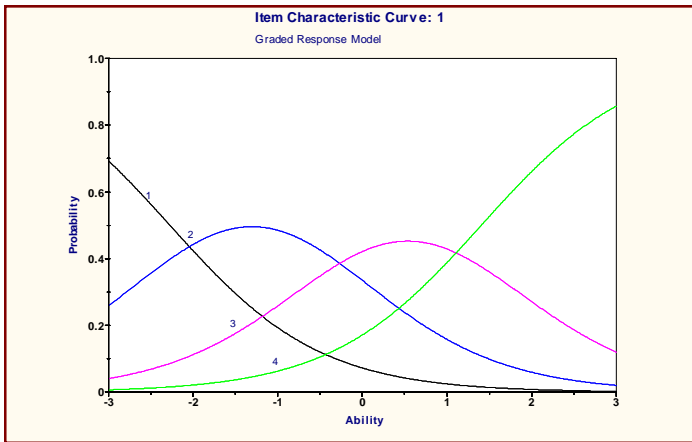
Item 5



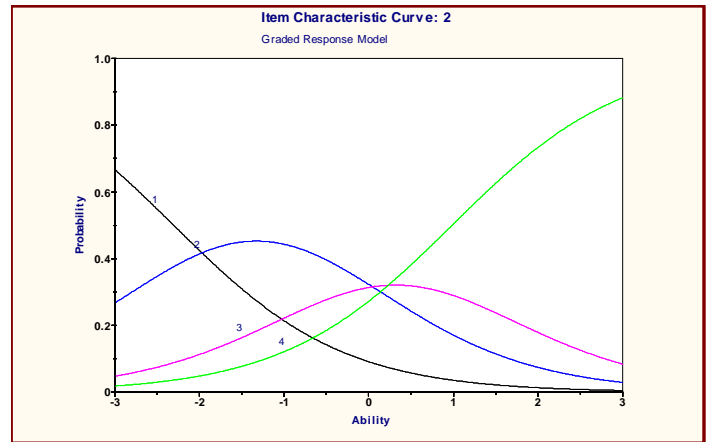
Item 6

## Bijlage 4

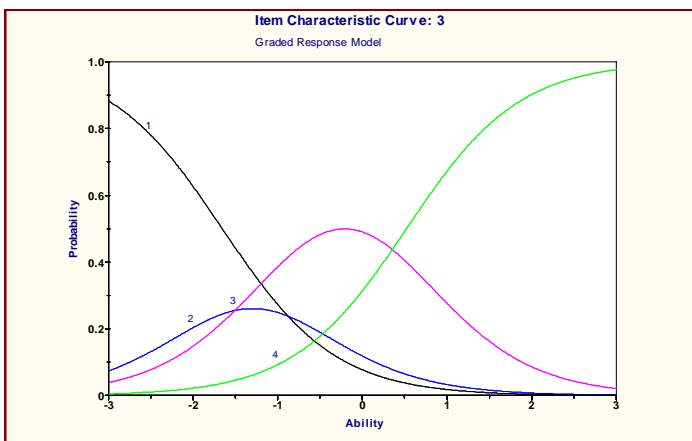
### Item Characteristic Curves voor de subschaal Sportieve Vaardigheden



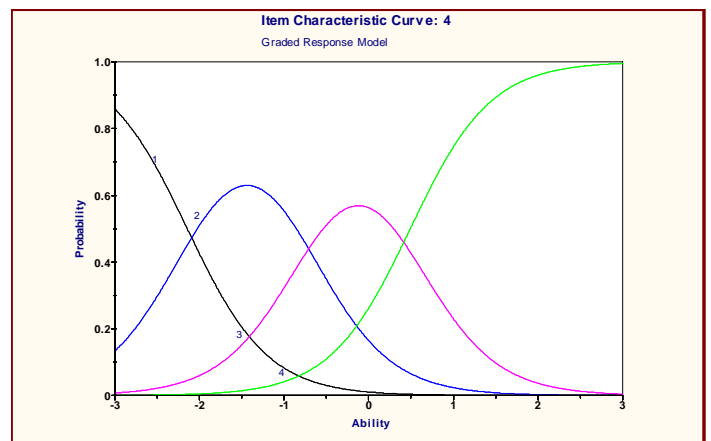
Item 1



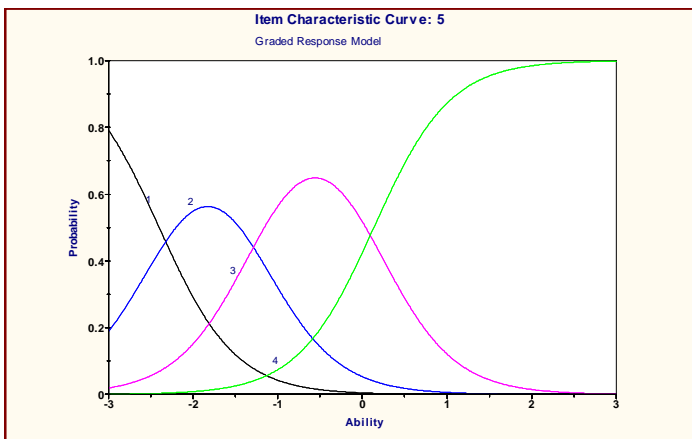
Item 2



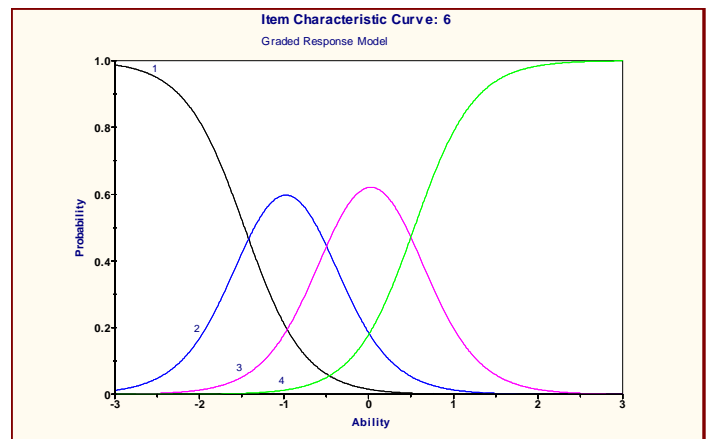
Item 3



Item 4



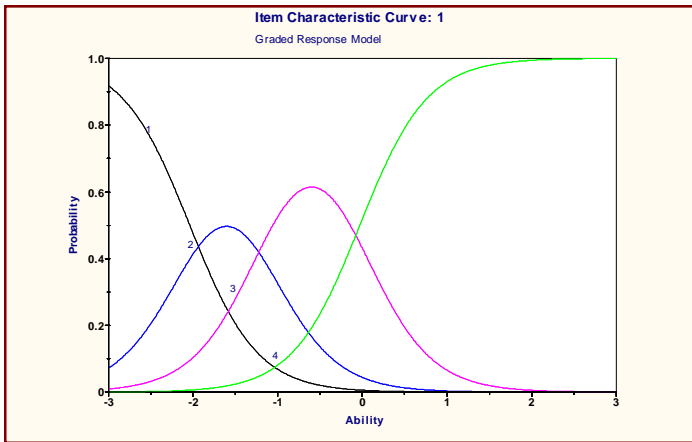
Item 5



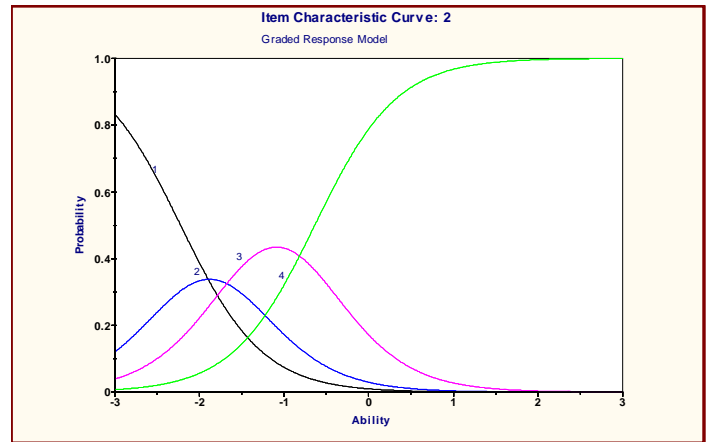
Item 6

## Bijlage 5

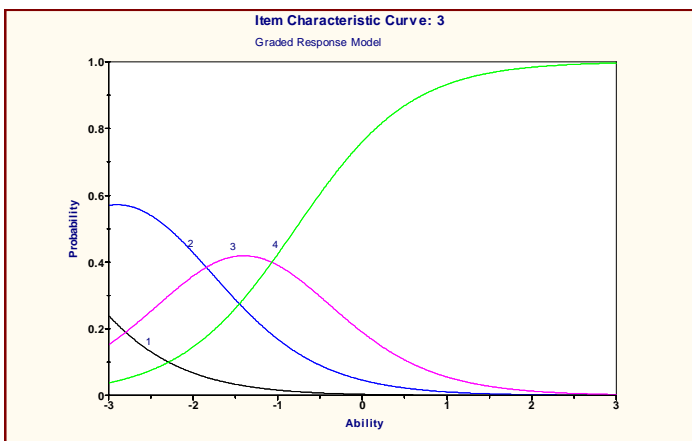
### Item Characteristic Curves voor de subschaal Gevoel van Eigenwaarde



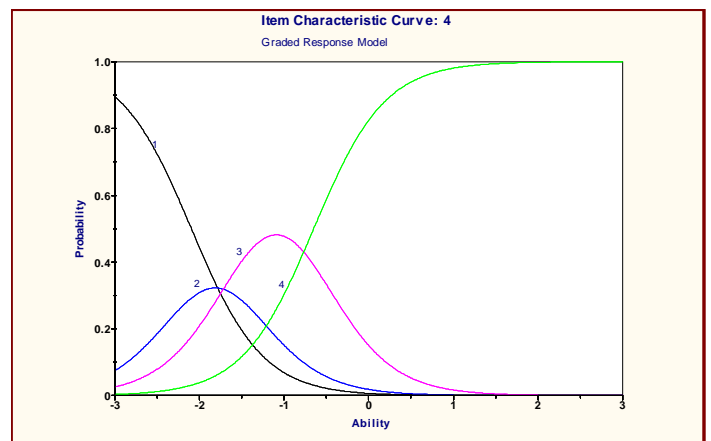
Item 1



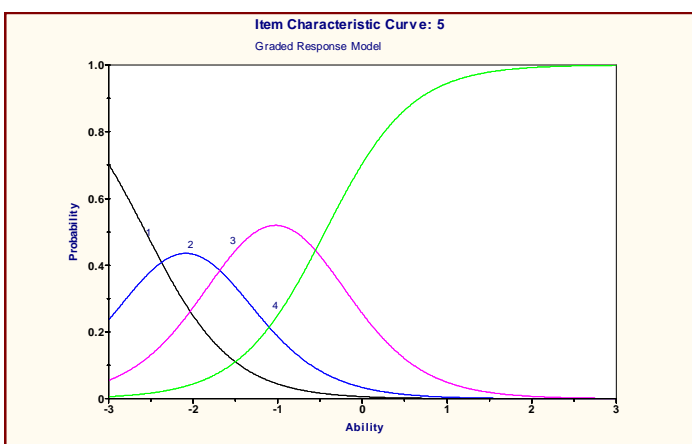
Item 2



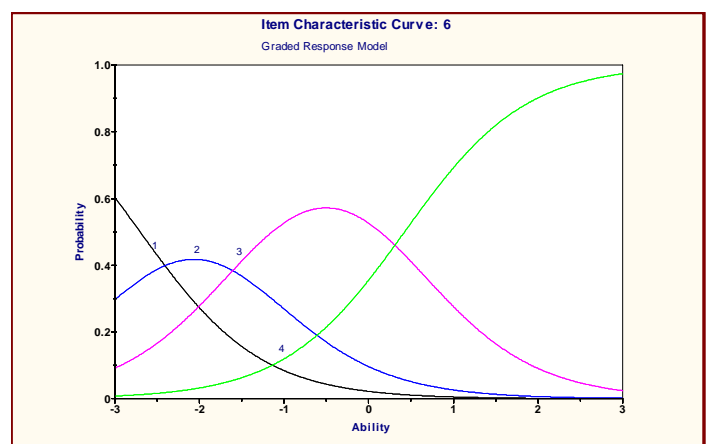
Item 3



Item 4



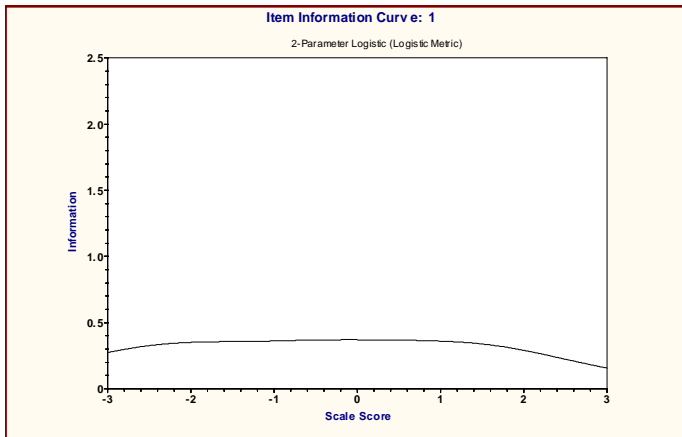
Item 5



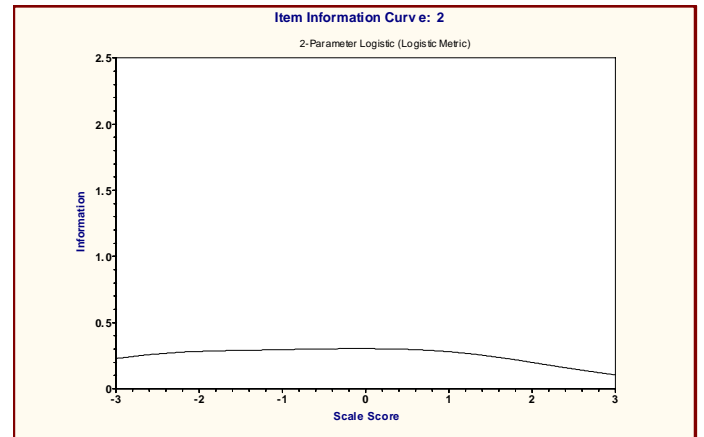
Item 6

## Bijlage 6

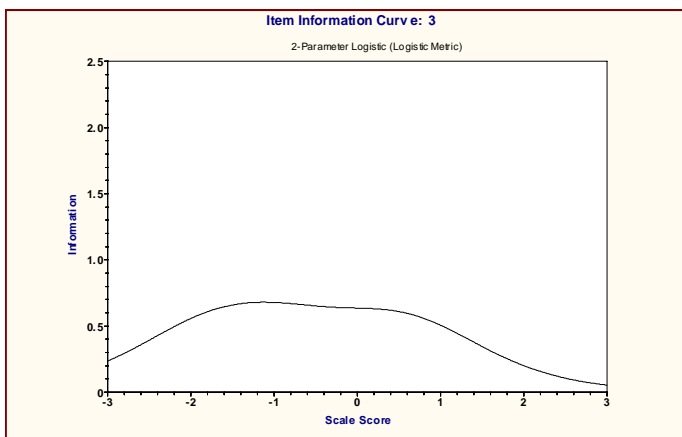
### Item-Informatiecurves voor de subschaal Schoolvaardigheden



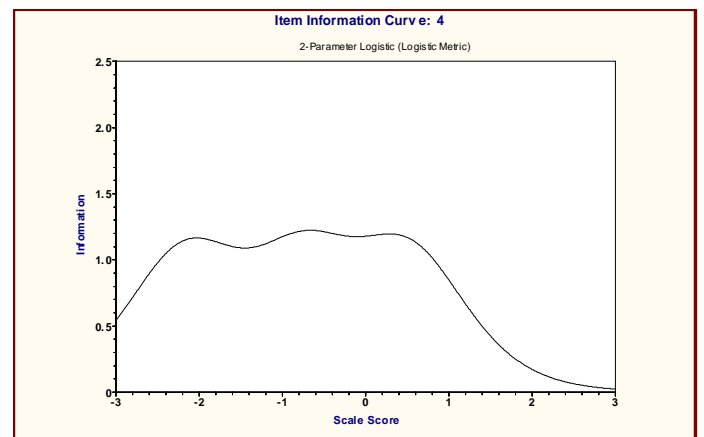
Item 1



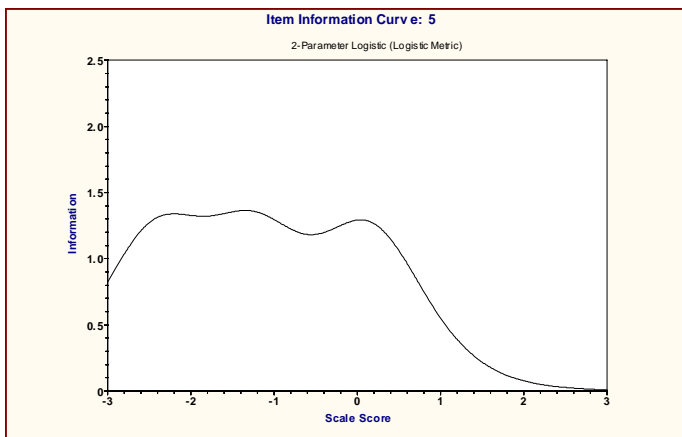
Item 2



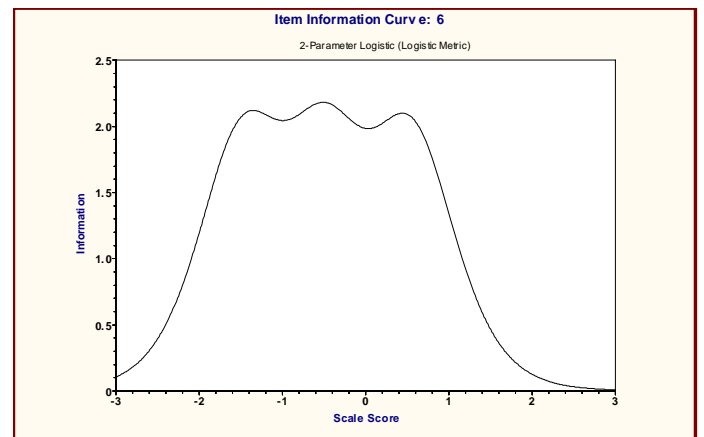
Item 3



Item 4



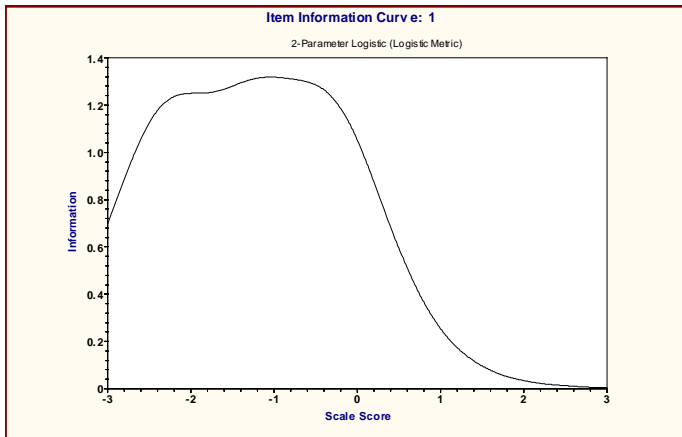
Item 5



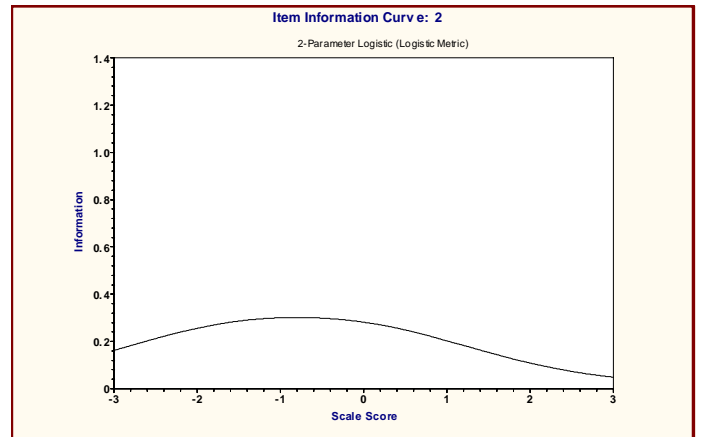
Item 6

# Bijlage 7

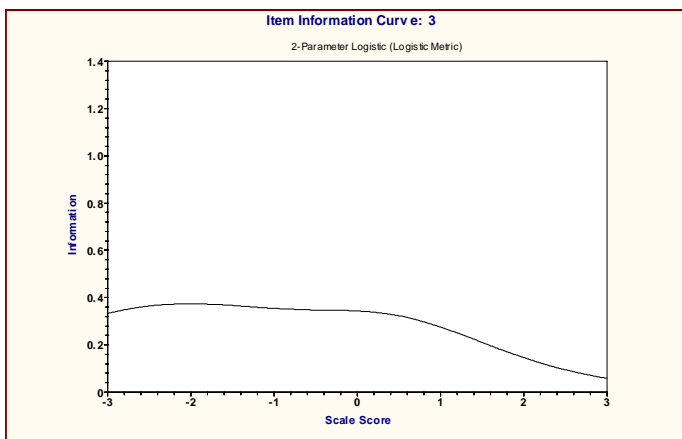
## Item-Informatiecurves voor de subschaal Sportieve Vaardigheden



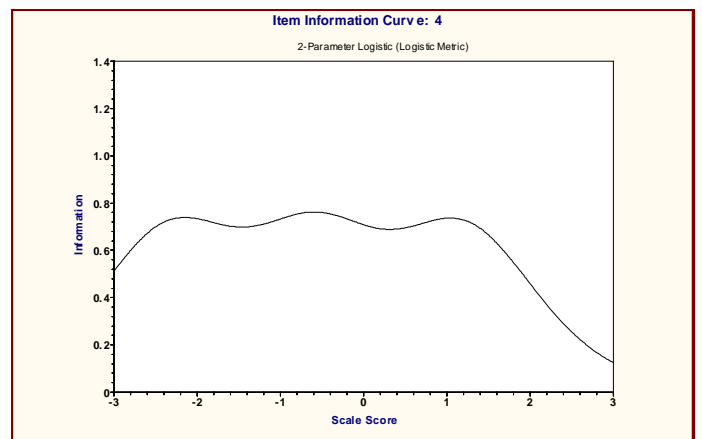
Item 1



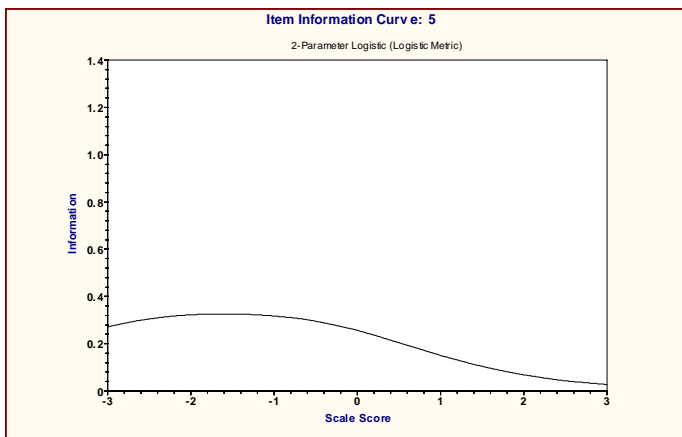
Item 2



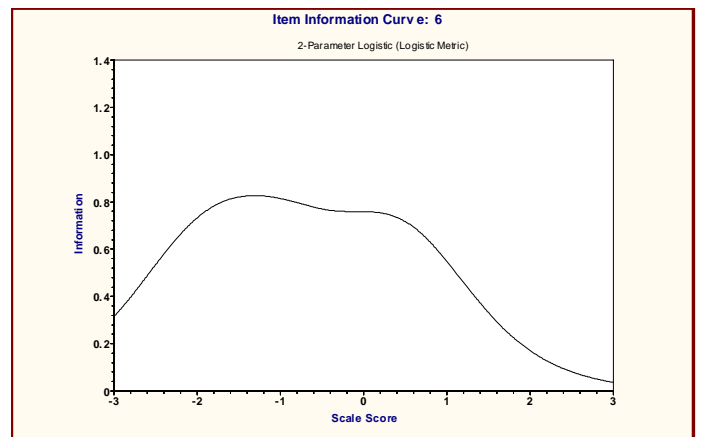
Item 3



Item 4



Item 5

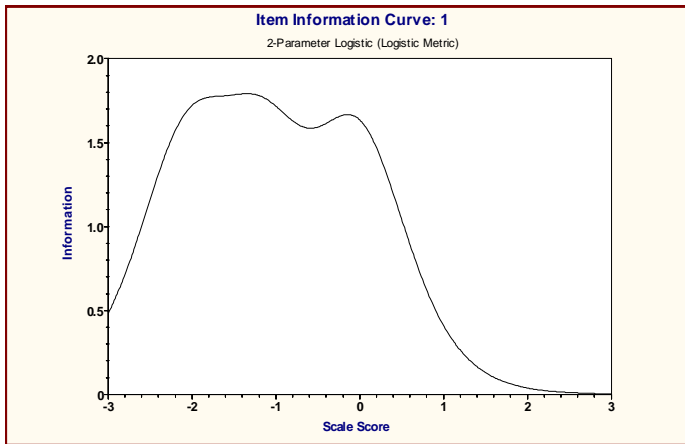


Item 6

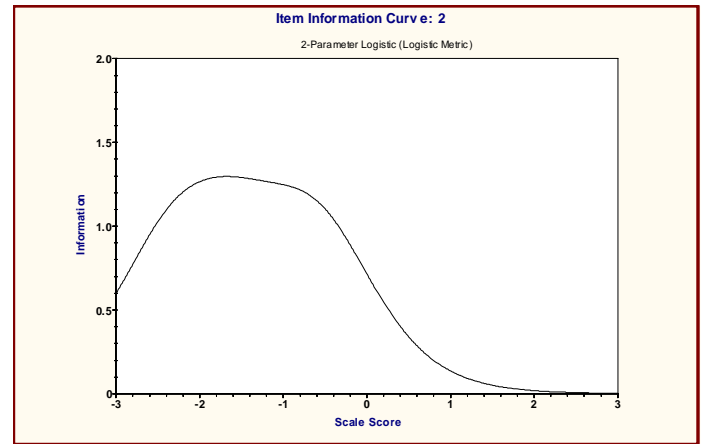


# Bijlage 8

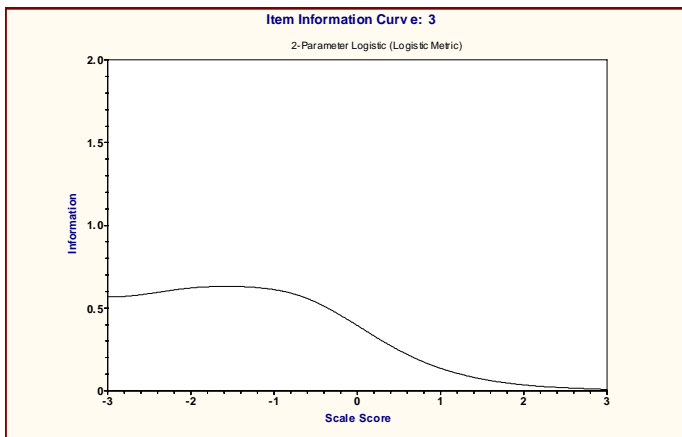
## Item-Informatiecurves voor de subschaal Gevoel van Eigenwaarde



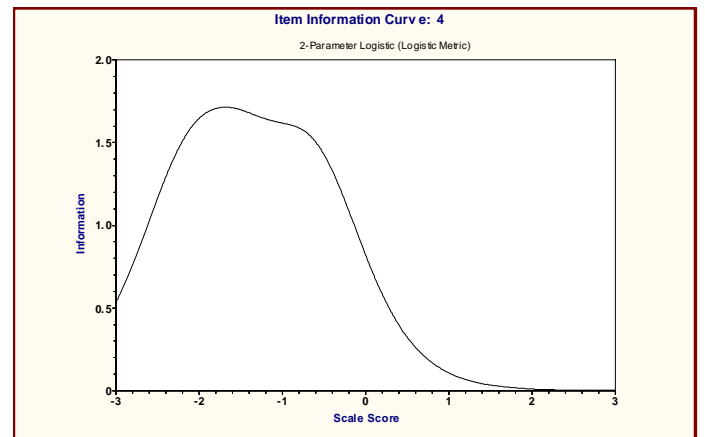
Item 1



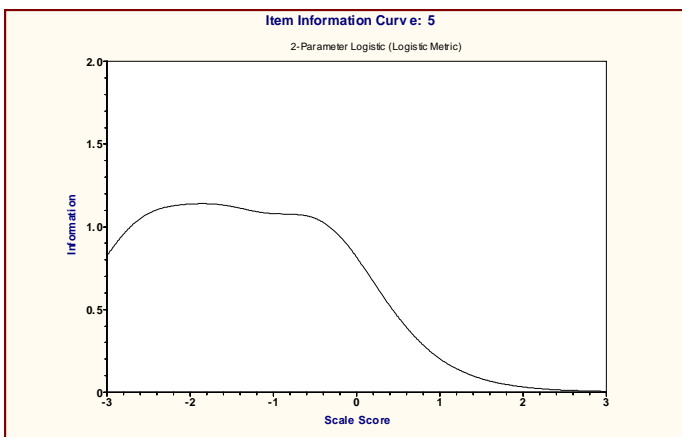
Item 2



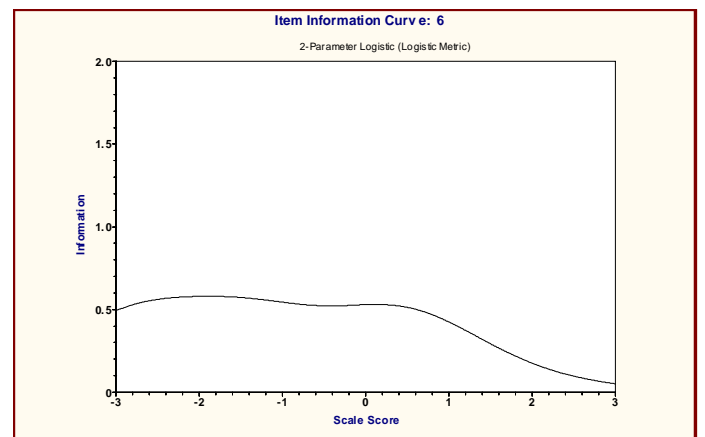
Item 3



Item 4



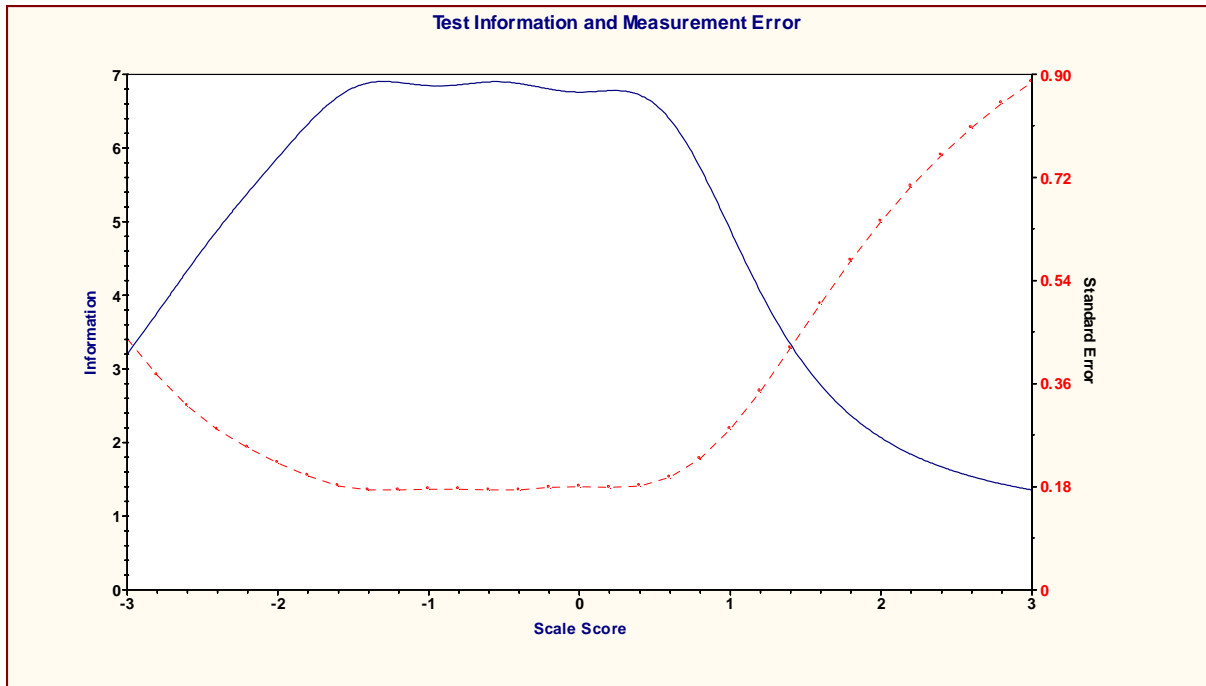
Item 5



Item 6

## Bijlage 9

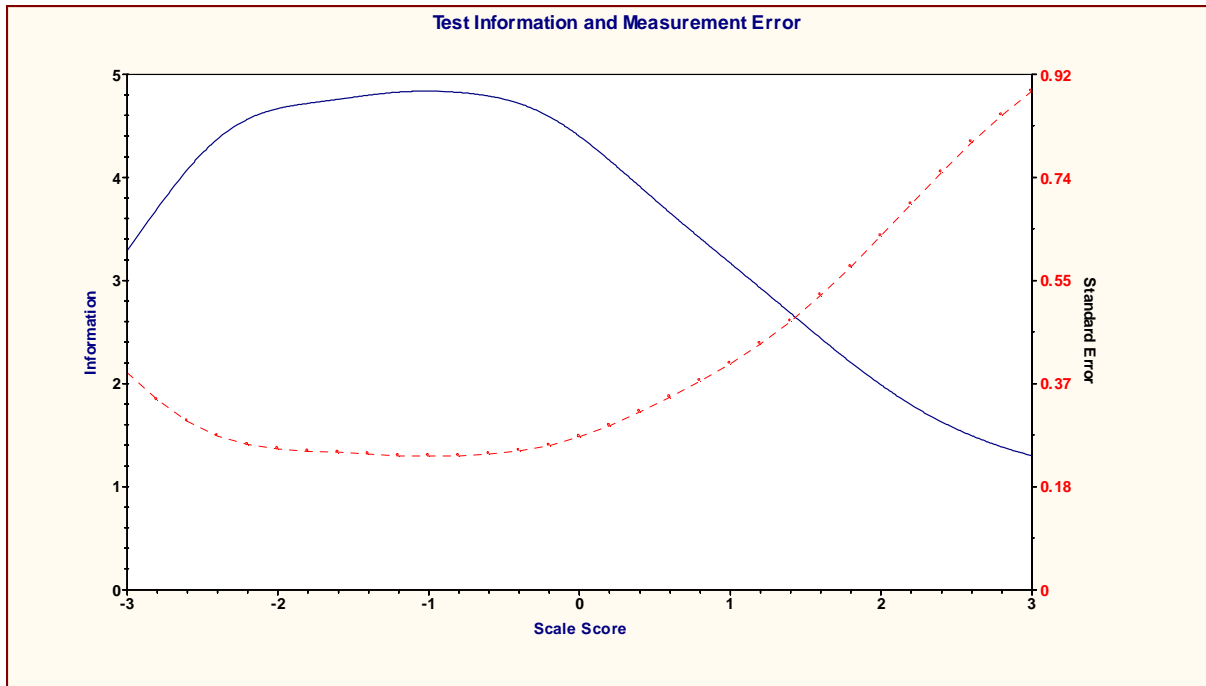
### Test Informatie Curve van de subschaal Schoolvaardigheden



- = Test Informatie Curve
- - - = Standaard fouten 'curve'

## Bijlage 10

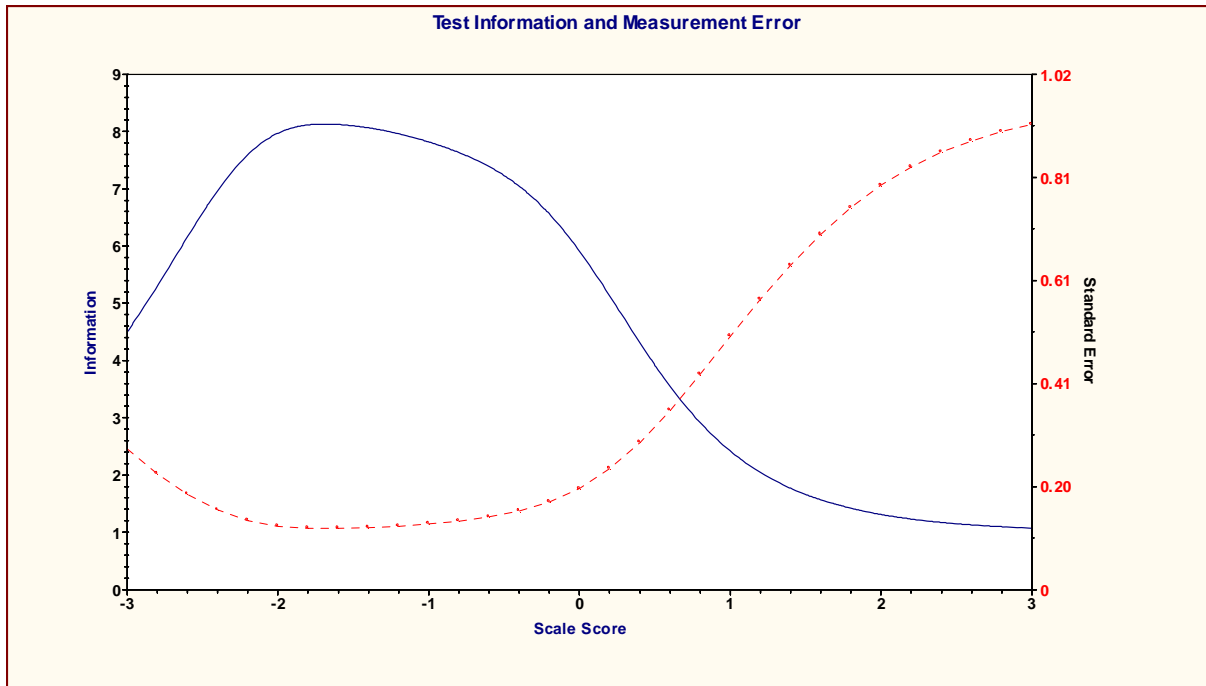
### Test Informatie Curve van de subschaal Sportieve Vaardigheden



- = Test Informatie Curve
- - - = Standaard fouten 'curve'

## Bijlage 11

### Test Informatie Curve van de subschaal Gevoel van Eigenwaarde



- = Test Informatie Curve
- - - = Standaard fouten 'curve'